



**Benemérita Universidad Autónoma de Puebla**  
**Facultad de Ciencias de la Computación**

Clasificación multi-etiqueta de comentarios al estudiante en  
cursos en línea

Tesis

Que para obtener el grado de:

**Doctor en Ingeniería del Lenguaje y Conocimiento**

Presenta:

**Dorian Ruiz Alonso**

Directores:

Claudia Zepeda Cortés  
Hilda Castillo Zacatelco  
José Luis García Cué

Puebla, Puebla

Noviembre del 2021

# Agradecimientos

Quiero agradecer a Dios y a la vida por permitirme llegar a esta etapa.

También, agradecer a la Dra. Claudia Zepeda Cortés, quien fue la directora de tesis, tuvo la confianza en que podría sacar adelante este reto, y me apoyó en el desarrollo de este trabajo con valiosos aportes y consejos.

Además, a Dra. Hilda Castillo Zacatelco, Dra. Maria Auxilio Medina Nieto, Dr. José García Cué, Dr. José Luis Carballido Carranza, Dra. Reyna Carolina Medina Ramírez, Dra. Patricia Silva Sánchez quienes con sus comentarios en las revisiones parciales me permitieron concluir con este trabajo.

Por último, a mi familia y amigos quienes han estado presentes en cada momento de mi vida.



# Resumen

Este trabajo trata sobre minería de textos educativa, un campo del procesamiento de lenguaje natural aplicado a la educación. La investigación tuvo por objetivo, proponer una metodología para analizar automáticamente los comentarios resultantes de la revisión y el análisis del docente a las actividades que envía el estudiante en cursos en línea de la licenciatura en derecho de la plataforma Blackboard en los niveles tarea, proceso, regulación, elogios y otros, categorizando los resultados mediante técnicas de minería de clasificación multi-etiqueta para estimar y evaluar la toma de decisiones respecto al desempeño docente.

El incremento de estudiantes que se matricula en cursos de educación en línea magnifica los retos de docentes para proveer retroalimentación informativa y de alta calidad que ayude a los estudiantes en su autorregulación. Este hecho, hace necesario la construcción de herramientas que permitan analizar automáticamente las retroalimentaciones que generan los docentes para identificar áreas de oportunidad que les permitan mejorar su práctica docente.

Se identificó que en la institución donde se realizó la investigación, no existía una herramienta para el análisis de comentarios al estudiante en cursos en línea quedando solo almacenados en el sistema de gestor de aprendizaje. También, de lo revisado en las referencias documentales se detectó que existe poca información sobre metodologías para el analizar automáticamente comentarios generados por docentes en cursos en línea escritos en lenguaje español. Los trabajos existentes utilizan técnicas de aprendizaje de máquina convencionales, que se restringen a problemas con clases mutuamente excluyentes. A diferencia de estos trabajos, en este, se propone utilizar estrategias de clasificación multi-etiqueta.

La metodología se aplicó en un caso de estudio, en la que se recopilaron del sistema gestor de aprendizaje Blackboard 11013 comentarios al estudiante de 121 cursos en línea de la licenciatura en derecho de una universidad pública de México. Los comentarios al estudiante fueron clasificados manualmente por expertos siguiendo el modelo de retroalimentación de Hattie y Timperley. Después, los comentarios al estudiante fueron preprocesados y se calcularon las frecuencias de términos con diferentes n-gramas. Luego, los comentarios al estudiante se partitionaron en dos subconjuntos para realizar el entrenamiento y prueba de los clasificadores multi-etiqueta.

En total, se entrenaron 122 clasificadores multi-etiqueta. En una etapa inicial se entrenaron 108 clasificadores para analizar el enfoque, algoritmo de clasificación, método

de transformación, características que mejor desempeño permiten alcanzar en la tarea de ubicar los comentarios al estudiante en los niveles tarea, proceso, regulación, elogios y otros con respecto a lo etiquetado manualmente por expertos. También, se analiza si la afinación de hiper-parámetros mejora el desempeño de los clasificadores multi-etiqueta.

Los resultados, muestran que las técnicas de clasificación multi-etiqueta permiten crear clasificadores capaces de ubicar automáticamente los comentarios al estudiante en los niveles tarea, proceso, regulación, elogios y otros propuestos por Hattie y Timperley de manera aproximada a como lo harían manualmente expertos. También, de la comparación de los enfoques de clasificación multi-etiqueta, se detecta que el enfoque de transformación tiene mejor desempeño que el enfoque de adaptación. De igual forma, de la comparación de los algoritmos de clasificación utilizados se concluye que son los de máquinas de soporte de vectores y bosques aleatorios los que mejor desempeño muestran para clasificar los comentarios al estudiante con respecto a lo clasificado manualmente por los expertos cuando se utiliza el enfoque de transformación. Mientras que el algoritmo de vecinos más cercanos multi-etiqueta es el de mejor desempeño cuando se opta por en el enfoque de adaptación.

Se concluye que todos los métodos de transformación permiten un buen desempeño de clasificadores multi-etiqueta, destacando la relevancia binaria y el conjunto potencia, con este último se debe asegurar que se tengan todas las posibles combinaciones de etiquetas en los conjuntos de prueba y entrenamiento.

Del estudio de características se concluye que el uso un-grama permite un mejor desempeño en la clasificación de los comentarios al estudiante en los niveles propuestos por Hattie y Timperley con respecto a lo clasificado manualmente por los expertos.

Por último, de la evaluación de la afinación de hiper-parámetros se concluye que aplicar esta etapa permite mejorar el desempeño de clasificadores multi-etiqueta para el conjunto de datos utilizado en este trabajo.

# Contenido

<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Justificación . . . . .	3
1.3. Objetivos . . . . .	4
1.4. Preguntas de investigación . . . . .	4
1.5. Hipótesis . . . . .	5
1.6. Propuesta de solución . . . . .	5
1.7. Organización del documento . . . . .	7
<b>2. Marco Teórico</b>	<b>9</b>
2.1. Educación en línea . . . . .	9
2.1.1. Evolución de la educación en línea . . . . .	10
2.1.2. Características de la educación en línea . . . . .	12
2.1.3. Ventajas y desventajas de la educación en línea . . . . .	12
2.1.4. Calidad de la educación en línea . . . . .	13
2.1.5. Sistemas gestores de aprendizaje . . . . .	14
2.1.6. Actividades de aprendizaje en educación en línea . . . . .	16
2.2. Retroalimentación . . . . .	17
2.2.1. Importancia de la retroalimentación . . . . .	18
2.2.2. Elementos de la retroalimentación . . . . .	20
2.2.3. Análisis de la retroalimentación . . . . .	20
2.3. Modelo de retroalimentación Hattie y Timperlay . . . . .	21
2.3.1. Retroalimentación sobre la tarea . . . . .	25
2.3.2. Retroalimentación sobre el procesamiento de la tarea . . . . .	25
2.3.3. Retroalimentación sobre auto-regulación . . . . .	25
2.3.4. Retroalimentación sobre el yo como persona . . . . .	26
2.4. Clasificación . . . . .	26
2.4.1. Clasificación binaria . . . . .	26
2.4.2. Clasificación multiclase . . . . .	27
2.4.3. Clasificación multi-etiqueta . . . . .	27
2.4.4. Clasificación multi-dimensional . . . . .	28
2.4.5. Aprendizaje de instancias múltiples . . . . .	28
2.5. Clasificación multi-etiqueta . . . . .	28

2.5.1.	Definición formal del problema . . . . .	28
2.5.2.	Enfoques del aprendizaje multi-etiqueta . . . . .	29
2.5.3.	Métodos de transformación . . . . .	31
2.5.4.	Características de los conjuntos de datos multi-etiqueta . . . . .	32
2.5.5.	Evaluación del desempeño de clasificadores . . . . .	35
2.6.	Afinación de hiper-parámetros . . . . .	38
2.6.1.	Estrategias existentes . . . . .	38
2.6.2.	Medición de la afinación . . . . .	39
<b>3.</b>	<b>Marco Contextual</b>	<b>41</b>
3.1.	Universidad donde se realizó la investigación . . . . .	41
3.1.1.	Oferta educativa . . . . .	41
3.1.2.	Admisión . . . . .	43
3.1.3.	Matrícula . . . . .	43
3.1.4.	Egreso . . . . .	44
3.1.5.	Docentes . . . . .	44
3.2.	Modalidad a distancia en BUAP . . . . .	45
3.2.1.	Modelo pedagógico de la modalidad a distancia en BUAP . . . . .	46
3.2.2.	Ingreso y matrícula de la modalidad a distancia en BUAP . . . . .	47
3.2.3.	Egreso de la modalidad a distancia en BUAP . . . . .	48
<b>4.</b>	<b>Metodología</b>	<b>51</b>
4.1.	Recolección e integración de comentarios al estudiante . . . . .	52
4.1.1.	Etiquetado manual de comentarios al estudiante . . . . .	53
4.1.2.	Análisis exploratorio de los comentarios al estudiante . . . . .	54
4.2.	Preprocesamiento y extracción de características . . . . .	58
4.2.1.	Conjuntos de entrenamiento y prueba . . . . .	60
4.3.	Afinación de hiper-parámetros . . . . .	62
4.4.	Clasificación . . . . .	63
4.5.	Análisis y Evaluación . . . . .	64
4.5.1.	Selección del enfoque de clasificación multi-etiqueta . . . . .	65
4.5.2.	Selección de algoritmos de clasificación . . . . .	65
4.5.3.	Selección de características . . . . .	66
4.5.4.	Evaluación de la afinación de hiper-parámetros . . . . .	66
<b>5.</b>	<b>Resultados</b>	<b>67</b>
5.1.	Resultados de los experimentos . . . . .	67
5.2.	Estudio comparativo de enfoques de clasificación multi-etiqueta . . . . .	75
5.3.	Estudio comparativo de algoritmos . . . . .	77
5.4.	Estudio comparativo de las características . . . . .	81
5.5.	Estudio comparativo de los métodos de transformación . . . . .	87
5.6.	Evaluación de la afinación de hiper-parámetros . . . . .	88

<i>CONTENIDO</i>	VII
<b>6. Conclusiones</b>	<b>91</b>
6.1. Trabajo futuro . . . . .	92
<b>7. Anexo A</b>	<b>99</b>



# Índice de figuras

2.1. Modelo de retroalimentación Hattie y Timperley (2007). Traducido por Dra. Katherina Gallardo. . . . .	23
2.2. Tipos de clasificación según el conjunto de salida . . . . .	27
3.1. Estudiantes admitidos por periodo escolar. . . . .	43
3.2. Distribución de estudiantes admitidos por nivel educativo. . . . .	43
3.3. Distribución de estudiantes matriculados por nivel educativo en diferentes periodos. . . . .	44
3.4. Distribución de estudiantes matriculados por nivel educativo en el periodo 2019-2020. . . . .	44
3.5. Distribución de tipo de contratación de docentes. . . . .	45
3.6. Estudiantes admitidos en modalidad a distancia. . . . .	48
3.7. Estudiantes admitidos en modalidad a distancia por programa educativo. . . . .	48
3.8. Estudiantes matriculados por programa educativo en 2019. . . . .	49
4.1. Metodología . . . . .	51
4.2. Estructura de los experimentos realizados . . . . .	52
4.3. Comentario al estudiante extraído del sistema gestor de aprendizaje Blackboard . . . . .	53
4.4. Comentario al estudiante después de eliminar etiquetas HTML/CSS . . . . .	54
4.5. Gráfica de caja de cantidad de caracteres en comentario al estudiante . . . . .	56
4.6. Gráfica de caja de cantidad de palabras en comentario al estudiante . . . . .	57
4.7. Gráfica de caja de cantidad de caracteres en comentario_limpio . . . . .	57
4.8. Gráfica de caja de cantidad de palabras en Comentario_limpio . . . . .	57
4.9. Distribución de comentarios por nivel . . . . .	58
4.10. Palabras de mayor frecuencia en comentarios al estudiante . . . . .	59
4.11. Palabras de menor frecuencia en comentarios al estudiante . . . . .	60

- 5.1. Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los enfoques transformación y adaptación utilizando un-grama. Para el enfoque de transformación se usan los métodos de relevancia binaria, cadenas de clasificadores y conjunto potencia con los algoritmos máquinas de soporte de vectores y bosques aleatorios. Para el enfoque de adaptación se utilizan los algoritmos máquinas de soporte de vectores multi-etiqueta y vecinos más cercanos multi-etiqueta . . . . . 75
- 5.2. Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los enfoques transformación y adaptación utilizando dos-grama. Para el enfoque de transformación se usan los métodos de relevancia binaria, cadenas de clasificadores y conjunto potencia con los algoritmos máquinas de soporte de vectores y bosques aleatorios. Para el enfoque de adaptación se utilizan los algoritmos máquinas de soporte de vectores multi-etiqueta y vecinos más cercanos multi-etiqueta . . . . . 76
- 5.3. Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los enfoques transformación y adaptación utilizando tres-grama. Para el enfoque de transformación se usan los métodos de relevancia binaria, cadenas de clasificadores y conjunto potencia con los algoritmos máquinas de soporte de vectores y bosques aleatorios. Para el enfoque de adaptación se utilizan los algoritmos máquinas de soporte de vectores multi-etiqueta y vecinos más cercanos multi-etiqueta . . . . . 76
- 5.4. Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los enfoques transformación y adaptación utilizando un-grama con afinación de hiper-parámetros. Para el enfoque de transformación se usan los métodos de relevancia binaria, cadenas de clasificadores y conjunto potencia con los algoritmos máquinas de soporte de vectores y bosques aleatorios. Para el enfoque de adaptación se utiliza el algoritmo vecinos más cercanos multi-etiqueta . . . . . 77
- 5.5. Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los algoritmos máquinas de soporte de vectores (SVM), bosques aleatorios (RF), regresión logística (LR) y naive bayes (NB) con método de transformación de relevancia binaria. . . . . 78
- 5.6. Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los algoritmos máquinas de soporte de vectores (SVM), bosques aleatorios (RF), regresión logística (LR) y naive bayes (NB) con método de transformación de cadenas de clasificadores. . . . . 79
- 5.7. Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los algoritmos máquinas de soporte de vectores (SVM), bosques aleatorios (RF), regresión logística (LR) y naive bayes (NB) con método de transformación de conjunto potencia. . . . . 79

5.8. Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los algoritmos máquinas de soporte de vectores (SVM), bosques aleatorios (RF), regresión logística (LR) y naive bayes (NB) con método de transformación de rakel-d. . . . .	80
5.9. Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los algoritmos adaptados máquinas de soporte de vectores multi-etiqueta (MLTSVN) y vecinos más cercanos multi-etiqueta (MLKNN). . . . .	80
5.10. Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método de relevancia binaria y algoritmo máquinas de soporte de vectores con diferentes n-gramas . . . . .	81
5.11. Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método de relevancia binaria y algoritmo bosques aleatorios con diferentes n-gramas . . . . .	82
5.12. Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método de cadenas de clasificadores y algoritmo máquinas de soporte de vectores con diferentes n-gramas . . . . .	82
5.13. Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método de cadenas de clasificadores y algoritmo bosques aleatorios con diferentes n-gramas . . . . .	83
5.14. Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método conjunto potencia y algoritmo máquinas de soporte de vectores con diferentes n-gramas . . . . .	83
5.15. Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método conjunto potencia y algoritmo bosques aleatorios con diferentes n-gramas . . . . .	84
5.16. Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método rakel-d y algoritmo máquinas de soporte de vectores con diferentes n-gramas . . . . .	84
5.17. Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método rakel-d y algoritmo bosques aleatorios con diferentes n-gramas . . . . .	85
5.18. Segundos para el entrenamiento de clasificadores multi-etiqueta utilizando un-grama, uno-dos-grama y uno-dos-tres-grama. . . . .	85
5.19. Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque adaptación utilizando el algoritmo vecinos más cercanos multi-etiqueta con diferentes n-gramas . . . . .	86
5.20. Segundos para el entrenamiento de clasificadores multi-etiqueta utilizando algoritmo MLKNN y diferentes n-gramas. . . . .	86
5.21. Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados utilizando los métodos de transformación relevancia binaria, cadenas de clasificadores, conjunto potencia y rakel-d con algoritmos máquinas de soporte de vectores (izquierda) y bosques aleatorios(derecha) . . . . .	87

5.22. Segundos para el entrenamiento de clasificadores multi-etiqueta utilizando diferentes métodos de transformación. . . . . 88

# Índice de tablas

3.1. Acontecimientos que modificaron nombre de la institución . . . . .	42
3.2. Distribución de estudiantes por nivel educativo . . . . .	42
3.3. Egresados por nivel educativo . . . . .	45
3.4. Distribución docentes por tipo de contratación . . . . .	45
3.5. Egresados por programa educativo . . . . .	49
4.1. Protocolo de etiquetado de comentarios al estudiante. . . . .	55
4.2. Frecuencias de caracteres y palabras por comentario al estudiante . . .	56
4.3. Frecuencias de comentarios al estudiante por niveles tarea, proceso, re- gulación, elogios y otros . . . . .	58
4.4. Cantidad de palabras únicas en conjunto de entrenamiento después de aplicar etapas de preprocesamiento . . . . .	61
4.5. Distribución de retroalimentaciones en conjuntos multi-etiqueta de en- trenamiento y prueba . . . . .	61
4.6. Características de los conjuntos de datos multi-etiqueta . . . . .	61
4.7. Hiper-parámetros y espacio de búsqueda. . . . .	62
4.8. Clasificadores multi-etiqueta entrenados en primera etapa . . . . .	63
4.9. Clasificadores multi-etiqueta entrenados en etapa 2 . . . . .	64
5.1. Métricas de desempeño de clasificadores utilizando un-gram . . . . .	68
5.2. Métricas de desempeño de clasificadores utilizando dos-gram . . . . .	69
5.3. Métricas de desempeño de clasificadores utilizando tres-gram . . . . .	70
5.4. Métricas de desempeño de clasificadores utilizando un-dos-tres-gram . .	71
5.5. Métricas de desempeño de clasificadores utilizando un-dos-gram . . . .	72
5.6. Métricas de desempeño de clasificadores utilizando dos-tres-gram . . . .	73
5.7. Parámetros por defecto y afinados . . . . .	89
5.8. Desempeño de los clasificadores multi-etiqueta utilizando hiper-paráme- tros por defecto y afinados . . . . .	89

---

# Capítulo 1

## Introducción

### 1.1. Motivación

Las industrias de los sistemas gestores de aprendizaje y de la educación en línea son las de mayor crecimiento actualmente (MarketandMarket, 2020) . En México, los cursos en línea también han mostrado un incremento como alternativa de estudio (Asociación de Internet MX, 2019) . En los cursos en línea, los comentarios que proporciona un docente a las actividades que entrega un estudiante son esenciales para evidenciar y cerciorarse que el proceso de aprendizaje está tomando lugar (García, 2014).

Los comentarios que proporciona un docente a las actividades que entrega un estudiante también llamados retroalimentaciones se refiere a la información que recibe alguien sobre su desempeño o nivel de entendimiento al realizar una tarea. En la literatura, se encuentran definiciones como la de Hattie y Timperley (2007 p. 81) que definen la retroalimentación como “información provista por un agente sobre aspectos del desempeño o entendimiento de alguien”; García (2014 p. 70), por su parte la define como “un producto resultante de la revisión y el análisis por parte del profesor a la actividad, aportación o proyecto que ha enviado el o los estudiantes”; Vives y Varela (2013 p. 113) la describen como una habilidad que desarrolla el docente al compartir información específica con el estudiante sobre su desempeño para lograr que el estudiante alcance su máximo potencial de aprendizaje.

La retroalimentación es un elemento importante en el proceso de aprendizaje de los estudiantes (Wisniewski *et al.*, 2020; Cavalcanti *et al.*, 2020; Aguerrebere *et al.*, 2018; Fui y Lían, 2018; García, 2014; Hattie y Timperley, 2007). Es parte central de la evaluación formativa (Quesada, 2006). Permite al estudiante identificar lo que le falta para lograr el éxito de acuerdo con lo que se espera de él (García, 2014). Ayuda a los estudiantes a evaluar su progreso de aprendizaje (Cavalcanti *et al.*, 2020). Empodera a los estudiantes a identificar y solventar sus concepciones equivocadas con lo cual mejora su desempeño (Fui y Lían, 2018). Provee al estudiante de reflexión sobre sus actos y consecuencias permitiéndole llegar a metas y objetivos (Vives y Varela, 2013). Se han identificado que la retroalimentación escasa puede causar problemas en el aprendizaje de estudiantes e incluso propiciar el abandono escolar (Wisniewski *et al.*, 2020).

---

En la educación en línea, donde los docentes y estudiantes están separados físicamente en tiempo y espacio, la retroalimentación se vuelve más crítica para la construcción del conocimiento y éxito académico (Cavalcanti *et al.*, 2020). La retroalimentación en la educación en línea pueden ser la principal o única forma de comunicación entre el docente y el estudiante, lo que la vuelve una práctica esencial ya que ayuda al andamiaje del aprendizaje (Hernández, 2007).

El incremento de estudiantes que se matricula en cursos de educación en línea magnifica los retos de los docentes para proveer retroalimentación informativa y de alta calidad que ayude a los estudiantes en su autorregulación (Cavalcanti *et al.*, 2020). Lo anterior, hace necesario la construcción de herramientas que permitan analizar automáticamente las retroalimentaciones para identificar áreas de oportunidad que permitan mejorar el desempeño docente.

Las investigaciones sobre la retroalimentación se centran en tratar de evaluar cuándo es efectiva, para determinar si se puede mejorar y cómo hacerlo (Van der Kleij *et al.*, 2015). Una retroalimentación efectiva es aquella que permite cerrar la brecha entre el desempeño actual y el esperado en el estudiante (Uribe y Baugman, 2017; Hattie y Timperley, 2007). Para ser efectiva, debe aportar al estudiante información específica, detallada, individualizada, en tiempo, enfocada a un nuevo aprendizaje, detectando áreas de oportunidad en las que se puede mejorar (García, 2014).

Existen modelos que facilitan la identificación de características que permiten analizar si una retroalimentación es efectiva o no. Uribe y Vaughan (2017), proponen distinguirlas entre retroalimentaciones correctivas, epistémicas, sugestivas y epistémica + sugestiva. Shute (2008), por su parte propone distinguirlas entre retroalimentaciones enfocadas al resultado, respuestas correctas y elaboradas. Hattie y Timperley (2007) plantean que se pueden ubicar en cuatro niveles: tarea, proceso, regulación y elogios.

El modelo de Hattie y Timperley (2007), establece que la retroalimentación puede ser diferenciada por la información que contiene en: nivel de tarea, nivel de proceso, nivel de regulación y nivel de elogios. La retroalimentación a nivel tarea responde a la pregunta ¿Cuán bien ha sido completada o entendida la tarea? ¿El resultado de la tarea es correcto o incorrecto? La retroalimentación a nivel proceso responde a la pregunta ¿Qué se necesita hacer para entender y ejecutar bien la tarea? La retroalimentación a nivel regulación responde a la pregunta ¿Qué puede hacerse para administrar, guiar y monitorear la forma de trabajo en la realización de la tarea? La retroalimentación a nivel elogios se centra en características personales del destinatario que generalmente son elogios. A partir de esta diferenciación, Hattie y Timperley (2007), señalan que las retroalimentaciones más efectivas son las ubicadas en el nivel proceso y nivel regulación, que las retroalimentaciones en nivel tarea solo son efectivas si se combinan con las de nivel proceso o de nivel regulación y que las de nivel elogios no son significativas para el aprendizaje.

Los niveles del modelo de Hattie y Timperley (2007) son una manera viable de analizar la retroalimentación (Brooks, *et al.*, 2019; Gan y Hattie 2014). Se han utilizado para el desarrollo de aplicaciones que proporcionan retroalimentación a un nivel específico (Pardo *et al.*, 2019), análisis de retroalimentaciones (Brooks *et al.*, 2019; Aj-

---

jawi y Boud, 2017; Harris *et al.*, 2015) y como estrategia para mejorar el desempeño de docentes (Ramírez y Valdés, 2019).

En general, los trabajos que hacen análisis de retroalimentaciones (Pardo, *et al.*, 2019; Brooks, *et al.*, 2019; Ajjawi y Boud 2017; Van der Kleig *et al.*, 2015; Harris *et al.*, 2015; Gan y Hattie 2014) se enfocan a la retroalimentación que se genera en cursos presenciales. Siguen una metodología de cuatro pasos (1) recopilación de retroalimentaciones, (2) diseño de protocolo de clasificación de retroalimentaciones, (3) clasificación manual de retroalimentaciones y (4) análisis de los resultados del etiquetado manual de retroalimentaciones. La etapa de etiquetado manual de las retroalimentaciones requiere de un costo y un nivel de conocimiento elevado, además de que se da de manera lenta y laboriosa.

Debido a la gran cantidad de retroalimentaciones que se pueden generar en un curso en línea han surgido propuestas para analizar automáticamente retroalimentaciones basados en buenas prácticas de lo que es un comentario efectivo (Aguerreberre *et al.*, 2018; Kwon *et al.*, 2019; Cavalcanti, 2020). Los trabajos extraen características de los comentarios y los pasan a algoritmos de aprendizaje de máquina para entrenar clasificadores que asignan a clases predefinidas retroalimentaciones escritas. Sin embargo, estos trabajos utilizan técnicas de aprendizaje de máquina convencionales, que se restringen a problemas con clases mutuamente excluyentes razón por la cual en este trabajo se considera el uso de estrategias de clasificación multi-etiqueta.

## 1.2. Justificación

Considerando el modelo de retroalimentación de Hattie y Timperley que establece que los comentarios que genera un docente a las actividades enviadas por estudiantes en cursos en línea pueden categorizarse en varios niveles simultáneamente, surge la necesidad de contar con herramientas que permitan analizar automáticamente los comentarios ya que hacerlo de manera manual requiere de costos y nivel de conocimiento elevados además de que se da de manera lenta y laboriosa.

Actualmente no existe alguna herramienta en la institución donde se realizó el estudio para el análisis de comentarios al estudiante en cursos en línea, además de que de lo revisado en las referencias documentales se detecta que existe poca información sobre metodologías para el analizar automáticamente comentarios generados por docentes en cursos en línea escritos en lenguaje español. Los trabajos existentes utilizan técnicas de aprendizaje de máquina convencionales, que se restringen a problemas con clases mutuamente excluyentes. En este trabajo se propone utilizar estrategias de clasificación multi-etiqueta.

La investigación planteada propone una metodología para analizar automáticamente las retroalimentaciones que se generan en cursos en línea y apoyar la toma de decisiones sobre el desempeño docente. El análisis de automático de retroalimentaciones ayudará a tener un mejor entendimiento de la retroalimentación y su impacto en el desempeño de estudiantes en cursos en línea que permitirá detectar áreas de oportunidad en la

---

calidad de comentarios que proporcionan los docentes a la revisión de actividades para asesorarlos y mejorar su práctica docente.

Los resultados, permiten hacer una comparación de los enfoques de clasificación multi-etiqueta presentes en el estado del arte para la categorización de retroalimentaciones. Se proponen estrategias para el preprocesamiento de las retroalimentaciones, selección de los algoritmos de clasificación, afinación de hiperparámetros y evaluación de los clasificadores multi-etiqueta.

### 1.3. Objetivos

**General:**

Proponer una metodología para el análisis automático de los comentarios resultantes de la revisión y el análisis del docente a las actividades que envía el estudiante en cursos en línea de la licenciatura en derecho de la plataforma Blackboard en los niveles tarea, proceso, regulación, elogios y otros, categorizando los resultados mediante técnicas de clasificación multi-etiqueta para estimar y evaluar la toma de decisiones respecto al desempeño docente.

**Específicos:**

- Recopilar e integrar los comentarios a los estudiantes del sistema gestor de aprendizaje Blackboard.
- Crear un corpus de comentarios a los estudiantes clasificados manualmente por expertos conforme al modelo de retroalimentación de Hattie y Timperley.
- Comparar los enfoques de clasificación multi-etiqueta transformación y adaptación para determinar cuál obtiene mejores desempeños.
- Examinar los algoritmos de clasificación y configuraciones previas que permiten obtener el mejor desempeño.
- Determinar las características a extraer de los comentarios que permiten obtener un mejor desempeño.

### 1.4. Preguntas de investigación

**General:**

- ¿Cómo analizar automáticamente los comentarios resultantes de la revisión y el análisis del docente a las actividades que envía el estudiante en cursos en línea de la plataforma Blackboard en los niveles tarea, proceso, regulación, elogios y otros, categorizando los resultados mediante técnicas de clasificación multi-etiqueta para estimar y evaluar la toma de decisiones respecto al desempeño docente?

**Particulares:**

- 
- ¿Qué tareas de integración y limpieza debe aplicarse a los comentarios previo al entrenamiento de clasificadores?
  - ¿Cómo evaluar el desempeño de clasificadores multi-etiqueta?
  - ¿Qué enfoque de clasificación multi-etiqueta permite mejor desempeño de los clasificadores multi-etiqueta?
  - ¿Qué algoritmos de clasificación permiten un mejor desempeño y qué hiperparámetros necesitan?
  - ¿Qué características de los comentarios al estudiante se utilizarán para el entrenamiento del clasificador multi-etiqueta?

## 1.5. Hipótesis

Las técnicas de clasificación multi-etiqueta sirven para analizar automáticamente los comentarios resultantes de la revisión y el análisis del docente a las actividades que envía el estudiante en cursos en línea de la licenciatura en derecho de la plataforma Blackboard en los niveles tarea, proceso, regulación, elogios y otros.

## 1.6. Propuesta de solución

Para la solución de la problemática se propone el uso de la clasificación multi-etiqueta que es una tarea predictiva que busca aprender de retroalimentaciones etiquetadas manualmente por expertos, para después ser capaz de predecir el nivel de nuevas retroalimentaciones.

En la literatura (Blanco *et al.*, 2020; Cabrera *et al.*, 2020; Al-Salemi *et al.*, 2019; Herrera *et al.*, 2016), la clasificación multi-etiqueta se aborda desde dos enfoques que son: transformación y adaptación. En este trabajo se hace una comparación de los enfoques para determinar el que mejor clasifica automáticamente las retroalimentaciones conforme los niveles propuestos de Hattie y Timperley. El enfoque de transformación utiliza algoritmos de clasificación existentes transformando los datos para que puedan ser procesados por dichos algoritmos (Herrera, *et al.*, 2016). Los algoritmos de clasificación utilizados en este trabajo fueron: máquinas de soporte de vectores, árboles aleatorios, regresión logística y Naive Bayes cada uno con cuatro técnicas de transformación: (1) relevancia binaria, que genera un clasificador binario por etiqueta, (2) conjunto potencia (Label Powerset), que genera un clasificador multi-clase al considerar cada subconjunto diferente de etiquetas como una sola etiqueta, (3) cadenas de clasificadores que genera un clasificador binario por etiqueta pero que a diferencia de relevancia binaria incorpora como entrada la salida de los clasificadores previos y (4) raketD que genera un ensamble de conjuntos potencia con subconjuntos pequeños de etiquetas (Herrera, *et al.*, 2016).

---

El enfoque de adaptación también utiliza algoritmos de clasificación conocidos, pero, que han sido adaptados para ser capaces de trabajar con datos multi-etiqueta sin transformar los datos (Herrera, *et al.*, 2016). Se utilizan las adaptaciones de los algoritmos vecinos más cercanos (Zhang y Zhou *et al.*, 2007) y máquinas de soporte de vectores (Chen *et al.*, 2016).

En esta propuesta de clasificación multi-etiqueta automática de retroalimentaciones conforme los niveles propuestos por Hattie y Timperley, se recopilaron 11013 retroalimentaciones del sistema gestor de aprendizaje Blackboard que corresponden a 121 cursos en línea de la licenciatura en derecho ofertados en el segundo semestre del año 2018 de una universidad pública de México. Las retroalimentaciones recopiladas fueron clasificadas manualmente por diseñadores instruccionales de la universidad donde se realizó la investigación, los cuales son responsables del diseño de materiales educativos utilizados en las licenciaturas en modalidad en línea además de contar con experiencia como docentes en línea a nivel licenciatura. Para la clasificación manual de los comentarios al estudiante los diseñadores instruccionales siguieron el modelo de retroalimentación de Hattie y Timperley, teniendo como resultado un conjunto de datos multi-etiqueta. El conjunto de datos multi-etiqueta se particionó en dos subconjuntos, uno para entrenamiento y otro para prueba. El conjunto de entrenamiento fue utilizado por los algoritmos de los enfoques de transformación y adaptación para inferir conocimiento que permitiera clasificar nuevas retroalimentaciones automáticamente, teniendo como resultado un clasificador multi-etiqueta por cada algoritmo. En una primera etapa de clasificación se entrenaron 108 clasificadores multi-etiqueta, 96 con el enfoque de transformación y 12 con el enfoque de adaptación para determinar el enfoque de clasificación multi-etiqueta y algoritmos que mejor desempeñaban la tarea de ubicar automáticamente las retroalimentaciones conforme los niveles propuestos por Hattie y Timperley. Una vez detectados los algoritmos que mejor desempeño tuvieron en la realización de la tarea se pasó por una etapa de afinación de hiper-parámetros y se analizó si el desempeño se mejoró. Para esta segunda etapa, se entrenaron 14 clasificadores multi-etiqueta, 12 con el enfoque de transformación utilizando los algoritmos de máquinas de soporte de vectores y bosques aleatorios con los métodos de transformación de relevancia binaria, cadenas de clasificadores y conjunto potencia. Los dos restantes se entrenaron con el enfoque de adaptación, utilizando el algoritmo de vecinos más cercanos multi-etiqueta. Para cada algoritmo, se comparan las métricas de desempeño de clasificación multi-etiqueta utilizando los valores por defecto de los hiper-parámetros y los encontrados en la etapa de afinación de hiper-parámetros.

El conjunto de prueba se utilizó para evaluar el desempeño de los clasificadores multi-etiqueta. Se proporcionó a cada clasificador las retroalimentaciones del conjunto de prueba sin las clasificaciones realizadas por los diseñadores instruccionales para que clasificaran automáticamente cada una conforme los niveles propuestos de Hattie y Timperley. Después, se compararon las clasificaciones automáticas realizadas por los clasificadores multi-etiqueta con las clasificadas por los diseñadores instruccionales y se calcularon las métricas tradicionalmente usadas en la literatura (Herrera, *et al.*, 2016) para la clasificación multi-etiqueta.

---

## 1.7. Organización del documento

Este trabajo se organiza en 6 capítulos, los cuales se distribuyen de la siguiente manera: en el capítulo 2 se incluye los conceptos de educación en línea y retroalimentación referente a los comentarios al estudiante, necesarios para entender el contenido del trabajo, posteriormente se introducen los conceptos de descubrimiento de conocimiento en bases de datos, minería de textos, clasificación y clasificación multi-etiqueta con una profunda revisión que ilustra los principales enfoques que se encuentran en el estado del arte para la clasificación multi-etiqueta. El capítulo 3 aborda el marco contextual que describe la institución donde se llevó a cabo la investigación, características y forma en que abordan la educación en línea. En el capítulo 4 se presenta la metodología propuesta y se describen los experimentos realizados. En el capítulo 5 se analizan y discuten los resultados obtenidos de los experimentos descritos en el capítulo anterior. En el capítulo, 6 se presentan las conclusiones y una propuesta para un trabajo futuro. Más adelante, aparecen las referencias documentales. Finalmente se muestran los diferentes anexos del documento.

---

# Capítulo 2

## Marco Teórico

En este capítulo se presenta el marco teórico del trabajo de investigación, se da un panorama de la educación en línea, la retroalimentación, su importancia en proceso de aprendizaje de los estudiantes, el modelo de retroalimentación de Hattie y Timperley, clasificación, clasificación multi-etiqueta y afinación de hiper-parámetros.

### 2.1. Educación en línea

Cuando se habla de enseñanza, se pueden encontrar en la literatura dos grandes vertientes: la enseñanza presencial y la enseñanza no presencial, cada una con sus propias características lineamientos y objetivos.

Rivera, Alonso y Sancho (2017) definen a la enseñanza presencial como el modelo tradicional que se imparte en un aula física, que permite una interacción y comunicación cara a cara entre docente y estudiante. Se resalta que en este tipo de enseñanza se pueden utilizar medios tecnológicos como apoyo didáctico.

Por otro lado, la enseñanza no presencial es un modelo de formación esencialmente asincrónico, donde no hay que acudir con periodicidad para el desarrollo de las clases a un determinado centro físico, se caracteriza en que el estudiante recibe material para cursar asignaturas, envía tareas para su evaluación y puede consultar al profesor a través de herramientas tecnológicas como el teléfono, fax, correo e Internet (Rivera, Alonso y Sancho, 2017).

Respecto a la enseñanza no presencial se identifican cuatro tipos de ambientes formativos (Bates, 2005; García 2009 citado en Rivera, Alonso y Sancho, 2017 ) que son la enseñanza a distancia, enseñanza a distancia con entornos virtuales, enseñanza en entornos duales o bimodales y enseñanza en entornos virtuales. A continuación se describe cada uno:

- **Enseñanza a distancia, sin entornos virtuales.** Se trata, del enfoque tradicional de la educación a distancia que se da frecuente en zonas de bajo desarrollo y escaso acceso de la población a redes telemáticas.

- 
- **Enseñanza a distancia con entornos virtuales.** Al igual que en la enseñanza a distancia desarrollan un modelo tradicional e introducen el uso de Internet en su oferta educativa.
  - **Enseñanza en entornos duales o bimodales.** Se trata de formación mixta, también denominadas como semipresencial, blended learning, etc. Son resultante de la coexistencia de los dos modelos clásicos: presenciales y a distancia en una misma institución.
  - **Enseñanza en entornos virtuales (e-learning).** Se trata de la enseñanza a distancia que utiliza como medio de transmisión fundamentalmente Internet mediante entornos virtuales.

Los términos, aprendizaje a distancia, educación a distancia, aprendizaje distribuido, e-learning y aprendizaje en línea se pueden utilizar sin mayores diferencias. García (2020) destaca que entre las diferentes denominaciones que se han asignado a propuestas educativas de carácter no presencial se encuentran: estudio por correspondencia, instrucción a lo lejos, estudio en casa, autoestudio guiado, estudiar sin dejar de producir, estudio independiente, forma industrializada de instrucción, estudios externos, teleformación, educación abierta educación/enseñanza o aprendizaje virtual, educación en línea, educación basada en web, e-Learning, blended learning (aprendizaje mixto), enseñanza/aprendizaje o educación distribuida, siendo el de mayor consenso el de educación a distancia.

Considerando los rasgos sustanciales de la educación a distancia como lo es la separación física del docente y alumno, el estudio independiente de los alumnos, el soporte de una organización/ institución y la comunicación e interacción síncrona o asíncrona entre docente y estudiantes y estos últimos entre sí, García (2020), define la educación a distancia como: ".aquella que se basa en un diálogo didáctico mediado entre docentes de una institución y los estudiantes que, ubicados en espacio diferente al de aquellos, pueden aprender de forma independiente o grupal".

Otra definición de la educación en línea es la propuesta por García-Peñalvo y Seoane-Parto (2015) que indican es un proceso formativo de naturaleza intencional o no intencional, orientado a la adquisición de una serie de competencias y destrezas en un contexto social, que se desarrolla en un ecosistema tecnológico en el que interactúan diferentes perfiles de usuarios que comparten contenidos, actividades y experiencias y que, en situaciones de aprendizaje formal, debe ser tutelado por docentes cuya actividad contribuya a garantizar la calidad de todos los factores involucrados.

### 2.1.1. Evolución de la educación en línea

En la literatura se pueden encontrar diferentes etapas sobre la evolución de la educación en línea. Salvat (2018), señala que la educación en línea nació en los años 90 como una modalidad de formación asociada a la educación a distancia y al uso del Internet

---

en la educación superior y la formación empresarial. Señala una diferencia entre los modelos tradicionales de educación a distancia y la educación en línea ya que la educación a distancia puede o no utilizar tecnología, siendo lo importante el garantizar el estudio independiente sin necesidad de que haya una intervención continua del docente. Por su parte en el caso de la educación en línea, señala que comparte la no presencialidad del modelo, pero el énfasis se produce en la utilización de Internet como sistema de acceso a los contenidos y a las actividades de formación, además de que la interacción y la comunicación son una parte fundamental en este tipo de formación.

La evolución de la educación en línea está relacionada con las transformaciones tecnológicas y los cambios relacionados con los espacios y tiempos para la formación y el aprendizaje. Al inicio, la educación en línea estaba vinculada al uso de los ordenadores personales y actualmente está sostenida por los sistemas distribuidos que han favorecido las redes de aprendizaje y el acceso a las plataformas desde cualquier dispositivo. Así el uso de dispositivos móviles se ha integrado en los diseños educativos y las actividades de aprendizaje en línea.

Downes (2012 citado en Salvat, 2018 ), describe la evolución de la educación en línea a partir de una serie de generaciones. En primer lugar, señala la existencia de una generación cero, caracterizada por el diseño y la publicación de los recursos multimedia en línea. En este primer momento lo importante es usar los ordenadores para transmitir contenidos instructivos y realizar actividades basadas en pruebas y cuestionarios. La generación uno de la educación en línea se inicia a partir del Internet y el uso del correo electrónico que facilita la comunicación virtual. La segunda generación tiene lugar a principios de los 90 y se caracteriza por la aplicación de los juegos de ordenador para el aprendizaje en línea. La tercera generación se da por el desarrollo de los sistemas gestores de aprendizaje (LMS) que permiten conectar los contenidos de la generación cero con las plataformas. La generación cuatro se fundamenta en el uso de la web 2.0 donde una de las características más significativas de esta fase es la interacción social entre los estudiantes, cambiando la naturaleza de la red subyacente donde los nodos son ahora personas en lugar de ordenadores. Esta orientación social también se ve impulsada gracias al uso de los dispositivos móviles. La generación cinco se caracteriza por la computación en la nube y el contenido abierto. Por último, la generación seis se caracteriza por los cursos abiertos masivos en línea.

Otra propuesta de ver la evolución de la educación en línea es la de García y Seoane (2015), que señalan tres generaciones. La primera se caracteriza por la aparición de las plataformas de aprendizaje en línea y la creación de las aulas y campus virtuales. Estos sistemas evolucionan y en la segunda generación, las plataformas permiten apoyar la socialización, la movilidad y la interoperabilidad de datos. Esta segunda generación subraya el factor humano, la interacción entre compañeros y la comunicación entre profesores y estudiantes son elementos esenciales para una educación en línea de calidad que pretende ir más allá de un simple proceso de publicación de contenidos. El desarrollo de las Web 2.0, las tecnologías móviles y los recursos abiertos para el aprendizaje también, son factores significativos de esta generación. La tercera y última generación se caracteriza porque la educación en línea deja de estar asociado de forma exclusiva a

---

las plataformas de aprendizaje en línea, los sistemas gestores de aprendizaje dejan de ser un componente único y monolítico y se convierten en un componente más de un ecosistema tecnológico orientado hacia el proceso de aprendizaje.

### **2.1.2. Características de la educación en línea**

Entre las características de la educación en línea se encuentran la desaparición de las barreras espacio-temporales al estar los contenidos habilitados cualquier día en cualquier hora, formación flexible a través de la diversidad de métodos y recursos empleados, enseñanza-aprendizaje centrado en el estudiante, el docente se transforma de un emisor de contenidos a un tutor que orienta, guía, ayuda y facilita procesos formativos, comunicación constante entre los participantes gracias a herramientas que incorporan los sistemas gestores de aprendizaje (Velazco *et al.*, 2017).

Cabero (2016) señala nueve características de la educación en línea que son: (1) separación espacial y temporal entre docente y discente, (2) formación mediada y, en consecuencia, apoyada en diferentes tecnologías que condicionan y matizan la relación que docente y discente establecen con los contenidos, (3) comunicación mediada entre el docente y discente, (4) posibilita tanto la comunicación sincrónica (chat, videoconferencia) como asincrónica (foros, correo electrónico, etc.), (5) por lo general, los alumnos son de más edad que los del sistema presencial, (6) existencia de una institución que organice la estructura educativa y la certificación académica, (7) incorporación en la acción educativa de diferentes personas que garanticen su calidad: profesores, alumnos, técnicos de producción de materiales, distribuidores de materiales, (8) formación fuertemente tutorizada, (9) comunicación bidireccional, multicódigo y multipersonal.

### **2.1.3. Ventajas y desventajas de la educación en línea**

Diversos autores han señalado ventajas y desventajas de la modalidad en línea, a continuación se describen algunas encontradas en las referencias documentales.

En Velazco *et al.*, (2017), destacan como ventaja la reducción de costos al permitir reducir o eliminar gastos de traslado, alojamiento y material didáctico. También, señalan la rapidez y agilidad debido al avance de las comunicaciones. De igual forma, el acceso justo-a-tiempo que permite a los estudiantes acceder al contenidos desde cualquier conexión a Internet cuando les surge la necesidad. Además, resaltan la flexibilidad al no requerir que un grupo de personas coincidan en tiempo y espacio. El aprendizaje permanente, interacción y colaboración con otros institutos y el incremento en el acceso a plataformas de enseñanza también son una ventaja.

Por su parte, Cabero (2006) señala que las ventajas de la educación en línea son: (1) Pone a disposición de los alumnos un amplio volumen de información. (2) Facilita la actualización de la información y de los contenidos. (3) Flexibiliza la información, independientemente del espacio y el tiempo en cual se encuentre el profesor y el estudiante. (4) Permite la deslocalización del conocimiento. (5) Facilita la autonomía del estudiante. (6) Propicia una formación justo en tiempo y justo para mi. (7) Ofrece

---

diferentes herramientas de comunicación sincrónica y asincrónica para los estudiantes y para los profesores. (8) Favorece la formación multimedia. (9) Facilita una formación grupal y colaborativa. (10) Favorece la interactividad en diferentes ámbitos: con la información, con el profesor y entre los alumnos. (11) Facilita el uso de los materiales, los objetos de aprendizaje, en diferentes cursos. (12) Permite que en los servidores pueda quedar registrada la actividad realizada por los estudiantes. (13) Ahorra costos y desplazamientos.

Respecto a las desventajas Velazco *et al.*, (2017) señalan la dependencia de medios digitales, la inversión en la infraestructura para el aprendizaje electrónico, altas tasas de deserción, difícil adaptación a la metodología de estudio, sobre carga de información, necesidad de autodisciplina y motivación.

También, Cabero (2006), indica como desventajas las siguientes: requiere más inversión de tiempo por parte del profesor, requiere de competencias tecnológicas por parte del profesor y de los estudiantes, necesidad de que los estudiantes tengan habilidades para el aprendizaje autónomo, riesgo de disminuir la calidad de la formación si no se da una interacción adecuada profesor-alumno, requiere más trabajo que el convencional, supone la baja calidad de muchos cursos y contenidos actuales, se encuentra con la resistencia al cambio del sistema tradicional, impone soledad y ausencia de referencias físicas, depende de una conexión a Internet, tiene profesorado poco formado, supone problema de seguridad y además de autenticación por parte del estudiante y existe una brecha digital.

#### **2.1.4. Calidad de la educación en línea**

En la literatura, se encuentran diferentes variables que impactan en la calidad de la educación en línea. Siemens (2014 citado en Salvat, 2018), señala la importancia del desarrollo de cursos bien diseñados que incorporen contenidos interactivos y atractivos, actividades que permitan la colaboración estructurada entre pares y plazos flexibles que permitan a los estudiantes organizar sus tiempos.

Darabit (2013 citado en Salvat, 2018) considera que el mayor impacto en el rendimiento del estudiante se obtiene a través de diseños pedagógicos ricos que incluyan la presencia de los docentes, la interacción con los estudiantes, la colaboración del estudiante y el seguimiento continuo. Destaca que el éxito en la educación en línea depende en gran medida de la capacidad del estudiante para dirigir y gestionar su propio proceso de aprendizaje, estableciendo objetivos y estrategias adecuadas para alcanzar sus objetivos.

Por su parte Cabero (2006), señala que las variables críticas a las que se les debe prestar atención para realizar una acción educativa de calidad en línea son: contenidos, papel del profesor, papel del alumno, e-actividad, aspectos organizativos, modelos de evaluación, herramientas de comunicación, estrategias didácticas y comunidad virtual.

Considerar las variables señaladas previamente permitirá mejorar la calidad de la educación en línea.

---

### 2.1.5. Sistemas gestores de aprendizaje

Uno de los elementos clave en la educación en línea son los sistemas utilizados para promover la interacción entre contenidos, estudiantes y docentes los cuales son denominados sistemas gestores de aprendizaje.

Un sistema gestor de aprendizaje es una aplicación de software basada en la web diseñada para gestionar contenidos de aprendizaje, interacción de estudiantes, herramientas de evaluación, reportes de aprendizaje y actividades de estudiante (Kasim y Khalid, 2016).

Los sistemas gestores de aprendizaje son herramientas que se pueden emplear para administrar las tareas de los estudiantes, en los que la comunicación es personalizada y efectiva, y con lo que se puede mejorar el seguimiento al proceso formativo (Ardila y Ruiz, 2015 ).

Los sistemas gestores de aprendizaje proveen de funciones para la administración de usuarios, gestión de recursos, control de calificaciones y seguimiento del proceso de aprendizaje, estas han sido clasificadas por Jaramillo y Cuasquer (2013) de la siguiente manera:

**Orientadas al aprendizaje.** Incluye herramientas como foros de discusión, intercambio de archivos, soporte multi-formato, comunicación síncrona y asíncrona, servicios multimedia, blogs, noticias, pizarra y wikis.

**Orientadas a la productividad.** Considera herramientas como: anotaciones personales, calendario, ayuda para utilizar la plataforma, sincronización de archivos, trabajo fuera de línea, control de publicaciones, páginas caducadas, enlaces rotos y aviso de actualización de páginas.

**Orientadas a la implicación de los estudiantes.** Herramientas que permiten la creación de grupos de trabajo, autoevaluaciones y presentación del perfil del estudiante.

**Orientadas al soporte.** Herramientas que permiten la autenticación de usuario, asignación de privilegios en función del rol del usuario, registro de estudiantes y auditoría.

**Orientadas a la publicación de cursos y contenidos.** Herramientas para incorporar pruebas y resultados automatizados, administración del curso, apoyo al creador de cursos, calificación en línea y seguimiento del estudiante.

**Orientadas al diseño de planes de trabajo.** Herramientas que facilitan la compartimento de contenidos y reutilización de los mismos, administración del currículo y personalización del entorno.

En la clasificación de funciones de los sistemas gestores de aprendizaje se resalta que estas proveen de herramientas que facilitan la interacción y comunicación entre estudiante-contenido, estudiante-estudiante y estudiante-docente tales como foros, blog, wikis, intercambio de archivos, pruebas, autoevaluaciones entre otras. Además, facilitan al estudiante la organización de actividades a través de herramientas de agenda y permiten llevar un control de usuarios a través de cuentas de usuario.

En la actualidad existen un gran número de sistemas gestores de aprendizaje que dan respuesta a necesidades específicas de aprendizaje que han sido desarrolladas por

---

organizaciones privadas, universidades o centros de investigación. El portal Capterra (<https://www.capterra.com/learning-management-system-software/>), que permite obtener comparativas de software sobre sus niveles de popularidad, costo y usabilidad, lista 410 sistemas gestores de aprendizaje diferentes sin considerar los desarrollos de universidades que han decidido no hacerlas públicas.

En la literatura se pueden encontrar varias categorizaciones de los sistemas gestores de aprendizaje, por ejemplo, el portal *Web hace* hace una agrupación basado en las posibilidades de pago destacando los sistemas gestores de aprendizaje de código abierto y comerciales de las que se desprenden las de pago por uso en la nube y de pago por licencia. *Clarenc et al.*, (2013) señalan tres categorías para clasificar los sistemas gestores de aprendizaje que son: comerciales, de software libre y en la nube.

Los sistemas gestores de aprendizaje de uso libre son desarrollados por instituciones educativas u organizaciones sin fines de lucro. Algunas de ellas son de tipo open source que como explica *González* (2011), los dueños del programa permiten utilizarlos como mejor le parezca al usuario, poder estudiar cómo funciona, modificarlo, redistribuirlos y distribuir copias modificadas, lo que permite cierto nivel de autonomía para la manipulación del software. Entre estas plataformas se encuentran:

**Moodle** . Descrito como un sistema gestor de aprendizaje único, robusto y seguro para crear ambientes de aprendizaje personalizados. Construido por el proyecto Moodle que está dirigido y coordinado por el cuartel general Moodle que está soportada financieramente por una red mundial de cerca de 80 compañías de servicio.

**Dokeos manager**. Uno de los softwares para e-learning que ofrece la empresa Dokeos, descrito en su documentación como un sistema gestor de aprendizaje completo, flexible y punto de partida para administrar un portal de e-learning. Permite administrar currículos de capacitación, personalizar su interfaz y adaptable para pc, tabletas y teléfonos inteligentes.

**Claroline** . En la documentación oficial se describe como un sistema gestor de aprendizaje que es software de libre descarga para la administración del aprendizaje. Pone a las personas en el centro de la capacitación ofreciendo la posibilidad de crear, compartir, elegir y organizar los elementos de su aprendizaje, pensada con una visión de colaboración, permitiendo la interacción entre usuarios dentro de un contexto de proyecto común.

**Sakai**. Definida en su portal oficial como un software de código abierto para ambientes de colaboración y aprendizaje. Provee de un ambiente flexible y variedad que características para la enseñanza, el aprendizaje, investigación y otras colaboraciones.

Los sistemas gestores de aprendizaje comerciales son desarrollados por compañías privadas, son de uso licenciado o mejor conocidos como software propietario que como explican *Culebro, Gómez y Torres* (citado en *Arriola, Tecuatl y Gozález* 2011) son programas informáticos en el que los usuarios tienen limitadas las posibilidades de usarlo, modificarlo o redistribuirlo (con o sin modificaciones), o cuyo código no esté disponible o el acceso a éste se encuentre restringido. Sus funcionalidades pueden expandirse de acuerdo a necesidades y presupuesto del proyecto. Cuentan con un soporte técnico sólido. Entre estas se encuentran:

---

**Blackboard** . Descrito en su documentación oficial como aplicación e enseñanza, aprendizaje, creación de comunidades y uso compartido de conocimientos en línea. Permite utilizar cualquier teoría o modelo para enseñar a través de un curso en línea, es flexible y se centra en mejorar los logros de los alumnos.

**Sumtotal** Descrita como un sistema gestor de aprendizaje que conecta habilidades y competencias que necesita la gente para realizar mejor su trabajo. Desarrollado en el contexto del logro de proyectos.

**E-ducativa.** Definida en su portal oficial como sistema gestor de aprendizaje que permite dictar cursos de e-learning y gestionar material educativo a través de internet, ideal para escuelas, instituciones, empresas o consultoras que deseen impartir cursos a distancia o apoyar la capacitación presencial.

**Schoology** Sistema gestor de aprendizaje que permite conectarse, comunicarse, compartir fácilmente con pares a través del campus y alrededor del mundo además de crear objetivos alineados a evaluaciones.

Por último los sistemas gestores de aprendizaje de la nube que son servicios que de acuerdo con Agut (2017), permiten al usuario final ejecutar aplicaciones de software y acceder a datos desde cualquier lugar hora, ordenador sin necesidad de instalar, actualizar, o solucionar problemas físicamente en un escritorio o servidor local. Almonte (2016), resalta que sus principales características son: la misma empresa que ofrece el servicio es la que programa y actualiza el sistema gestor de aprendizaje, tienen una amplia oferta, y continuamente aparecen nuevas soluciones, ofrecen un servicio completo, que incluye alojamiento, mantenimiento, copias de seguridad y actualizaciones además de que las mejores suelen ofrecer un “ecosistema” de productos y servicios asociados. Entre este tipo de plataformas se encuentran las siguientes:

**Canvas.** Definido en su portal oficial como un sistema gestor de aprendizaje basado en la nube que permite la enseñanza y aprendizaje más fácil.

**Talensoft LMS** Sistema gestor de aprendizaje que permite desarrollar contenido de formación, mientras promueve la interacción y el intercambio de conocimientos entre sus empleados para, de esta manera, desarrollar sus competencias en cualquier momento, desde cualquier parte.

### 2.1.6. Actividades de aprendizaje en educación en línea

Las actividades de aprendizaje en la educación en línea son aquellas propuestas de trabajo dirigidas a los estudiantes que le ayudan a comprender, analizar, sintetizar y valorar los contenidos propuestos en los diferentes materiales y convertir la información librada en bruto en un conjunto de conocimientos, habilidades y actitudes relativas a la materia trabajada (Almenara y Gray, 2006).

Las actividades de aprendizaje son una de las variables críticas de la educación en línea como se mencionó en secciones anteriores. De acuerdo con Almera y Gray (2006), las actividades de aprendizaje son el elemento que facilita la interconexión entre la enseñanza y el aprendizaje, sus características y funcionalidad serán las mismas que las realizadas en contextos presenciales.

---

Los autores destacan que las funciones que pueden desempeñar las actividades de aprendizaje en los cursos en línea son diversas, e irán desde las que persiguen con ellas objetivos claramente cognitivos, hasta las que únicamente pretendan en su realización la motivación y socialización del estudiante con el resto de compañeros de la acción formativa.

Entre las funciones que deben cubrir las actividades de aprendizaje en cursos en línea se encuentran:

- Clarificación de los contenidos presentados.
- Transferencia de la información a contextos y escenarios diferentes en los cuales fueron presentados.
- Profundización en la materia.
- Adquisición de vocabulario específico.
- Socialización.
- Aplicación de los contenidos a su actividad profesional actual.

Las actividades de aprendizaje y la retroalimentación van de la mano debido a la forma de trabajo, en la que los estudiantes envían sus actividades electrónicamente y suelen recibir retroalimentación de la misma. Así, en cuanto el profesor envía el comentario de retroalimentación en el primer intento de evaluación formativa, se espera que el alumno lo utilice para realizar ajustes en la actividad con el fin de alcanzar las expectativas y los objetivos de aprendizaje (García, 2014).

## 2.2. Retroalimentación

En la literatura se pueden encontrar diversas definiciones sobre retroalimentación. García (2014), la define como un producto resultante de la revisión y el análisis por parte del profesor a la actividad, aportación o proyecto que ha enviado el (los) estudiante(s), destaca que en ese producto se deben reflejar: observaciones, correcciones, aportaciones, reforzamiento y una evaluación cuantitativa basada en una rubrica y la puntual descripción de la actividad. Por su parte Vives y Varela (2013) especifican que se refiere a la habilidad que desarrolla el docente al compartir información específica con el estudiante sobre su desempeño, para lograr que el educando alcance su máximo potencial de aprendizaje según su etapa de formación. Hattie y Timperley (2007 cit en Martínez y Vargas, 2014) la describen como la información que provee un agente como podría ser: un profesor, un compañero de equipo, un libro, uno mismo, sobre el desempeño académico de una actividad de aprendizaje.

En las definiciones se destaca que la retroalimentación se refiere a información explícita dirigida al estudiante en el cual se le informa sobre su desempeño en la realización de una actividad.

---

En la educación en línea, la retroalimentación juega un papel importante debido a su forma de trabajo, en la que los estudiantes envían sus tareas electrónicamente y suelen recibir retroalimentación de la misma. Así, en cuanto el profesor envía un comentario en el primer intento de evaluación formativa, se espera que el alumno lo utilice para realizar ajustes en la actividad con el fin de alcanzar las expectativas y los objetivos de aprendizaje (García, 2014).

En los modelos de educación en línea esta información por lo general se transmite de manera electrónica y queda almacenada en los sistemas gestores de aprendizaje solo para su consulta desaprovechando su contenido.

### **2.2.1. Importancia de la retroalimentación**

La retroalimentación busca enterar a los estudiantes sobre su desempeño en la realización de una actividad destacando qué realizó bien y en qué se debe mejorar. Hattie y Timperley (2007 cit. En Martínez y Vargas, 2014) explican que esta busca que el alumno se dé cuenta de la discrepancia que hay entre lo que comprendió y lo que debió haber comprendido, o cómo se ha desempeñado para cumplir con el objetivo de aprendizaje de cada actividad.

La importancia de la retroalimentación radica en que mejora el desempeño de los estudiantes y el logro de objetivos de aprendizaje. En la literatura se señala que es la parte central de la evaluación formativa por lo que deberá ser constante, clara, oportuna, suficiente y pertinente (Quesada, 2006). Es el medio a través del cual el alumno identifica lo que le hace falta para lograr el éxito de acuerdo a lo que se espera de él (García (2014)).

La retroalimentación, empodera a los estudiantes a identificar y los ayuda solventar sus concepciones equivocadas con las cuales puede mejorar su desempeño ( Fui y Lian, 2018).

Se ha identificado a la retroalimentación como una actividad clave en el proceso de enseñanza-aprendizaje del estudiante, ya que implica darle información que le ayude a cumplir con los objetivos de aprendizaje (Martínez y Vargas, 2014).

También, se ha detectado como un componente esencial en el proceso de enseñanza y aprendizaje que provee al estudiante de reflexión sobre sus actos y consecuencias, le permite llegar a las metas y objetivos (Vives y Varela, 2013).

La importancia de la retroalimentación se sustenta en los principios del constructivismo en el cual se afirma que se aprende a través de la interacción con otros individuos. Vives y Varela (2013), explican al respecto que el conocimiento se considera como un proceso de relación entre el sujeto y el medio, pero el medio entendido como algo social y cultural, no solamente físico. Este principio se desprende de las aportaciones de Lev Vygotsky (zona de desarrollo próximo) quien argumenta que, con el apoyo de una persona más capaz, ya sea de mayor conocimiento y/o experiencia, se puede desarrollar mejor el potencial personal, que sin ayuda o guía que puede quedar sin avances.

La interacción se refiere al intercambio de comunicación entre docente-estudiante y estudiante-estudiante, es un mecanismo esencial para evidenciar y cerciorarse de que el

---

proceso de aprendizaje está tomando lugar en los ambientes de educación en línea. En el caso específico de esta investigación interacción es caracterizada por ser asíncrona y comunicación escrita.

Vives y Varela (2013) destacan nueve puntos sobre la relevancia de la retroalimentación los cuales son: (1) Ofrece introspección, que ayuda a los estudiantes a realizar ajustes en su desempeño para mejorarlo. (2) Favorece la autoevaluación al recibir una crítica constructiva. (3) Aclara las metas y expectativas que se espera sean alcanzadas. (4) Refuerza las acciones y prácticas correctas. (5) Provee las bases para corregir errores. (6) Incrementa la capacidad de incorporar al desempeño la opinión y sugerencias de expertos. (7) Brinda información que no se encuentra en los textos o en los exámenes. (8) Demuestra compromiso e interés del docente en sus aprendices. (9) Promueve una comunicación positiva entre profesores y estudiantes.

Por su lado, Walker (cit. en Vives y Varela, 2013) destaca que la retroalimentación ayuda a clarificar lo que se hizo bien (objetivos, criterios y estándares esperados), facilita la reflexión en el aprendizaje, brinda información de calidad sobre el aprendizaje de los estudiantes, promueve el diálogo entre pares sobre el aprendizaje, promueve la motivación y autoestima, da oportunidades para cerrar la brecha entre desempeño actual y el esperado además de proveer información a los docentes que puede ser usadas para modelar la enseñanza.

Como puede notarse, la retroalimentación no solo es útil para el estudiante sino para el docente pues le permite hacer ajustes a sus procesos de enseñanza. Sobre esto García (2014) menciona que permiten identificar aspectos en los que los estudiantes están fallando permitiendo hacer intervenciones didácticas. En el mismo sentido esta podría considerarse una especie de andamiaje y por ello ser el paso inicial en dirección hacia una sutil estrategia tutorial.

Desde el punto de vista de la educación en línea la retroalimentación juega un papel importante por la forma de trabajo en la que los estudiantes envían sus tareas electrónicamente y normalmente reciben retroalimentación de manera electrónica. En cuanto el docente envía la retroalimentación y quizás una calificación en el primer intento de una evaluación formativa se espera que el estudiante utilizará la retroalimentación para hacer ajustes a la actividad con el objeto alcanzar expectativas y objetivos de aprendizaje.

Al respecto, Hernández (2007) indica que los comentarios a las actividades en la educación en línea pueden ser el principal o única comunicación de aprendizaje entre el tutor y el estudiante. Uribe y Vaughan (2017) indican que en un ambiente en línea la retroalimentación es una práctica docente esencial pues ayuda al andamiaje del aprendizaje y establece una conexión entre la distancia profesor-estudiante. Martínez y Vargas (2014) mencionan que por cada actividad o tarea que el estudiante realice se necesita que su desempeño y/o producción académica sea evaluada y retroalimentada por un profesor tutor quien será su guía en el proceso de aprendizaje.

---

### 2.2.2. Elementos de la retroalimentación

Una retroalimentación efectiva es aquella que aporta al estudiante información específica, detallada, individualizada, en tiempo, enfocada a un nuevo aprendizaje, detectando áreas de oportunidad en las que puede mejorar. Una buena retroalimentación reduce la distancia entre lo que comprendió el estudiante y lo que debió haber comprendido.

De acuerdo con García (2014) para que una retroalimentación esté completa debe incluir tres conceptos: “Feed– Up” que se refiere a qué dirección está tomando el alumno, si se está dando cuenta hacia dónde va, es básicamente no perder de vista el objetivo de la actividad que se está revisando y hacer referencia, en síntesis, de la actividad previa para ligar el conocimiento previo con el actual; el “Feed – Forward” que debe contestar al alumno las interrogantes sobre ¿qué sigue ahora? ¿cómo puede mejorar para la siguiente actividad? Y el “Feed– Back” que le ayuda al alumno a darse cuenta cómo se está desempeñando.

Para que una retroalimentación sea efectiva debe incluir ciertas características. Uribe y Vaughan (2017) indican que la más efectiva no solo refleja el desempeño actual de los estudiantes sino también como mejorar su desempeño futuro y aprendizajes subsecuentes para esto debe clarificar objetivos, criterios o estándares, facilitar la autoevaluación y reflexión, proporcionar información de calidad sobre el aprendizaje de los estudiantes, buscar cerrar la brecha entre desempeño actual y el esperado, proveer información para ayudar a los profesores a mejorar sus prácticas de enseñanza.

Por su parte Sadler (cit. En , Martínez y Vargas, 2014) encuentra tres premisas que los estudiantes necesitan tener en las retroalimentaciones: (1) Deben saber el objetivo de la actividad, (2) recibir retroalimentación con referencia a algún estándar, (3) conocer algunas estrategias para modificar la actividad de manera que se cumpla con las expectativas.

### 2.2.3. Análisis de la retroalimentación

En la literatura se encuentran diferentes propuestas para analizar la efectividad de la retroalimentación. Uribe y Vaughan (2017) proponen para analizarlas clasificarlas en cuatro tipos de retroalimentación (1) correctiva la cual se enfoca a los requerimientos y contenido, (2) epistémica que clarifica o explica a través de indicaciones y cuestionamientos, (3) sugestiva que incluye consejos o ideas para mejorar y (4) epistémica más sugestiva la cual incluye indicaciones, cuestionamientos e ideas para el andamiaje de los aprendizajes de los estudiantes.

Por su parte, Vives y Varela (2013) proponen clasificar la retroalimentación considerando el momento en que surgen que puede ser la Formal (se incorpora como parte de la evaluación formativa, ocurre de manera estructurada y con un plan establecido) y la informal (generalmente es más breve, se administra de acuerdo a las necesidades de profesores y alumnos, en especial cuando el estudiante está demostrando alguna habilidad procedimental o cognitiva).

---

Para evaluar la calidad de las retroalimentaciones se han propuesto modelos que facilitan la identificación de características que permitan analizar si son efectivas o no. Algunos de estos modelos se describen a continuación.

- Propuesta de Uribe y Vaughan (2017) en la que indican que existen cuatro tipos de retroalimentación: correctiva, epistémica, sugestiva y epistémica + sugestiva, destacan que las últimas dos son las que mayor impacto tienen para el aprendizaje.
- Propuesta de Shute (2008), se destacan tres tipos de retroalimentaciones: resultado, respuestas correctas y elaboradas estas últimas descritas con tener un mayor impacto positivo en el aprendizaje de los estudiantes.
- Propuesta de Hattie y Timperley (2007), destaca que una retroalimentación efectiva debe responder a tres cuestionamientos ¿de dónde vengo?, ¿dónde estoy? ¿hacia dónde voy? La cual puede estar en cuatro niveles: tarea, proceso, autorregulación y el yo (elogios). Se destaca que los que producen una mejora en el aprendizaje son los tres primeros.

### 2.3. Modelo de retroalimentación Hattie y Timperley

El modelo seleccionado para este trabajo es el propuesto por Hattie y Timperley al ser de los más citados en la literatura y existir evidencia de la viabilidad de analizar las retroalimentaciones a través de los niveles propuestos.

El modelo fue presentado por los autores en su artículo “The power of feedback” (Hattie y Timperley, 2007), actualmente este trabajo es uno de los más citados de acuerdo a las estadísticas del portal Web of Science, en el que reportan 2040 citas.

Ha sido utilizado con diferentes fines como el desarrollo de aplicaciones que ofrecen retroalimentaciones en un nivel específico (Pardo, *et al.*, 2019), análisis de comentarios (Brooks, *et al.*, 2019; Ajjawi y Boud, 2017; Harris, *et al.*, 2015) y como estrategia para mejorar el desempeño de docentes (Ramírez y Lozano 2019). Se destaca que codificar las retroalimentaciones con los niveles propuestos en el modelo es una manera viable de analizar (Brooks *et al.*, 2019; Gan y Hattie, 2014).

En su trabajo, señalan que, la retroalimentación es una de las más poderosas influencias para el aprendizaje y logro de objetivos, pero su impacto puede ser positivo o negativo. Evidencian que, aunque la retroalimentación tiene un impacto el tipo y la forma en que se proporciona puede ser diferencialmente efectiva.

Describen la retroalimentación como información provista por un agente (profesor, par, pariente, uno mismo, experiencia) respecto aspectos de rendimiento y desempeño de alguien. Resaltan que la retroalimentación es una consecuencia del desempeño, es decir que sucede después de que un estudiante ha realizado una actividad o demostrado una habilidad.

---

Señalan que, para asumir una retroalimentación con propósito instruccional (aquella que ayude al estudiante a mejorar su desempeño), esta necesita proveer información específica sobre la tarea o proceso de aprendizaje que cierre la brecha entre lo que es entendido y lo que se espera haya entendido. Esta brecha, indican puede reducirse a través de procesos afectivos como un mayor esfuerzo, motivación o compromiso, con el uso de procesos cognitivos como la reestructuración de entendimientos, confirmación a los estudiantes de que están en lo correcto o incorrecto, indicando que más información está disponible o es necesaria, señalando direcciones que el estudiante puede seguir e indicar estrategias alternativas para entender una información particular.

Como resultados de su meta-investigación señalan que las formas más efectivas de retroalimentación proveen pistas o refuerzos para los estudiantes, son en la forma de video, audio, o retroalimentación efectiva asistida por computadora y están relacionadas con los objetivos. Destacan que esta es más efectiva cuando da información sobre las respuestas correctas en lugar de las incorrectas y cuándo se basa en cambios de tareas previas.

A partir de estos hallazgos realizan un modelo en el que indican que las retroalimentaciones efectivas deben responder tres grandes cuestionamientos realizados por el profesor o estudiante: ¿A dónde voy? (¿cuáles son los objetivos?), ¿cómo llego? (¿Qué progreso se tiene para lograr el objetivo?) y ¿A dónde continuar? (¿Qué actividades necesitan hacerse para tener un mejor progreso?). Estas preguntas corresponden a las nociones de feed up, feed back y feed forward. Los autores, indican que la efectividad de las respuestas a estas preguntas para reducir la brecha de aprendizaje depende en parte del nivel en el que opera la retroalimentación. Esto incluye el nivel de desempeño de tarea, el nivel del proceso de entendimiento de cómo realizar la tarea, el proceso regulatorio o meta-cognitivo y el nivel personal o del yo (no relacionado a las especificaciones de la tarea).

Hattie y Timperley, indican que la retroalimentación tiene diferentes efectos a través de estos niveles y por tanto es importante asegurar que esté dirigida a estudiantes en el nivel apropiado, porque algunos comentarios son efectivos en reducir la discrepancia entre entendimiento actual y el deseado mientras otros no lo son.

El modelo construido por los autores parte de que el propósito de la retroalimentación debe ser reducir las discrepancias entre lo que se ha entendido o se ha desempeñado un estudiante y el objetivo deseado. Como se mencionó anteriormente, indican que existen varias formas posibles para que los estudiantes reduzcan la brecha entre lo entendido y lo que se espera entiendan, respecto a la retroalimentación señalan que no son siempre efectivas en mejorar el aprendizaje y aquellas que lo son deben responder a tres preguntas ¿a dónde voy? ¿cómo llego? ¿qué sigue después?, a continuación, se describe a detalle cada una.

**¿A dónde voy?** Esta pregunta está relacionada con los objetivos de aprendizaje. Los autores describen que un aspecto crítico de la retroalimentación es la información proporcionada a estudiantes y sus profesores sobre el logro de objetivos de aprendizaje relacionados a la tarea o desempeño.

Para juzgar el éxito del logro del objetivo se pueden considerar varias dimensiones

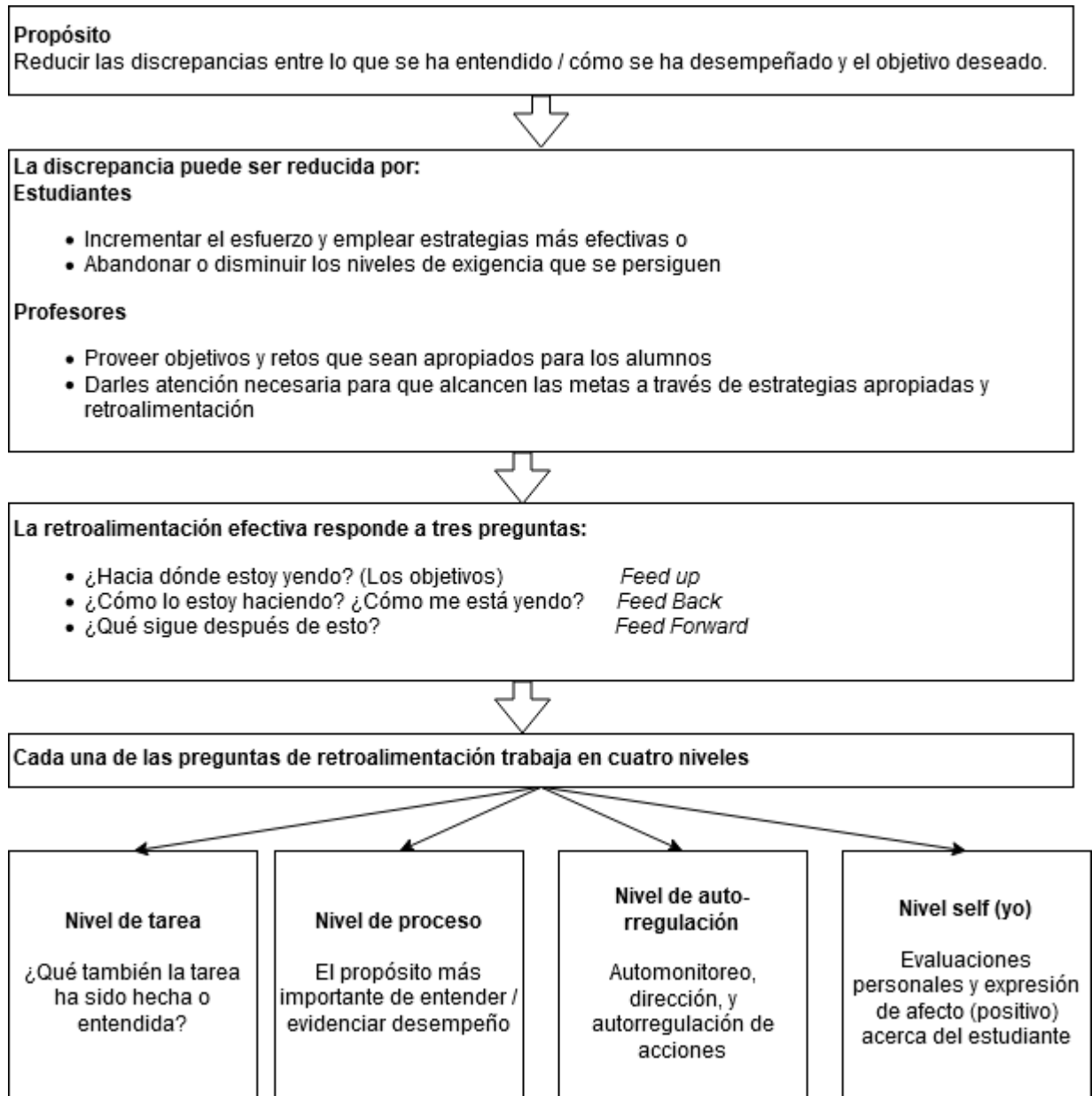


Figura 2.1: Modelo de retroalimentación Hattie y Timperley (2007). Traducido por Dra. Katherina Gallardo.

---

como lo son: directa (aprobación de un examen), comparativa (comentarios como lo hiciste mejor que . . . ), social, relacionada con el nivel de compromiso, o automática y fuera de un contexto específico.

Se señala que los objetivos están relacionados a logros y entendimientos específicos o a diferentes calidades de experiencia y que involucran dos dimensiones: la primera informar a los individuos (como el tipo o nivel de desempeño que debe ser logrados) y la segunda permitir a los estudiantes a plantearse objetivos futuros apropiados y retadores.

**¿Cómo llego?** De acuerdo con los autores, responder esta pregunta implica que un maestro provea información respecto a una tarea o desempeño a menudo en relación a un estándar esperado, rendimiento previo y/o al éxito o fracaso en una parte específica de la tarea. Esta información es efectiva cuando consiste de información sobre el progreso y/o sobre cómo proceder.

**¿Qué sigue?** Respecto a esta pregunta los autores describen que la instrucción a menudo es secuencial, con profesores dando información, tareas o intenciones de aprendizaje, después estudiantes atendiendo tareas y por último alguna consecuencia subsecuente donde dicha consecuencia es frecuentemente más información, más tarea y más expectativas. Destacan que el poder de la retroalimentación puede ser utilizado para dirigir esta pregunta a proveer información que permita ampliar las posibilidades de aprendizaje, lo cual incluye mejorar los retos, más autorregulación del proceso de aprendizaje, mayor fluidez y autonomía, más estrategias y procesos para trabajar en las tareas, entendimiento a profundidad y más información sobre que se ha entendido y que no.

Las preguntas descritas previamente, trabajan de manera conjunta en los cuatro niveles señalados en el modelo: tarea, proceso, autorregulación y del yo, el nivel al que se dirige la retroalimentación influye en la efectividad. Sobre el primer nivel indican que los comentarios pueden ser sobre una tarea o producto por ejemplo si un trabajo es correcto o incorrecto, este nivel de comentario debe incluir indicaciones para adquirir más información diferente o correcta. En el segundo, los comentarios pueden estar dirigidos al proceso usado para crear un producto o completar una tarea, esta clase de retroalimentación está más dirigida al procesamiento de la información o procesos de aprendizaje que requieren comprender o completar la tarea. En el tercero, la retroalimentación a los estudiantes puede estar enfocada al nivel de la auto-regulación, incluyendo estrategias para mejorar la autoevaluación o confianza para comprometerse más en una tarea, estos comentarios pueden tener mayor influencia en auto-eficacia, competencias auto-reguladoras y creencias propias sobre estudiantes como aprendices, con estas los estudiantes son alentados o informados sobre cómo mejorar una tarea. Por último, la cuarta, la retroalimentación puede ser personal en el sentido de que es dirigida a el “yo”, que a menudo no está relacionada con el desempeño en la tarea.

Con lo anterior, Hattie y Timperley señalan que hay una distinción entre retroalimentación sobre la tarea , sobre el procesamiento de la tarea, sobre la auto-regulación y sobre el “yo” como una persona.

De esta clasificación se destaca que la retroalimentación en el nivel del yo es la

---

menos efectiva, que las ubicadas en el nivel de regulación y nivel proceso son poderosas en términos de procesamiento profundo y dominio de las tareas y las de nivel tarea son eficaces cuando la información de la tarea subsecuente es útil para mejorar estrategias de procesamiento o aumentar la autorregulación.

A continuación, se describe a detalle cada nivel.

### **2.3.1. Retroalimentación sobre la tarea**

Este nivel incluye comentarios sobre cuán bien una tarea ha sido realizada, incluye la distinción de las respuestas correctas e incorrectas, señala el adquirir más o diferente información y construir más aprendizaje base que es información que responde a preguntas de qué, cuándo, dónde y quién, principalmente explícita y que representa elecciones visibles que requieren comprensión mínima.

Hattie y Timperley (2007), señalan que los comentarios en el nivel de tarea son poderosos cuando informan sobre malas interpretaciones y no solo sobre la falta de información. Utilizar solo este tipo de comentario provoca que no se generalice a otras tareas.

### **2.3.2. Retroalimentación sobre el procesamiento de la tarea**

Este nivel se relaciona con las estrategias de los estudiantes para la detección de errores, los cuales pueden indicar fallos y una necesidad de replantear estrategias, elegir otras, ser más efectivo en su aplicación y/o buscar ayuda.

La información sobre los procesos para realizar una tarea también puede actuar como un mecanismo de señalización y llevar a una búsqueda de información más efectiva y al uso de tareas estratégicas.

De acuerdo con Hattie y Timperley (2007), estos comentarios son más útiles cuando ayudan a los estudiantes a refutar hipótesis erróneas y proveen de dirección para búsqueda y estrategias. También, la retroalimentación a este nivel es más efectiva que las del nivel de tarea por mejorar el aprendizaje profundo.

### **2.3.3. Retroalimentación sobre auto-regulación**

Los comentarios en el nivel de la autorregulación involucran una interacción entre compromiso, control y confianza, dirigen la forma en que los estudiantes monitorean, dirigen y regulan acciones lograr el objetivo de aprendizaje, implica autonomía, auto-control, auto-dirección, y auto-disciplina (Hattie y Timperley, 2007).

Dicha regulación involucra pensamientos auto-generados, sentimientos y acciones que son planeadas y cíclicamente adaptadas para el logro de objetivos personales.

Hattie y Timperley (2007) señalan que existe evidencia de que la retroalimentación que se atribuye el rendimiento y el esfuerzo o la habilidad aumentan el compromiso y el desempeño en las tareas.

---

### 2.3.4. Retroalimentación sobre el yo como persona

La retroalimentación en este nivel incluye comentarios personales como “buena chica” o “gran esfuerzo”, expresan evaluaciones positivas y afecto sobre el estudiante. Contienen poca información relacionada a la tarea y rara vez se convierte en más compromiso con el aprendizaje.

De acuerdo con Hattie y Timperley (2007), este nivel de retroalimentación puede tener un impacto en el aprendizaje solo si permite un cambio en el esfuerzo del estudiante, compromiso o sentimientos de eficacia en relación al aprendizaje o estrategias que usa cuando intenta entender tareas. Se señala que los efectos al nivel del yo son diluidos, a menudo no informativos sobre el desempeño de la tarea y muy influenciados por auto-concepto de los estudiantes para ser efectivo.

La información en este nivel tiene poco valor para el aprendizaje, no es efectivo porque tiene poca información y a menudo desvía la atención de la tarea. Hattie y Timperley (2007) indican que es importante distinguir entre los elogios que alejan la atención de la tarea y elogios dirigidos al esfuerzo, autorregulación, compromiso, o procesos relacionados con la tarea y su desempeño ya que estos últimos pueden ayudar a mejorar la auto-eficacia.

## 2.4. Clasificación

La clasificación, es una tarea predictiva, realizada con técnicas de aprendizaje supervisado, que busca aprender de patrones etiquetados un modelo capaz de predecir la etiqueta (o clase) para ejemplos futuros y nunca antes vistos (Herrera *et al.* 2016). Se utiliza para clasificar cada elemento de un conjunto de datos en uno de los conjuntos predefinidos de clases (Kesavaraj y Sukumaran, 2013).

Para esta tarea, se divide el conjunto de datos en dos subconjuntos, el primero contiene las características de entrada y el segundo almacena los atributos de salida llamadas clases o etiqueta asignadas a cada instancia.

De acuerdo con Herrera et al. (2016), los algoritmos de clasificación inducen el modelo analizando la correlación entre características de entrada y clases de salida. Una vez que se obtiene el modelo entrenado, puede ser usado para procesar el conjunto de características de nuevos ejemplos obteniendo una clase predicha.

Dependiendo de la naturaleza del segundo subconjunto de atributos, que contiene la clase, varios problemas de clasificación pueden identificarse dependiendo del número de salidas y sus tipos los cuales pueden ser binaria, multiclase, multi-etiqueta, multi-dimensional y aprendizaje de instancias múltiples.

### 2.4.1. Clasificación binaria

En el problema de clasificación binaria, las instancias en un conjunto de datos binario tienen un solo atributo de salida y esta solo puede tomar dos valores diferentes (positivo/negativo, verdadero/falso, 1/0, ...).

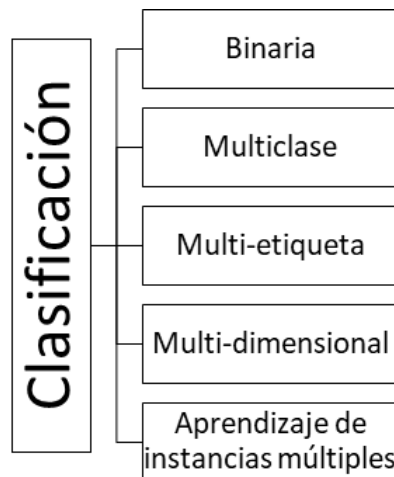


Figura 2.2: Tipos de clasificación según el conjunto de salida

Herrera et al. (2016) señala que un clasificador binario tiene por objetivo encontrar un límite capaz de separar las instancias en dos grupos, una perteneciente a la clase positiva y la otra a la negativa.

### 2.4.2. Clasificación multiclase

En la clasificación multiclase, las instancias de un conjunto de datos también tienen un atributo de salida, como los binarios, pero puede almacenar alguno de un conjunto de valores predefinidos no solo positivo y negativo.

El problema de clasificación multiclase puede resolverse ampliando de forma natural la técnica de clasificación binaria para algunos algoritmos (Aly, 2005). Herrera et al. (2016) indica que varios algoritmos de clasificación multiclase se basan en la binarización, un método que iterativamente entrena un clasificador binario para cada clase contra las otras, siguiendo un enfoque uno contra todos o por cada par de clases, usando un enfoque uno contra uno.

### 2.4.3. Clasificación multi-etiqueta

La clasificación multi-etiqueta permite asociar las instancias a más de una clase, su objetivo es aprender de un conjunto de instancias donde cada instancia pertenece a una o más clases (Sorower, 2010).

En la clasificación multi-etiqueta cada una de las instancias tiene asociado un vector de salidas, en lugar de solo un valor. El tamaño de este vector es ajustado de acuerdo al número de diferentes etiquetas en el conjunto de datos. Cada elemento del vector será un valor binario, indicando si la etiqueta correspondiente es relevante para el ejemplo o no. (Herrera et al., 2016).

---

#### 2.4.4. Clasificación multi-dimensional

Los conjuntos de datos multi-dimensionales también tienen un vector de salida asociado a cada instancia, en vez de un solo valor. Las técnicas multi-dimensionales están estrechamente relacionadas con las usadas en la clasificación multi-etiqueta, con métodos de transformación considerando un rol relevante para atacar la tarea (Herrera et al., 2016).

#### 2.4.5. Aprendizaje de instancias múltiples

El paradigma de instancias múltiple, aprende una etiqueta de clase común para un conjunto de vectores de características de entrada. Este es un problema diferente, dado que cada instancia lógica es definida no solo por un vector de características de entrada, sino también por una colección de instancias físicas, cada una con un conjunto de atributos de entrada. Cada grupo de instancias es conocido bolsa. La etiqueta de clase asociada pertenece a la bolsa en vez de a una instancia individual (Herrera et al., 2016).

### 2.5. Clasificación multi-etiqueta

La clasificación multi-etiqueta como se mencionó anteriormente, es una tarea predictiva de minería de datos con aplicaciones múltiples en el mundo real incluyendo el etiquetado automático de varios recursos como textos, imágenes, música y video.

El aprendizaje de datos multi-etiqueta puede ser logrado a través de diferentes enfoques, como transformación de datos, adaptación de métodos y el uso de conjunto de clasificadores (Herrera et al., 2016). Para llevarlo a cabo se deben considerar aspectos como la dependencia de etiquetas, alta dimensionalidad y etiquetas desbalanceadas.

#### 2.5.1. Definición formal del problema

La principal diferencia entre clasificación tradicional y multi-etiqueta es la salida esperada de los modelos entrenados. Donde un clasificador tradicional devolverá solo un valor, un multi-etiqueta tiene que producir un vector de valores de salida.

Herrera et al. (2016) define formalmente del problema de clasificación multi-etiqueta de la siguiente manera:

*Definición 2.1* Sea  $\chi$  el espacio de entrada, con datos de ejemplo  $X \in A_1 \times A_2 \times \dots \times A_f$ , siendo  $f$ , el número de atributos de entrada y  $A_1, A_2, \dots, A_f$  conjuntos arbitrarios.

*Definición 2.2* Sea  $L$  el conjunto de todas las posibles etiquetas.  $P(L)$  es el conjunto de  $L$ , conteniendo todas las posibles combinaciones de etiquetas  $l \in L$ , incluyendo el conjunto vacío y  $L$  mismo.  $k = |L|$  es el número total de etiquetas en  $L$ .

---

*Definición 2.3* Sea  $\varphi$  el espacio de salida, con todos los posibles vectores  $Y$ ,  $Y \in P(L)$ . El tamaño de  $Y$  siempre será  $k$ .

*Definición 2.4* Sea  $D$  un conjunto de datos multi-etiqueta, conteniendo un subconjunto finito de  $A_1 \times A_2 \times \dots \times A_f \times P(L)$ . Cada elemento  $(X, Y) \in D \mid X \in A_1 \times A_2 \times \dots \times A_f$ ,  $Y \in P(L)$ , será una instancia o dato de ejemplo.  $n = |D|$  será el número de elementos en  $D$ .

*Definición 2.5* Sea  $F$  un clasificador multi-etiqueta, definido como  $F : \chi \rightarrow \varphi$ . La entrada de  $F$  será cualquier instancia  $X \in \chi$ , y la salida será la predicción  $Z \in \varphi$ . Por lo tanto la predicción del vector de etiquetas asociado con cualquier instancia puede ser obtenido como  $Z = F(X)$ .

### 2.5.2. Enfoques del aprendizaje multi-etiqueta

En la literatura se detectan dos enfoques para aprender de conjuntos de datos multi-etiqueta (1) transformación de datos y (2) Adaptación de métodos. Herrera et al. (2016) explica que el primero está basado en técnicas de transformación que aplicadas a los conjuntos de datos multi-etiqueta originales, son capaces de producir uno o más conjuntos de datos binarios o multiclase que pueden ser procesados con clasificadores tradicionales. El segundo, busca adaptar algoritmos existentes, para que puedan trabajar con datos multi-etiqueta, produciendo varias salidas en vez de una sola. Una tercera alternativa, que emerge de la primera, es el uso de conjuntos de clasificadores.

A continuación se detalla cada uno de los enfoques.

#### El enfoque de transformación de datos

Uno de los primeros enfoques para resolver este problema es transformarlo en uno o más problemas simples. La idea de transformación busca partir de un conjunto de datos multi-etiqueta y generar conjuntos de datos que puedan ser procesados por clasificadores binarios o multiclase. Una vez que cada clasificador produce una salida, tienen que ser transformada con objeto de obtener una predicción multi-etiqueta.

Herrera et al. (2016) indica que entre los métodos propuestos en la literatura se encuentran:

- Selección de una sola etiqueta. Transformación llamada Modelo-s, s refiriéndose a etiqueta única. Cuando un ejemplo es asociado con un conjunto de etiquetas, una de estas es asignada como una sola clase. Esta selección puede ser al azar o estar basada en algún método heurístico. El resultado es un conjunto de datos multiclase teniendo el mismo número de instancias que el original pero cada ejemplo tiene solo una clase.
- Ignorar instancias multi-etiqueta. Descartando todos los ejemplos asociados con más de una etiqueta, esta transformación obtiene un nuevo conjunto de datos de

---

naturaleza multiclase con una sola etiqueta por instancia. El conjunto de datos resultado tendrá menos instancias que el original.

- Despliegue de ejemplos con etiqueta múltiples. Este método descompone cada instancia en tantas instancias como etiquetas contenga, clonando los atributos de entrada y asignando a cada ejemplo una de las etiquetas. Un peso puede ser opcionalmente asignado a cada etiqueta, dependiendo de su distribución en el conjunto de datos multi-etiqueta. La salida de esta transformación es también un conjunto de datos multiclase, en este caso se tendrán más ejemplos que el original.
- Usando el conjunto de datos como un identificador de clase. En vez de eliminar etiquetas o ejemplos, este modelo propone usar cada combinación diferente de etiqueta como un identificador de una nueva clase. El conjunto de datos resultante tendrá el mismo número de instancias pero solo una clase. Por tanto, puede ser procesada con cualquier clasificador multiclase. Este método de transformación es mejor conocido como LP (Label PowerSet).
- Aplicando técnicas de binarización. El enfoque más usual consiste en entrenar  $k$  clasificadores, uno por cada etiqueta, tomando las instancias en las cuales la etiqueta aparece como positiva y todos los demás como negativa. Otra propuesta, llamada entrenamiento cruzado, también entrena  $k$  clasificadores pero prefiere usar ejemplos con etiquetas múltiples siempre como positivas. Esta transformación es conocida como BR(Binary Relevance).

Herrera et al. (2016) señala que en los últimos años, las transformaciones LP y BR han sido la base de múltiples soluciones de clasificación multi-etiqueta.

## Enfoque de adaptación de métodos

Un segundo enfoque para la clasificación multi-etiqueta es la adaptación de métodos. Es decir, que algunos modelos de clasificación que fueron inicialmente diseñados para resolver problemas binarios después fueron extendidos para también considerar casos de multi-clase. En la literatura se pueden encontrar ejemplos como las máquinas de vectores de soporte multi-etiqueta, clasificador kNN multi-etiqueta que puede predecir la clase más frecuente a través de los vecinos de la nueva instancia, donde hay dos o más clases (Herrera et al., 2016).

## Ensamblaje de clasificadores

El ensamblaje de clasificadores es una técnica que busca mejorar el desempeño obtenido por clasificadores individuales. Un ensamble está compuesto por un conjunto de clasificadores, cuyas entradas son usualmente combinadas por promedios con o sin peso (Herrera et al., 2016).

---

### 2.5.3. Métodos de transformación

En este apartado se describen los métodos de transformación utilizados en este trabajo.

#### Relevancia binaria

De acuerdo con Giraldo (2015), es una estrategia conocida como uno contra todos, en la que el número de clasificadores entrenados es igual al número de clases. Para entrenar cada clasificador, se usa toda la matriz de características, mientras que solamente se considera la  $q$ -ésima columna de la matriz de etiquetas.

Cada clasificador binario, predice una de las etiquetas asociadas a la instancia. Es una estrategia utilizada pero criticada por no tener en cuenta la correlación entre clases, de igual forma, esta estrategia condice a problemas de desbalance, lo que da como resultados bajos niveles de sensibilidad cuando técnicas de balance no son incorporadas.

#### Conjunto potencia (LP)

Descrita en Giraldo (2015), como método de transformación que genera nuevas clases teniendo en cuenta todas las muestras que compartan un mismo conjunto de etiquetas. Por ejemplo, un conjunto de datos con instancias multi-etiqueta que pertenecen simultáneamente a las clases A y B, estas clases son combinadas en una nueva etiqueta. Con esta nueva combinación, el número de clases involucradas en el problema puede aumentar considerablemente, y algunas clases pueden contener pocas muestras, lo que conlleva al principal problema con el que cuenta este método, el desbalance de clases. Este método no debe ser utilizado cuando el conjunto de datos posee un gran número de clases, y cuando las relaciones entre estas son muy variadas, ya que el número de clasificadores que se deben utilizar se incrementa exponencialmente, con una complejidad de  $2^Q$  en el peor de los casos siendo  $Q$  el número de clases.

#### Conjuntos aleatorios Rakel

Es una alternativa para tener en cuenta las correlaciones entre las clases, de una manera sencilla como en el caso de la clasificación mediante LP, pero sin sufrir el problema del aumento desmedido de la cantidad de clasificaciones que se deben efectuar en LP. Este método propuesto por Tsoumakas propone partir un conjunto grande de etiquetas en varios grupos de menor tamaño, donde para entrenar cada subconjunto es utilizado el método LP. Para esta alternativa se proponen dos tipos de subconjuntos, Rakel-d y Rakel-o, en el primero los subconjuntos construidos son disjuntos, mientras que para el segundo se presenta un cierto solapamiento (Giraldo, 2015).

#### Clasificadores en cadena

Estos clasificadores modelan las correlaciones entre etiquetas manteniendo una complejidad del mismo orden que el método de relevancia binaria. En este esquema, los cla-

---

sificadores son entrenados en un orden predefinido, al igual que en relevancia binaria, es necesarios construir un clasificador por etiqueta, pero esta vez, la matriz de características para el  $k$ -ésimo clasificador es enriquecido con la salida de los clasificadores anteriores.

Los principales inconvenientes de este enfoque son: el orden de la cadena influye en los desempeños de clasificación, emplear un método heurístico de búsqueda de la cadena tendría un costo computacional demasiado alto cuando el número de clases es elevado.

#### 2.5.4. Características de los conjuntos de datos multi-etiqueta

Una de las tareas previo a intentar construir un clasificador multi-etiqueta para resolver un problema específico, se debe analizar las características principales de los datos disponibles. Herrera et al. (2016) indica que entender los rasgos internos de los datos permitirá la selección del mejor algoritmo, parámetros, etc. En la literatura se encuentran métricas para analizar las características de los conjuntos de datos multi-etiqueta como la cardinalidad y densidad que se describen a continuación.

##### Métricas básicas

Las métricas de análisis de conjuntos de datos multi-etiqueta tienen como propósito evaluar el nivel de multiplicidad de etiquetas en los datos, en otras palabras, determinan el grado en que los ejemplos del conjunto de datos tienen más de una etiqueta (Herrera et al., 2016).

##### Cardinalidad

Esta métrica propuesta por Tsoumakos *et al.*, (2007), consiste en contar el número de etiquetas relevantes para cada instancia en el conjunto de datos, promediando la suma para conocer el número promedio de etiquetas por instancia ver Fórmula 2.1.

$$Card(D) = \frac{1}{n} \sum_{i=1}^n |Y_i| \quad (2.1)$$

En este contexto,  $n$  denota el número de instancias en el conjunto de datos multi-etiqueta  $D$ ,  $Y_i$  el conjunto de etiquetas de la  $i$ -ésima instancia, y  $k$  el número total de etiquetas consideradas en  $D$ .

(Herrera et al., 2016) señala que mientras mayor sea el nivel de la cardinalidad mayor será el número de etiquetas activas por instancia.

##### Densidad

El valor de  $Card$  es una métrica influenciada por el tamaño del conjunto de etiquetas usado por cada conjunto de datos multi-etiqueta, y es expresado usando el número de etiquetas como unidad de medida, una versión normalizada de esta métrica es la densidad que se obtiene de Dividir  $Card$  por el número de etiquetas en el conjunto de datos multi-etiqueta, ver Formula 2.2.

Herrera *et al.*, 2016 señala que un valor alto en  $Dens$  indica que las etiquetas en el conjunto de datos multi-etiqueta están bien representadas en cada instancia. En contraste valores bajos en  $Dens$  denotan más dispersión, con solo un conjunto pequeño de las etiquetas presente en la mayoría de las instancias.

$$Dens(D) = \frac{1}{k} \frac{1}{n} \sum_{i=1}^n |Y_i| \quad (2.2)$$

$P_{min}$

Otra forma de medir el nivel de multi-etiqueta de un conjunto de datos sería a través de promedios de la métrica  $P_{min}$  propuesta por Turner *et al.*, (2013) que es un simple porcentaje de instancias en el conjunto de datos multi-etiqueta con solo una etiqueta activa, Ver formula 2.3. Un valor alto de  $P_{min}$  denotaría que una larga proporción de instancias tienen una sola etiqueta (Herrera *et al.*, 2016).

$$P_{min}(D) = \sum_{y' \in Y \mid |y'|=1} \frac{|y'|}{n} \quad (2.3)$$

### Diversidad de etiquetas

De acuerdo con Herrera et al. (2016), subconjuntos de etiqueta en el conjunto  $L$  aparecen en las instancias de  $D$  formando conjuntos de etiquetas. Teóricamente  $2^k$  conjuntos diferentes de etiquetas diferentes podrían existir, pero en la práctica el número único de conjuntos de datos (diferentes) está limitado por el número de instancias en  $D$ . Por tanto el número de combinaciones únicas es limitado por la expresión  $\min(n, 2^k)$ . El número efectivo de distintos conjuntos de etiquetas en un conjunto de datos multi-etiqueta es un indicador de la uniformidad en la distribución de etiquetas entre los ejemplos. Lo mayor que este valor sea, lo más irregular que las etiquetas aparecen en las instancias de datos. El número de distintos conjuntos de etiquetas es también conocido como diversidad de etiquetas ( $Div$ ).

### Métricas de desbalanceo

Una característica común de los conjuntos multi-etiqueta es que algunas etiquetas son más frecuentes que otras, y debe ser un aspecto a valorar dado su impacto en los resultados de clasificación. Tres diferentes métricas para evaluar el desbalanceo de etiquetas se encuentran en la literatura llamadas  $IRLbl$ ,  $MaxIR$  y  $MeanIR$ . En las fórmulas siguiente el operador  $[[expression]]$  denota el corchete de inversión. Devolverá 1 si la expresión dentro es verdadera o 0 de otra forma.

$$IRLbl(l) = \frac{\max_{l' \in L} (\sum_{i=1}^n [[l' \in Y_i]])}{\sum_{i=1}^n [[l \in Y_i]]} \quad (2.4)$$

Con el valor de  $IRLbl$  es posible conocer el nivel de desbalanceo de una etiqueta específica (Herrera et al., 2016). Es computada como la proporción entre el número de apariciones de la etiqueta más común y la etiqueta considerada, ver Fórmula 2.4. Para

la etiqueta más común se tendrá  $IRLbl = 1$ . Para etiquetas menos frecuentes, el nivel siempre será mayor que 1. Mientras mayor sea el  $IRLbl$ , más rara es la presencia de la etiqueta en conjunto de datos multi-etiqueta.

El objetivo del valor  $MaxIR$  es obtener la proporción máxima de desbalanceo, ver Fórmula 2.5.

$$MaxIR = \max_{l \in L} (IRLbl(l)) \quad (2.5)$$

$$MeanIR = \frac{1}{k} \sum_{l \in L} IRLbl(l) \quad (2.6)$$

Una evaluación global del desbalanceo en conjunto de datos multi-etiqueta es  $MeanIR$ , calculada promediando el  $IRLbl$  de todas las etiquetas, ver Fórmula 2.6. De acuerdo con Herrera et al. (2016) un alto valor de  $MeanIR$  podría ser debido a valores relativamente en altos  $IRLbl$  para varias etiquetas, pero también por valores extremos de desbalanceo para solo algunas etiquetas.

## Otras métricas

En la literatura se pueden encontrar otras métricas para análisis de los conjuntos de datos multi-etiqueta. Una de estas es el valor  $SCUMBLE$  propuesta por Charte et al., (2014), que mide la concurrencia entre etiquetas frecuentes y raras. Un valor es computado individualmente para cada instancia ver Fórmula 2.7.

Herrera et al. (2016) explica que este valor está basado en el índice Atkinson y el valor  $IRLbl$ . El primero es una medición econométrica que busca evaluar entradas desiguales entre la población. En este contexto, cantidades monetarias han sido remplazadas por proporciones desbalanceadas, provistas por el valor  $IRLbl$ . El resultado es un valor en el rango  $[0, 1]$  indicando si todas las etiquetas en la instancia tienen frecuencias similares en el conjunto de datos multi-etiqueta, valores bajos, o por el contrario hay diferencias significativas, el resultado sería un valor alto.

El valor global SCUMBLE es obtenido promediando el valor de todas las instancias en el conjunto de datos multi-etiqueta ver Fórmula 2.8. Como regla general, altos valores de SCUMBLE denotan complejos conjuntos de datos multi-etiqueta para aprender de ellos.

$$SCUMBLE_{ins}(i) = 1 - \frac{1}{IRLbl_i} \left( \prod_{l \in L} IRLbl_{il} \right)^{\frac{1}{k}} \quad (2.7)$$

$$SCUMBLE(D) = \frac{1}{n} \sum_{i=1}^n SCUMBLE_{ins}(i) \quad (2.8)$$

Otra métrica propuesta en Charte et al., (2016), es el valor  $TCS$  que es indicador teórico de complejidad, es calculado como el producto del número de características de entrada, número de etiquetas y el número de diferentes combinaciones de etiquetas,

---

ver fórmula 2.9. Para evitar trabajar con valores demasiado grandes que dificulte la interpretación y comparación, la función log es usada para ajustar la escala del producto previo. El objetivo es determinar cuál conjunto de datos multi-etiqueta presentará mayor esfuerzo de trabajo para el pre-procesamiento y algoritmos de aprendizaje.

$$TCS(D) = \log(f \times k \times |conjuntodedatosunicos|) \quad (2.9)$$

### 2.5.5. Evaluación del desempeño de clasificadores

La salida de un clasificador multi-etiqueta consiste en un conjunto de etiquetas predichas para cada instancia de prueba. De acuerdo con Herrera et al. (2016) una predicción multi-etiqueta puede ser completamente correcta, parcialmente correcta/incorrecta o totalmente incorrecta por lo que aplicar las mismas métricas usadas en clasificación tradicional es posible, pero suelen ser excesivamente estrictas.

Herrera et al. (2016) señala que todas las métricas de evaluación pueden ser agrupadas conforme a dos criterios:

- Cómo se computó la predicción. Una medida puede ser realizada por instancia o por etiqueta, dando como resultado dos diferentes grupos de métricas:
  - Métricas basadas en ejemplos. Estas métricas son calculadas de manera independiente para cada instancia y después promediadas dividiendo entre el número de ejemplos.
  - Métricas basadas en etiqueta. En contraste a las anteriores, estas son computadas independientemente para cada etiqueta previa a que son promediadas. Para hacerlo, dos diferentes estrategias pueden aplicarse:
    - Macro-promedio: La métrica es calculada individualmente para cada etiqueta y el resultado promediado dividiendo por el número de etiquetas ( $k$ ).
    - Micro-promedio: El conteo de éxitos y errores por cada etiqueta son agregadas al principio, y después la métrica es computada solo una vez.
- Cómo el resultado es provisto. La salida producida por un clasificador multi-etiqueta puede ser una bi-partición binaria de etiquetas o una jerarquización de etiquetas. Algunas de estas proveen ambos resultados.
  - Bi-Partición binaria: Una bi-partición binaria es un vector de 0s y 1s indicando cuales de las etiquetas pertenecientes al conjunto de datos multi-etiqueta son relevantes al ejemplo procesado. Hay métricas que operan sobre estas bi-particiones, usando los conteos de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos.
  - Ranking de etiquetas: La salida es una lista de etiquetas ordenadas de acuerdo a alguna medida de relevancia. Una bi-partición binaria puede ser obtenida de una etiqueta ordenada a través de aplicar un umbral, proporcionada

---

por el clasificador mismo. Sin embargo, hay métricas de desempeño que trabajan con jerarquías en bruto para computar la medida, en lugar de usar conteos de predicciones correctas o incorrectas.

## Métricas basadas en ejemplos

**Hamming Loss** Mide cuán bien el clasificador predice cada una de las etiquetas promediando por etiquetas y luego por instancias (Szymanski y Kajdanowicz, 2017). Se calcula con la fórmula 2.10.

$$HammingLoss = \frac{1}{n} \frac{1}{k} \sum_{i=1}^n |y_i \Delta z_i| \quad (2.10)$$

El operador  $\Delta$  devuelve la diferencia entre  $y_i$ , el conjunto real de la instancia  $i$ -ésima y  $z_i$ , la predicha. El operador  $|r|$  cuenta el número de 1s en esta diferencia, en otras palabras, el número de predicciones erróneas. El número total de errores en las  $n$  instancias es agregada y después normalizada tomando en cuenta el número de etiquetas y número de instancias.

**Exactitud (Accuracy)** Mide cuán bien el clasificador predice las combinaciones de etiquetas promediadas por instancias (Szymanski y Kajdanowicz, 2017). La fórmula 2.11 muestra la forma de calcularla:

$$Accuracy = \frac{1}{n} \sum_{i=1}^n \frac{|y_i \cap z_i|}{|y_i \cup z_i|} \quad (2.11)$$

**Precisión, Recall y F-measure** La precisión (Precision) es considerada una de las métricas más intuitivas para evaluar el desempeño predictivo multi-etiqueta. Es calculada como la proporción entre el número de etiquetas correctamente predichas y el número total de etiquetas predichas, ver Fórmula 2.12. Esto puede interpretarse como el porcentaje de etiquetas predichas que son verdaderamente relevantes para la instancia. Esta métrica es generalmente usada en conjunto con la exhaustividad (Recall), que devuelve el porcentaje de etiquetas correctamente predichas entre todas las etiquetas verdaderamente relevantes, ver Fórmula 2.13. Esto es, la proporción de etiquetas verdaderas que es proporcionada como salida por el clasificador.

$$Precision = \frac{1}{n} \sum_{i=1}^n \frac{|y_i \cap z_i|}{|z_i|} \quad (2.12)$$

$$Recall = \frac{1}{n} \sum_{i=1}^n \frac{|y_i \cap z_i|}{|y_i|} \quad (2.13)$$

F-measure mide el promedio ponderado de la precisión y exhaustividad (Szymanski y Kajdanowicz, 2017). La fórmula 2.14 muestra cómo calcularla.

$$F - measure = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (2.14)$$

---

**Subset Accuracy** Mide la exactitud de la clasificación (Herrera *et al.*, 2016). Se calcula con la fórmula 2.15.

$$SubsetAccuracy = \frac{1}{n} \sum_{i=1}^n [[y_i = z_i]] \quad (2.15)$$

### Métricas basadas en orden

Herrera *et al.*, (2016) explica que las métricas descritas previamente trabajan sobre particiones binarias de etiquetas, así que necesitan un conjunto de etiquetas como salida del clasificador. En contraste, las explicadas en este apartado, necesitan un ordenamiento de etiquetas, es decir, es necesario un grado de confianza o de probabilidad de pertenecer a cada etiqueta.

En las siguientes ecuaciones,  $rank(x_i, l)$  es definido como una función para las instancias  $x_i$  y la etiqueta relevante  $l \in \Upsilon$ , cuya posición es conocida, devuelve  $l$ 's grados de confianza dentro de la predicción  $Z_i$  regresada por el clasificador.

La métrica, promedio de precisión (AvgPrecision) determina para cada etiqueta en una instancia, la proporción de etiquetas relevantes que son ordenadas sobre ella en el ordenamiento predicho, ver Fórmula 2.16 (Herrera *et al.*, 2016).

$$AveragePrecision = \frac{1}{n} \sum_{i=1}^n \frac{1}{|y_i|} \sum_{y \in Y_i} \frac{|y'|rank(x_i, y') \leq rank(x_i, y), y' \in Y_i|}{rank(x_i, y)} \quad (2.16)$$

La métrica coverage (cobertura) cuenta el número de pasos para ir a través del ordenamiento provisto por el clasificador hasta que todas las etiquetas relevantes son encontradas, ver Fórmula 2.17 (Herrera *et al.*, (2016)). Mientras más pequeño sea el número de pasos para el conjunto de datos multi-etiqueta, mejor es el desempeño del clasificador. Mientras más grande es el conjunto más grande será número promedio de pasos para pasar a través del ordenamiento.

$$Coverage = \frac{1}{n} \sum_{i=1}^n \operatorname{argmax}_{y \in Y_i} (rank(x_i, y)) - 1 \quad (2.17)$$

La métrica RLoss (Ranking loss) mide el número de etiquetas ordenadas incorrectamente con respecto al número de ordenadas correctamente (Herrera *et al.*, 2016). La fórmula 2.18 muestra cómo calcularlo.

$$RLoss = \frac{1}{n} \sum_{i=1}^n \frac{1}{|y_i| \cdot |\bar{y}_i|} |y_a, y_b : rank(x_i, y_b), (y_a, y_b) \in Y_i \times \bar{Y}_i| \quad (2.18)$$

### Métricas basadas en etiqueta

Las métricas de desempeño explicadas previas secciones son evaluadas individualmente por cada instancia, y después promediadas dividiendo por el número de instancias

consideradas. Por lo tanto, a cada ejemplo se les da el mismo peso en el resultado final. Por el contrario, en las métricas basadas en etiquetas puede ser computado por el promedio de dos diferentes estrategias de promediado. Estas son conocidas como macro-averaging y micro-averaging.

Para hacerlo, las fórmulas genéricas son usadas (ver fórmulas 2.19, 2.20). *EvalMet* sería una de las métricas mencionadas. En este contexto, TP se refiere a verdaderos positivos, FP para falsos positivos, TN para verdaderos negativos y FN para falsos negativos.

$$MacroMet = \frac{1}{k} \sum_{l \in L} EvalMet(TP_l, FP_l, TN_l, FN_l) \quad (2.19)$$

$$MicroMet = EvalMet\left(\sum_{l \in L} TP_l, \sum_{l \in L} FP_l, \sum_{l \in L} TN_l, \sum_{l \in L} FN_l\right) \quad (2.20)$$

Herrera *et al.*, (2016) describe que en el enfoque de macro-averaging, la métrica es evaluada una vez por etiqueta, usando los conteos acumulados para ello, y después el promedio es obtenido dividiendo por el número de etiquetas. De esta forma el mismo peso es asignado a cada etiqueta, donde es muy frecuente o muy rara. Por el contrario, la estrategia de micro-averaging primero agrega los contadores para todas las etiquetas y después computa la métrica una sola vez. Dado que las predicciones donde etiquetas raras aparecen son combinadas con aquellas hechas por las más frecuentes, los primeros son diluidos entre los segundos. La contribución de cada etiqueta en la medición no es la misma.

## 2.6. Afinación de hiper-parámetros

Los hiper-parámetros son parámetros que no son aprendidos directamente en el entrenamiento de clasificadores y deben ser pasados como argumentos antes de utilizar un algoritmo de clasificación (Mantovani *et al.*, 2015).

La afinación de hiper-parámetros, se refiere a la tarea de encontrar los hiper-parámetros óptimos de un algoritmo de aprendizaje para un conjunto de datos considerado. Un hiper-parámetro óptimo será aquel que logre mejores valores para una métrica de desempeño especificada Probst *et al.* (2019).

La importancia de la afinación de hiper parámetros radica en que se ha demostrado que el desempeño de varios algoritmos de clasificación depende de las configuraciones de los hiper parámetros que pueden hacer la diferencia entre una buena o mala clasificación (Hutter *et al.*, 2015; Weerts *et al.* 2020; Mantovani *et al.*, 2015).

### 2.6.1. Estrategias existentes

Para seleccionar una configuración de hiper parámetros adecuada para un conjunto de datos específico, se suele recurrir a los valores por defecto que dan los paquetes de

---

software o se configuran manualmente basándose en recomendaciones de la literatura, la experiencia o el método de ensayo y error (Probst *et al.*, 2019 ).

Otra alternativa, es el uso de estrategias para la afinación de hiper parámetros que son dependientes de los datos. Estas estrategias, intentan minimizar el error de generalización esperado del algoritmo sobre un espacio de búsqueda de hiper parámetros de configuraciones candidatas. Evalúan las predicciones en un conjunto de pruebas independiente, o ejecutando un esquema de muestreo como la validación cruzada (Probst *et al.*, 2019 ).

Para realizar una búsqueda de hiper parámetros óptimos se debe considerar: un algoritmo, en este caso de clasificación, un espacio de búsqueda de parámetros, un método de búsqueda o selección de candidatos, un esquema de validación cruzada y una métrica a optimizar.

Entre los métodos para realizar la búsqueda de hiper parámetros se encuentra la búsqueda de malla, búsqueda aleatoria y optimización secuencial basado en modelos. En Probst *et al.*, (2019) ) se describe a cada una de la siguiente manera:

- **Búsqueda de malla.** Estrategia en la que se evalúan todas las combinaciones posibles de espacios de parámetros discretos dados.
- **Búsqueda aleatoria.** Estrategia en la que los valores de los hiper parámetros se obtienen de forma aleatoria de un espacio específico de hiper parámetros.
- **Optimización secuencial basada en modelos.** Estrategia de ajuste que iterativamente intenta encontrar los mejores valores de los hiper parámetros basado en evaluaciones de los hiper-parámetros que fueron obtenidos previamente.

En este trabajo se utiliza la búsqueda de malla como estrategia para la afinación de hiper-parámetros que genera exhaustivamente candidatos a partir de una malla de valores de parámetros especificados previamente. Evalúa todas las posibles combinaciones de valores de los parámetros devolviendo la que mejor desempeño tuvo para la métrica seleccionada.

### 2.6.2. Medición de la afinación

Para evaluar la afinación de hiper-parámetros de un algoritmo para un conjunto de datos específico, se calcula la diferencia entre el desempeño de una configuración general de referencia como pueden ser los valores por defecto de algún programa de software y el desempeño de la mejor configuración posible de hiper parámetros en ese conjunto de datos (Probst *et al.*, 2019 ).

Se utiliza la validación cruzada, que es un método de muestreo para evaluar la habilidad de generalización de modelos predictivos y prevenir el sobre ajuste (Hutter *et al.*, 2015; Probst *et al.*, 2019).

También, se debe elegir una métrica que será dependiente del problema de aprendizaje. De acuerdo con Probst *et al.*, (2019) cuando se trata de clasificación las métricas más utilizadas son la exactitud y error bajo la curva AUC.

---

# Capítulo 3

## Marco Contextual

En este capítulo se desarrolla el contexto dónde se llevó a cabo la investigación. Se hace una descripción de la Universidad Pública donde se realizó el estudio, la forma en que aborda la modalidad a distancia y estadísticas de admisión, matrícula y egreso.

### 3.1. Universidad donde se realizó la investigación

La investigación se realizó en una institución mexicana pública y autónoma ubicada en la Ciudad de Puebla, México que tiene como misión la formación integral de profesionales y ciudadanos críticos y reflexivos en los niveles de educación media superior, superior y posgrado, que son capaces de generar, adaptar, recrear, innovar y aplicar conocimientos de calidad y pertinencia social.

Las raíces de la universidad, se remontan al siglo XVI, cuando un grupo de jesuitas realizan la fundación de la institución la cual ha modificado su nombre debido a los acontecimientos descritos en la Tabla 3.1.

Actualmente la institución, constituye un gran pilar de la educación superior y la investigación científica en la región y ocupa un destacado sitio entre las universidades públicas del país, gracias al esfuerzo conjunto de todos los miembros de la institución.

#### 3.1.1. Oferta educativa

Para el periodo escolar 2019-2020 la universidad donde se realizó la investigación ofertó 2 programas de bachillerato, 1 de preparatoria a distancia, 3 carreras técnicas, 3 técnico superior universitario, 83 licenciaturas y 91 posgrados (BUAP, 2020). La matrícula que tuvo en el mismo periodo fue de 105722 estudiantes. La distribución de estudiantes se muestra en la Tabla 3.2.

Las áreas de conocimiento que abarca la oferta educativa de nivel licenciatura son: (1) ciencias de la salud, (2) ciencias exactas, (3) ciencias naturales, (4) ciencias sociales y administrativas, (5) educación y Humanidades e (6) ingeniería y tecnología.

Tabla 3.1: Acontecimientos que modificaron nombre de la institución

Año	Evento
1578	Un grupo de religiosos jesuitas establecen su residencia en Puebla y a petición expresa del cabildo de la ciudad fundaron el Seminario de la Compañía de Jesús de San Jerónimo.
1587	Fundación de la institución con el nombre Colegio del Espíritu Santo.
1670	Se inicia la construcción del edificio Carolino.
1767	Todos los colegios dirigidos por jesuitas se unieron en uno solo al que se le denominó Real Colegio Carolino.
1820	Regresan los jesuitas a Puebla y el instituto se renombra como Real Colegio del Espíritu Santo de San Jerónimo y San Ignacio de la Compañía de Jesús.
1825	El congreso local lo transforma en Colegio del Estado lo que le da un carácter público, laico y gratuito.
1917, 1923 y 1932	Los estudiantes del colegio del Estado demandan la transformación de éste en universidad con plena autonomía.
1937	El gobernador del estado Maximino Ávila Camacho anunció la transformación del colegio en Universidad. Se expide la Ley Orgánica de la Universidad de Puebla y se nombró como primer rector de la institución al licenciado Manuel L. Márquez.
1951	El gobernador Rafael Ávila Camacho intenta militarizar la institución.
1956	El Diario Oficial del Estado de Puebla publica la ley de la Universidad Autónoma de Puebla.
1968	El gobierno del estado entrega las instalaciones de Ciudad Universitaria, construida por la Fundación Mary Street Jenkins.
1970	Se impone un modelo de Universidad Crítica, Democrática y Popular. Se crea el Instituto de Ciencias Se consolida la escuela de físico matemáticas Nacen los primeros estudios de posgrado: maestría y doctorado en física.
1987	El congreso del estado le otorga el título de Benemérita.

Tabla 3.2: Distribución de estudiantes por nivel educativo

Nivel educativo	Matricula
Bachillerato o preparatoria	19484
Técnico	914
Técnico superior universitario	452
Licenciatura	81599
Posgrado	3270

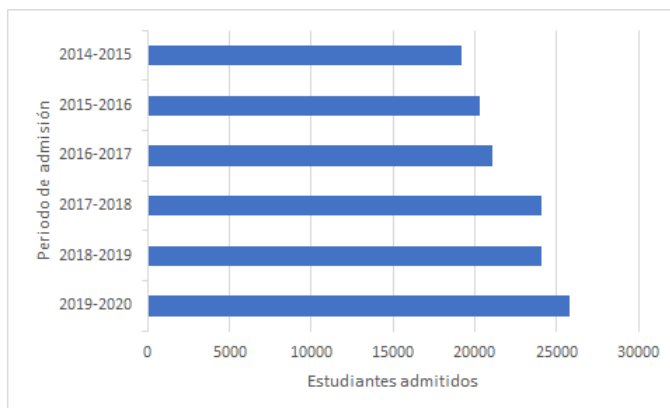


Figura 3.1: Estudiantes admitidos por periodo escolar.

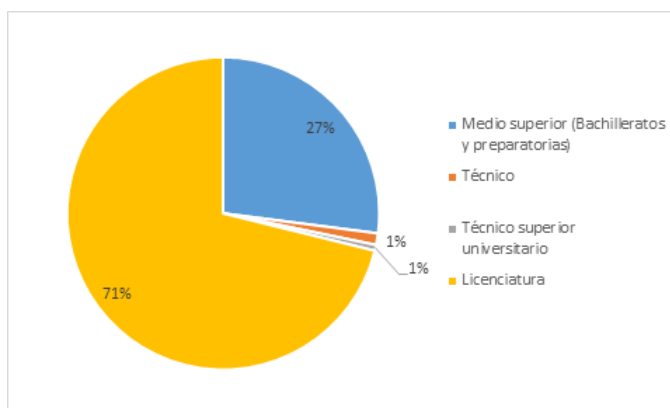


Figura 3.2: Distribución de estudiantes admitidos por nivel educativo.

### 3.1.2. Admisión

De la revisión de los anuarios estadísticos institucionales de los últimos 5 años de la universidad donde se realizó el estudio (BUAP, 2020; BUAP, 2019; BUAP, 2018; BUAP, 2017; BUAP, 2016), se detecta que recibe en promedio 22408 estudiantes por proceso de admisión, cantidad que va en aumento año con año ver Figura 3.1.

La mayor parte de los estudiantes admitidos se concentra en programas de licenciatura con 71 % de los admitidos, 27 % de nivel medio superior ver Figura 3.2.

### 3.1.3. Matrícula

La matrícula de estudiantes en promedio de los últimos cinco años es de 92552 estudiantes (BUAP, 2020; BUAP, 2019; BUAP, 2018; BUAP, 2017; BUAP, 2016). La cantidad de estudiantes matriculados crece año con año. La mayor concentración de la matrícula se encuentra en el nivel licenciatura seguido del nivel medio superior ver Figura 3.3.

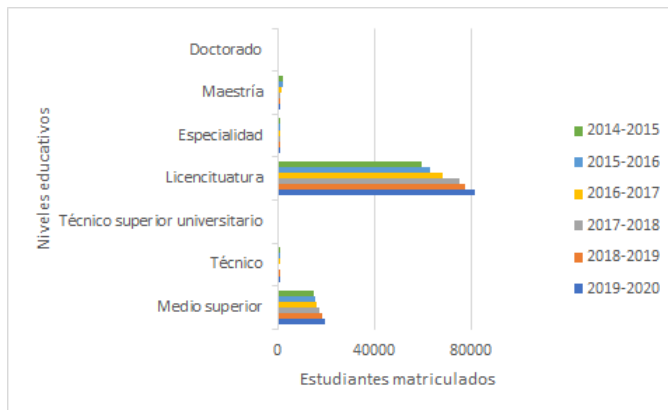


Figura 3.3: Distribución de estudiantes matriculados por nivel educativo en diferentes periodos.

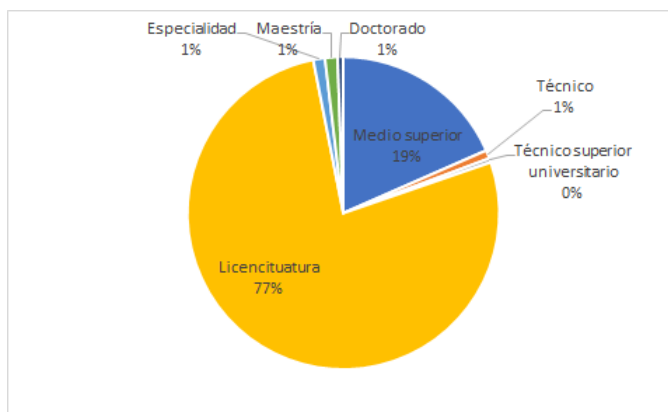


Figura 3.4: Distribución de estudiantes matriculados por nivel educativo en el periodo 2019-2020.

La concentración de la matrícula del periodo 2019-2020 se muestra estar en 77% para licenciatura, 19% para nivel medio superior y el resto en programas de posgrado ver Figura 3.4.

### 3.1.4. Egreso

En promedio, en los últimos cinco años 12134 estudiantes egresan de la institución (BUAP, 2020; BUAP, 2019; BUAP, 2018; BUAP, 2017; BUAP, 2016). La Tabla 3.3 muestra las cifras por periodo y nivel educativo.

### 3.1.5. Docentes

Durante el año 2019-2020 la planta académica estaba conformada por 5157 docentes de los cuales 957 forman parte del nivel medio superior y 4115 del nivel superior esto

Tabla 3.3: Egresados por nivel educativo

Nivel educativo	2019-2020	2018-2019	2017 -2018	2016-2017	2015-2016	2014-2015
Nivel medio superior	5815	5094	4750	4229	4488	4536
Técnico	35	111	119	78	95	175
Técnico superior universitario	41	24	44	15	20	18
Licenciatura	4930	6208	6689	5890	6427	5883
Especialidad	87	343	390	388	401	305
Maestría	480	349	936	953	1088	857
Doctorado	112	50	107	88	98	59
Total	11500	12179	13035	11641	12617	11833

Tabla 3.4: Distribución docentes por tipo de contratación

	PHC	Tiempo completo	Medio tiempo	Total
Nivel medio superior	596	217	144	957
Nivel superior	1784	1985	346	4115
Total	2380	2202	490	5,072

conforme lo señalado en el anuario estadístico del año señalado (BUAP, 2020).

Las Figuras 3.4 muestran la distribución de docentes por tipo de contratación.

## 3.2. Modalidad a distancia en BUAP

En la institución dónde se llevó a cabo la investigación se define la educación a distancia como "Modalidad educativa centrada en el aprendizaje que promueve el desarrollo autónomo del alumno como un proceso mediado por las tecnologías de la información y de la comunicación, con estrategias y recursos que permiten la interacción a los diferentes actores y que tiene como fin propiciar el aprendizaje. Se realiza en tiempos y espacios distintos entre docentes facilitadores y alumnos favoreciendo la comunicación multidireccional" (BUAP, 2015).

Las modalidades alternativas de educación tienen su inicio en la universidad en 1997, cuando se crea el Sistema de educación a Distancia, en el 2000 se transforma en el Sistema Universitario de Educación Semiescolarizado y a Distancia, para 2005 se

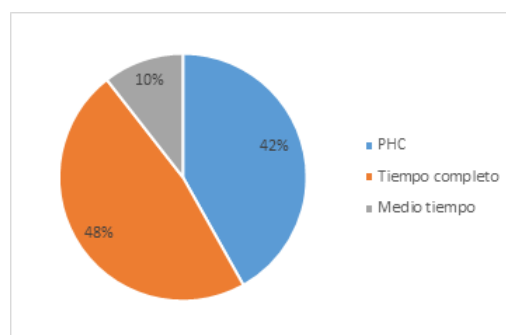


Figura 3.5: Distribución de tipo de contratación de docentes.

---

constituye la Dirección General de Modalidades Alternativas de Educación, para 2007 con el fin de integrar en una coordinación a las distintas opciones se crea la Dirección General de Innovación Educativa la cual cambia su nombre a Dirección de Tecnología educativa en 2019.

En la institución, los programas educativos en modalidades alternativas o no convencionales e representan la oportunidad para:

- Ampliar la cobertura educativa.
- Extender la oferta educativa a regiones que no disponen de sistemas presenciales.
- Favorecer la inclusión al permitir el acceso a grupos en desventaja social, capacidades diferentes, amas de casa y privados de su libertad entre otros.
- Detonar la investigación educativa y la innovación en competencias tecnológicas y trabajo colaborativo virtual.
- Contribuir a los procesos de globalización internacionalización de la ciencia y la tecnología en beneficio de la formación profesional, para lograr una alta competitividad de los alumnos.
- Retroalimentar la labor docente en la Universidad.
- Currículo flexible.

Las primeras licenciaturas en modalidad a distancia se abren en agosto de 2011 entre las que se encuentran: administración, comunicación, contaduría y derecho. En 2014 se amplía la oferta al incorporar otras tres licenciaturas (negocios internacionales, mercadotecnia y medios digitales, administración y dirección de PyMEs). En 2017 se abre la licenciatura en contaduría y finanzas públicas. En 2021 se incorpora a la oferta la licenciatura en procesos educativos.

Actualmente esta modalidad abarca el nivel de educación media superior con la preparatoria a distancia y el nivel de educación superior con 9 licenciaturas entre las que se encuentra: administración de empresas, comunicación, contaduría pública, derecho, mercadotecnia y medios digitales, negocios internacionales, administración y dirección de pymes, contaduría y finanzas públicas y procesos educativos.

### **3.2.1. Modelo pedagógico de la modalidad a distancia en BUAP**

En el modelo pedagógico para la educación a distancia de la universidad, se especifica que la línea a seguir serán los principios educativos del modelo universitario minerva, el cual está centrado en el estudiante con un enfoque socio-constructivista donde se favorece el desarrollo de habilidades en el uso de las tecnologías de información y comunicación, la formación integral de los estudiantes y la construcción de conocimientos y experiencias.

---

Los principales elementos que conforman el modelo son los estudiantes, docentes, ambientes virtuales, un modelo de comunicación educativa, la comunidad en general y los servicios de apoyo, a continuación, se describe cada uno:

**Alumno.** Se establece que desarrollan sus saberes y competencias profesionales a través de ambientes virtuales diseñados para el aprendizaje autogestivo además de interactuar con profesores y otros compañeros favoreciendo el aprendizaje social.

**Profesor.** Facilitador del proceso de aprendizaje, que ayuda a que el estudiante construya su proceso de formación.

**Ambientes virtuales de aprendizaje.** Espacios virtuales donde aplican diferentes tecnologías educativas y que son diseñados para desarrollar saberes o competencias profesionales específicas a distintos niveles de granularidad. Los componentes de estos ambientes son: materiales educativos multimedia diseñados para el autoaprendizaje (lecturas, objetos de aprendizaje, referencias, sitios de interés), actividades de aprendizaje, evaluaciones y autoevaluaciones y herramientas para la gestión del curso. Sobre las actividades de aprendizaje, describen que en los ambientes virtuales de aprendizaje son de diferente índole, algunas son modeladas por el experto para promover directamente los saberes en los estudiantes, en otras se resaltan las actividades de forma individual o en grupos pequeños donde se abordan tareas que implican descubrimiento y solución de problemas abiertos. Estas actividades permiten la evaluación del grado de desarrollo del aprendizaje, y deben poner en juego diferentes capacidades intelectuales en un nivel de complejidad creciente. Estos ambientes de aprendizaje corresponden con los planes de estudios establecidos por la institución y se encuentran montados en sistemas gestores de aprendizaje como Blackboard o Moodle.

**Modelo de comunicación educativa.** Generalmente se desarrolla un ambiente virtual de aprendizaje para cada asignatura o módulo de un plan de estudios en el que se incluye un plan de clase y un cronograma que guían a los estudiantes sobre la distribución en el tiempo de los temas a revisar y de las actividades de aprendizaje a realizar. Las actividades de aprendizaje contienen instrucciones para su realización. Los facilitadores orientan el proceso de aprendizaje a lo largo del programa manteniendo una comunicación continua a través del correo electrónico; revisando sus trabajos, evaluando y retroalimentando los avances con apoyo del sistema gestor de aprendizaje.

**Comunidad.** Alumnos, profesores y personal administrativo, cambian su estilo de comunicación, logrando interactuar a distancia a través de diferentes herramientas tecnológicas como el sistema gestor de aprendizaje, herramientas de videoconferencia o correo electrónico.

**Servicios de apoyo.** Con el fin de atender las necesidades de los estudiantes se les provee de servicios de tecnología y comunicación, apoyo y soporte tecnológico, administración del sistema gestor de aprendizaje y servicios escolares.

### 3.2.2. Ingreso y matrícula de la modalidad a distancia en BUAP

De la revisión de los anuarios estadísticos de la institución se calculó que, en promedio, la cantidad de alumnos que ingresa a estudiar una licenciatura en modalidad

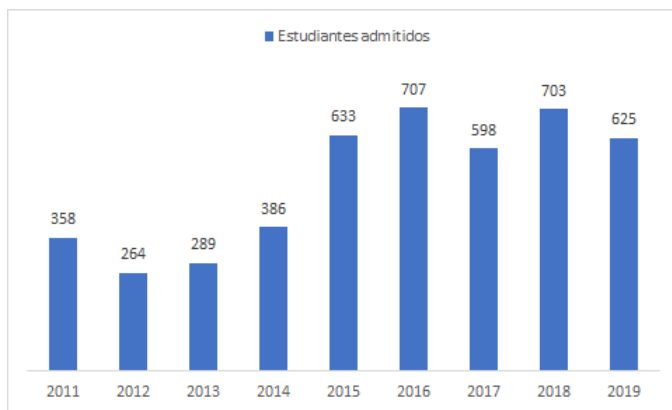


Figura 3.6: Estudiantes admitidos en modalidad a distancia.

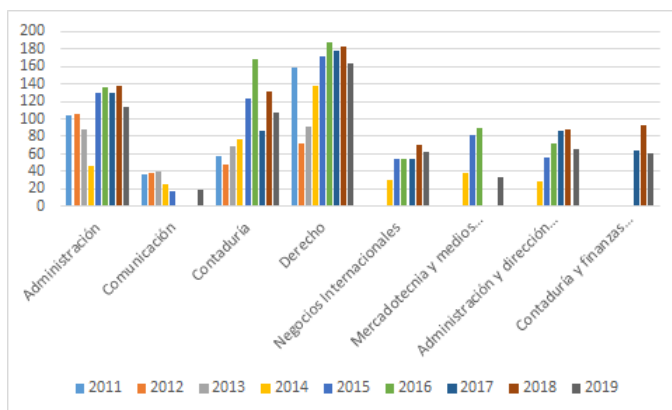


Figura 3.7: Estudiantes admitidos en modalidad a distancia por programa educativo.

a distancia es de 507 (BUAP, 2020; BUAP, 2019; BUAP, 2018; BUAP, 2017; BUAP, 2016) . En la Figura 3.6 se muestra el aumento del ingreso entre 2012 y 2016, también se destaca una reducción del ingreso después del 2017 que puede deberse a que en ese año la institución no ofertó dos licenciaturas (comunicación y mercadotecnia y medios digitales).

Las licenciaturas con mayor demanda en esta modalidad de estudio son derecho y contaduría seguidas por administración esto se puede deber a que fueron de las primeras licenciaturas en ofertarse ver Figura 3.7.

En el año 2019 la matrícula total de estudiantes era de 2016 (BUAP, 2019), de los cuales 60 % se encuentra en los programas de derecho, contaduría y administración, ver Figura 3.8.

### 3.2.3. Egreso de la modalidad a distancia en BUAP

La Tabla 3.5 muestra el número de egresados por programa educativo y periodos escolar. En el periodo 2016-2017 se muestra el signo *¿za* que no se pudo obtener la

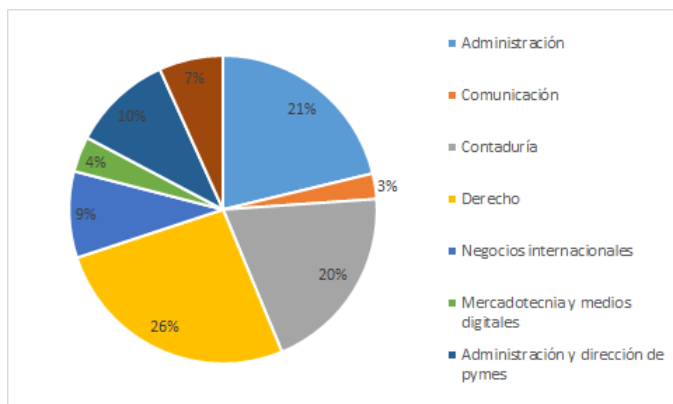


Figura 3.8: Estudiantes matriculados por programa educativo en 2019.

Tabla 3.5: Egresados por programa educativo

	2014-2015	2015-2016	2016-2017	2017-2018	2018-2019	2019-2020
Administración	-	3	?	13	20	16
Comunicación	1	6	?	6	10	4
Contaduría	-	6	?	12	12	12
Derecho	-	8	?	15	21	13
Negocios internacionales	-	-	-	-	-	2
Mercadotecnia y medios digitales	-	-	-	-	-	6
Administración y dirección de PyMES	-	-	-	-	4	-
Contaduría y finanzas públicas	-	-	-	-	-	-

información.

---

# Capítulo 4

## Metodología

En este capítulo, se presenta la metodología seguida para entrenar clasificadores multi-etiqueta que ubican automáticamente retroalimentaciones en los niveles tarea, proceso, regulación, elogios y otros considera cinco etapas que son: (1) recolección e integración de datos, (2) Preprocesamiento y extracción de características, (3) afinación de hiper-parámetros, (4) clasificación y (5) análisis y evaluación ver Figura 4.1.

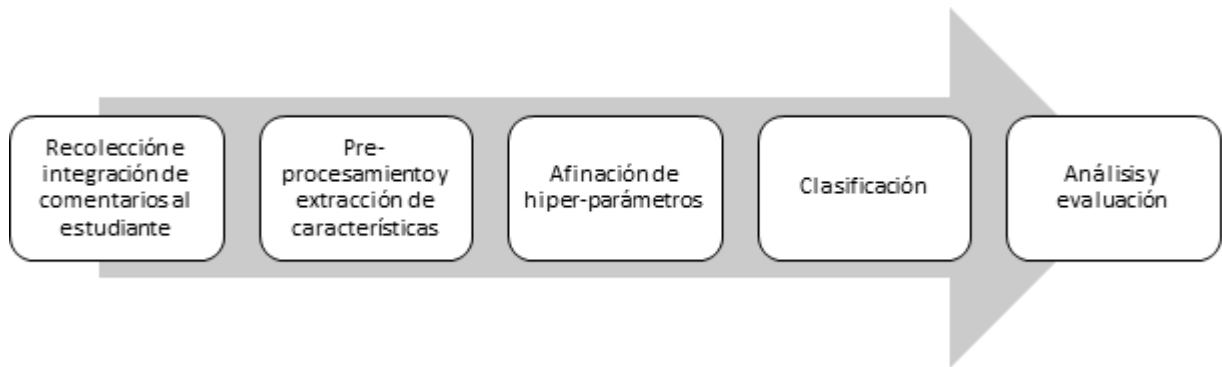


Figura 4.1: Metodología

La metodología se aplicó en un caso de estudio en el que se analizaron las retroalimentaciones generadas por docentes en cursos en línea de la facultad de derecho de una universidad pública de México. Se entrenaron 122 clasificadores multi-etiqueta utilizando los enfoques de transformación y adaptación los cuales están organizados en la estructura de la figura 4.2. A continuación, se describe cada una de las etapas de la metodología.

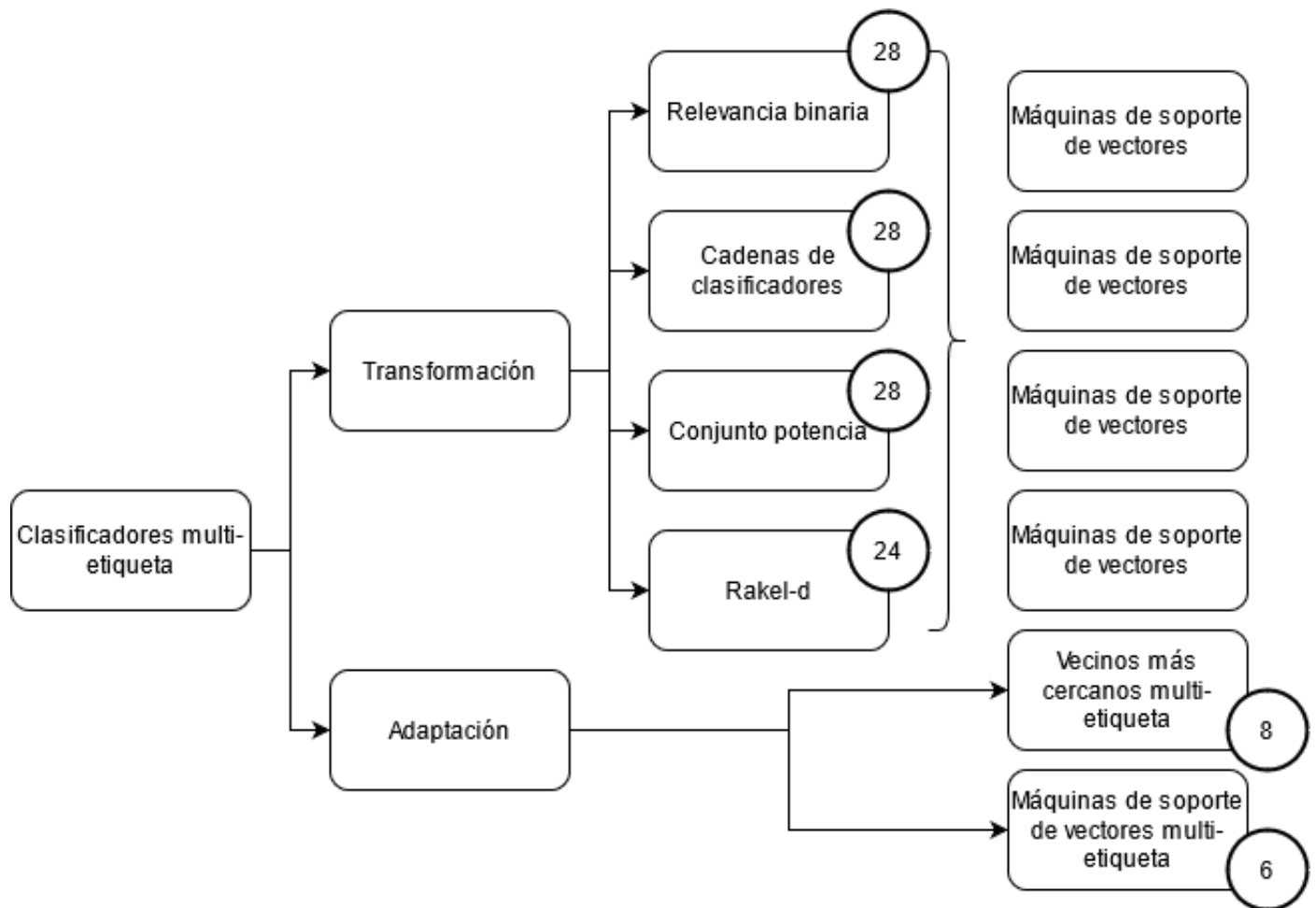


Figura 4.2: Estructura de los experimentos realizados

## 4.1. Recolección e integración de comentarios al estudiante

Los comentarios al estudiante utilizados en este trabajo se obtuvieron del sistema gestor de aprendizaje Blackboard a través del historial de calificaciones que es una herramienta que realiza el seguimiento de todos los cambios realizados en el centro de calificaciones el cual muestra una tabla de datos en los que se detallan todos los cambios de calificaciones realizados.

Se descargó el historial de calificación de 121 cursos en línea de la licenciatura en derecho ofertados en el periodo de otoño 2018 de una universidad pública de México. Cada historial se descargó en formato CSV con la opción todos para la condición "Mostrar entradas del pasado".

En el Anexo A se muestran los cursos en línea de los que se extrajeron los historiales



---

```
Estimado Edgar adiciono lo siguiente:
El contrato de promesa puede ser clasificado como:
Unilateral- bilateral
Oneroso-gratuito
Commutativo
Consensual
Formal
Instantáneo
La obligación que genera el contrato de promesa es solo la de HACER.
Código Civil del Estado de Puebla.
Artículo 2111 Son elementos esenciales de la promesa bilateral de contratar: I. El consentimiento; II. Que se expresen los elementos esenciales y 1
cláusulas que según la voluntad de los promitentes, contendrá el contrato, que se obligan a celebrar o contrato definitivo; III. Que se determine e
plazo en que habrá de otorgarse el contrato definitivo.
```

Figura 4.4: Comentario al estudiante después de eliminar etiquetas HTML/CSS

con experiencia en la docencia en línea procedió a leer cada uno de los comentarios e indicar si se ubicaba o no en los niveles tarea, proceso, regulación, elogios y otros. Se incorporó al documento maestro las columnas tarea, proceso, regulación, elogios y otros para que los expertos marcaran si cada comentario pertenecía al nivel o no. Para el etiquetado, se utilizó el modelo de Hattie y Timperley el cual se les fue explicado y se les proporcionó un protocolo para identificar el tipo de las retroalimentaciones. El protocolo de etiquetado se puede consultar en la tabla 4.1.

El resultado de la etapa de etiquetado manual por parte de los expertos fue un conjunto de datos multi-etiqueta.

#### 4.1.2. Análisis exploratorio de los comentarios al estudiante

El siguiente paso fue realizar un análisis exploratorio de los comentarios al estudiante con el fin de mejorar el entendimiento del problema que se tenía, utilizando técnicas estadísticas y representaciones gráficas para conocer los datos (retroalimentaciones).

Para el análisis de datos exploratorio de las retroalimentaciones, se realizó una limpieza de las retroalimentaciones que consistió en eliminar comentarios al estudiante que estuvieran vacíos o repetidos, se reemplazaron direcciones web, dígitos y nombres de archivos por identificadores únicos (URL, DP, AP) ya que para el estudio no era necesario conocer cuáles son solo si se menciona. Después, se pasaron los textos de cada comentario a minúsculas y se eliminaron espacios dobles entre palabras y signos de puntuación como ( , . ¡ ¿ ? ) a través de la librería re de Python que facilita el reemplazo de textos que cumplen con alguna expresión regular.

Posteriormente, se extrajeron los lemas de las palabras de cada comentario, se obtuvieron sus características que consisten en el número de caracteres, palabras de cada retroalimentación antes y después de aplicar la limpieza. También, fue de interés la extracción de características de partes del discurso y se identificaron los siguientes tipos de palabras: sustantivas, verbos y adjetivos. Por último, se realizó un análisis de las características a través de gráficas de barras y gráficas de caja.

Para los valores numéricos de cantidad de caracteres y palabras se calcularon las medidas básicas de dispersión, es decir: promedio, desviación estándar y el resumen de 5 números el cual se refiere al conjunto de estadística descriptiva que provee información sobre el conjunto de datos, consiste de los 5 más importantes percentiles.

Tabla 4.1: Protocolo de etiquetado de comentarios al estudiante.

Nivel tarea	Información sobre cuán bien una tarea fue completada o realizada.	Se distinguen respuestas correctas de incorrectas, indican que se debe obtener más o diferente información, se construye conocimiento superficial y se clarifican indicaciones
Nivel proceso	Información más específica a los procesos para la realización y ampliación de una tarea.	Incluye estrategias de los estudiantes para la detección de errores, señala la necesidad de plantear nuevas estrategias, elegir diferentes o ser más efectivo, aplicación de estrategias y/o buscar ayuda, pistas para apoyar a estudiantes a rechazar hipótesis erróneas, proporcionar ayudas de búsqueda y estrategias además de que mejora el conocimiento profundo.
Nivel regulación	Información que implica una interacción entre compromiso, control y confianza y que promueve autonomía, autocontrol, autodirección y autodisciplina.	Promueve pensamientos, sentimientos y acciones autogeneradas que son planeados y cíclicamente adaptados al logro de objetivos personales, la capacidad de crear retroalimentación interna y autoevaluación, voluntad para invertir esfuerzo en búsqueda y tratamiento de la retroalimentación, el grado de confianza o certeza en la veracidad de la respuesta y el nivel de competencia para buscar ayuda
Nivel de elogios	Información que expresa principalmente evaluaciones positivas (a veces negativas) y afecto hacia los estudiantes.	Elogios
Nivel otro	Información que no está relacionada con la tarea o características de los estudiantes.	Información descontextualizada.

Tabla 4.2: Frecuencias de caracteres y palabras por comentario al estudiante

	Promedio	Desviación estándar	Mínimo	25 %	50 %	75 %	Máximo
Caracteres	502.20	944.53	1	28	167	360	6586
No_palabras	77.18	150.29	0	11	24	54	1082
ComentarioL_len	459.04	893.25	0	66	143	316	6046
ConteoL_palabra	75.99	148.78	0	10	23	52	1074
ComentarioLemma_len	459.04	893.25	0	66	143	316	6046
ConteoLemma_palabra	75.99	148.78	0	10	23	52	1074

En la tabla 4.2 se observa que en promedio los comentarios sin limpieza tienen 502 caracteres y 77 palabras respectivamente. Cuando se les aplica limpieza y lemas el promedio es de 459 para caracteres y 75 para las palabras.

Se observa que el tamaño de los comentarios medidos por número de caracteres y palabras después de la limpieza se reduce. Se observa también que no hay cambios entre un comentario con lemas y uno solo con la limpieza.

El resumen de 5 números revela que para los comentarios al estudiante sin limpieza el número de caracteres está entre 1 y 6586 mientras que para los comentarios limpios y con lemas se encuentra entre 0 y 6046. Sin embargo, 75 % de estos tienen menos de 360 caracteres.

Respecto a número de palabras en el comentario sin limpieza se tiene un rango que va de 0 a 1082 palabras en comparación con los limpios y con lemas que están entre 0 y 1074. Se observa que el 75 % tiene menos de 54 palabras.

En las figuras 4.5 y 4.6, se muestran las gráficas de caja del número de caracteres y palabras para los comentarios sin limpieza el cual permite observar que la mayor concentración de comentarios se encuentra en el rango de 120 y 2300 caracteres mientras que para las palabras se encuentran entre 20 y 380.

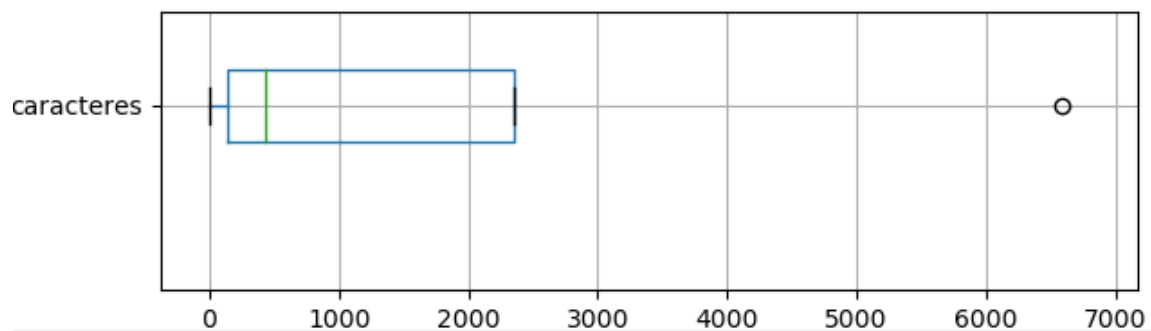


Figura 4.5: Gráfica de caja de cantidad de caracteres en comentario al estudiante

Por su parte para los comentarios con lemas y limpios en las figuras 4.7 y 4.8 se observa que la mayoría de los comentario se encuentra en el rango de 131 y 2100 para caracteres y 19 y 380 para palabras.

Sobre el etiquetado manual realizado por los expertos, en la Tabla 4.3 se muestra la

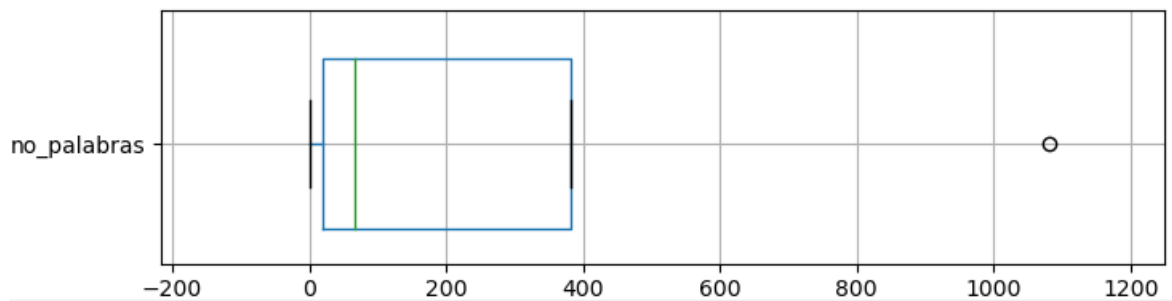


Figura 4.6: Gráfica de caja de cantidad de palabras en comentario al estudiante

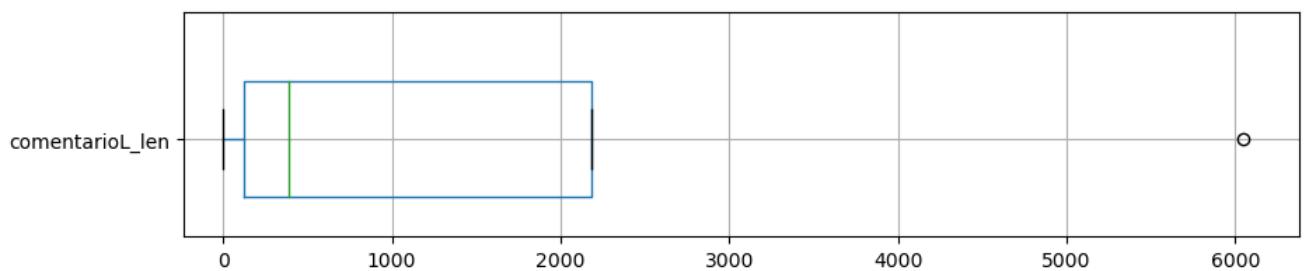


Figura 4.7: Gráfica de caja de cantidad de caracteres en comentario\_limpio

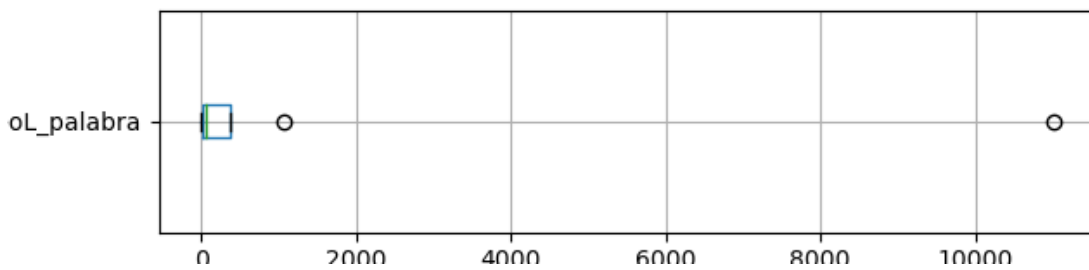


Figura 4.8: Gráfica de caja de cantidad de palabras en Comentario\_limpio

cantidad de comentarios al estudiante que fueron ubicados en los niveles tarea, proceso, regulación, elogios y otros. Por su parte, la Figura 4.9 muestra de manera gráfica la distribución de comentarios etiquetados por cada nivel, se destaca un desbalance en la frecuencia de comentarios que tienen o no cada uno de los niveles, este es mayor para el nivel de regulación.

De la exploración de frecuencia de palabras se obtuvo que el conjunto de comentarios al estudiante contenía un total de 14421 palabras diferentes, las Figuras 4.10 y 4.11

Tabla 4.3: Frecuencias de comentarios al estudiante por niveles tarea, proceso, regulación, elogios y otros

Clase	Tarea	Proceso	Regulación	Elogios	Otros
1	4216	7133	10960	6746	9720
0	6807	3890	63	4277	1303
Total	11023	11023	11023	11023	11023

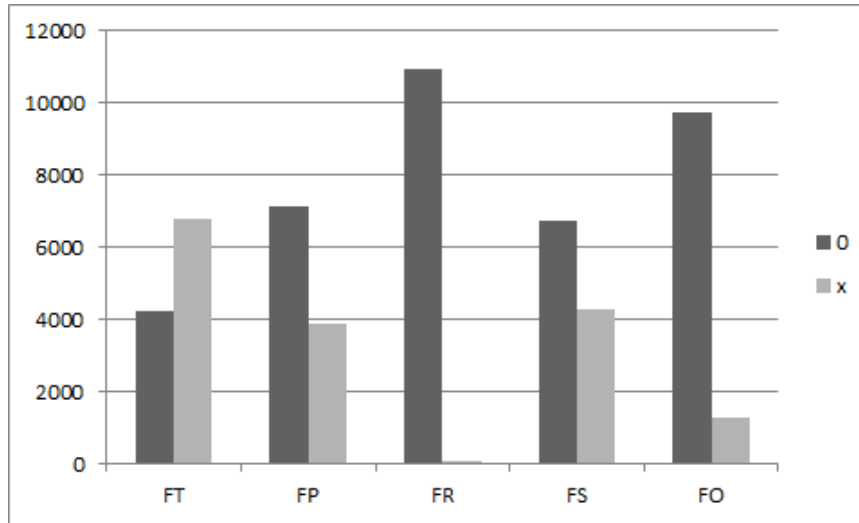


Figura 4.9: Distribución de comentarios por nivel

muestran las palabras de mayor y menor frecuencia en las que se visualiza que las primeras están relacionadas con las palabras que detienen (stop words) como son "de", "la", "que", ... mientras que las segundas son por lo general errores ortográficos.

A partir del análisis exploratorio de retroalimentaciones se determinó que era necesario aplicar un corrector ortográfico y buscar una manera de reducir el número de palabras en el conjunto de datos ya que sería las que se utilizarían como características para la futura fase de clasificación en la que a menor número de características menor costo computacional.

## 4.2. Preprocesamiento y extracción de características

Siguiendo las recomendaciones de Herrera et al (2016) para el tratamiento previo que deben tener los textos antes de pasarlos a un algoritmo de clasificación, se procesaron las retroalimentaciones del conjunto de datos multi-etiqueta.

A cada comentario además de la limpieza y su paso a minúsculas mencionados

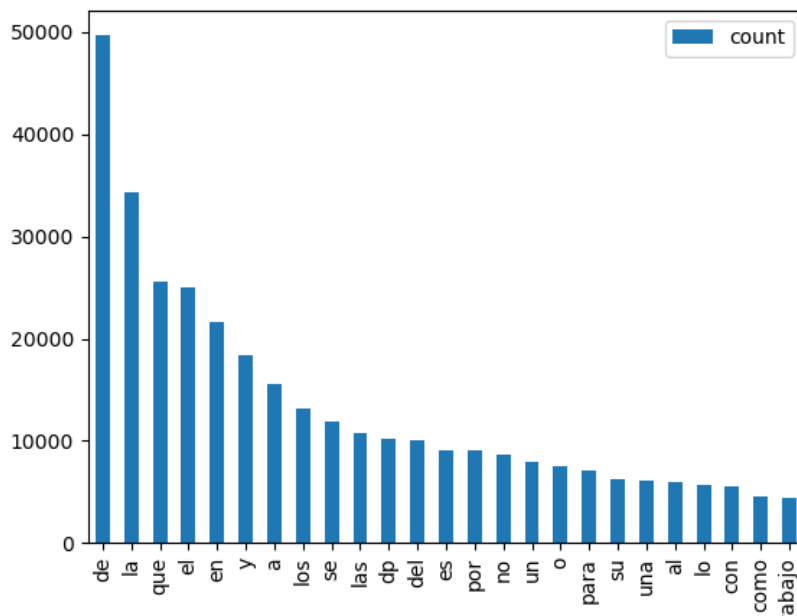


Figura 4.10: Palabras de mayor frecuencia en comentarios al estudiante

previamente, se aplicó la lematización que consiste el análisis morfológico de las palabras para agrupar las diversas formas de inflexión de una palabra para que puedan analizarse como un solo elemento a través de la librería de Python “spaCY” que proporciona herramientas para el procesamiento del lenguaje natural.

También, se implementó un corrector ortográfico que utiliza la distancia de Levenshtein que mide la distancia entre dos palabras, mientras mayor sea, mayor será la diferencia, de manera que buscará en un diccionario las que tengan la menor y las mostrará como candidatas de reemplazo. El diccionario de palabras utilizado para el corrector fue el libro en español de las “21 leyes de liderazgo” de John C. Maxwell. Las palabras que pasaron por el corrector ortográfico fueron las que tenían frecuencias menores a 10 al identificarse en la etapa de análisis exploratorio de retroalimentaciones con errores ortográficos. Después a cada comentario al estudiante se le eliminaron las palabras que detienen que se refiere a aquellas que aparecen con mayor frecuencia en las retroalimentaciones sin tener mucha información sobre el contenido como pueden ser las preposiciones, para este trabajo las que se eliminaron fueron: “de”, “lo”, “el”, “que”, “en”, “y”, “a”, “ser”, “uno”, “se”, “los”, “dp”, “del”, “su”, “por”, “o”, “parir”, “al”, “poder”, “derecho”, “tu”, “este” y “la”. Posteriormente se eliminaron los signos diacríticos y palabras de baja frecuencia de cada comentario al estudiante. Un signo diacrítico es cualquier marca que se utiliza para modificar algún otro signo de la escritura como son los acentos y las diéresis. Las palabras de baja frecuencia (menor a 10) fueron eliminadas al aportar poca información para diferenciar las retroalimentaciones.

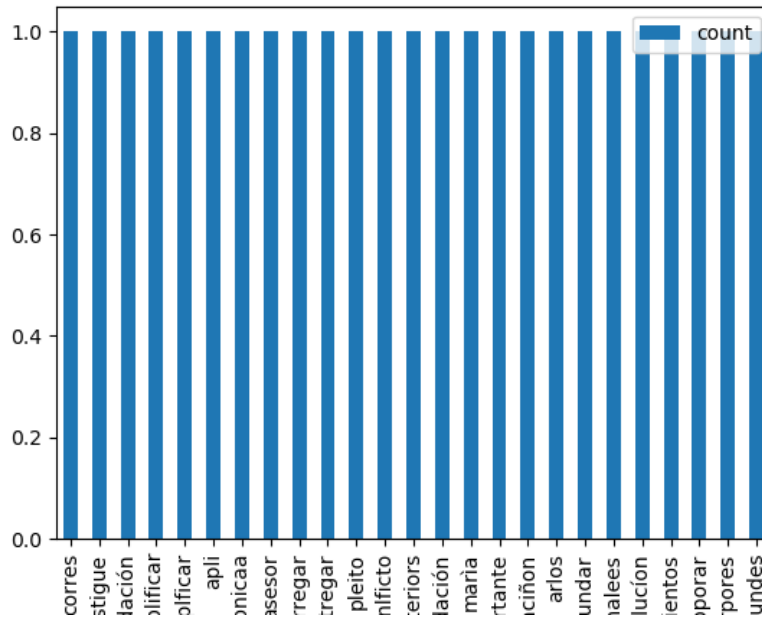


Figura 4.11: Palabras de menor frecuencia en comentarios al estudiante

Por último, se aplicó truncamiento a todas las palabras contenidas en el conjunto de datos multi-etiqueta el cual consiste en obtener la raíz de la palabra derivada a través de la librería de Python “Snowball Stemmer” que implementa el algoritmo de Porter para identificar la raíz.

En la tabla 4.4 se muestran la cantidad de palabras existentes en el conjunto de datos multi-etiqueta después de la aplicación de las técnicas de procesamiento. Se muestran la cantidad de palabras antes y después de aplicar el método de procesamiento (f), la cantidad de etiquetas que tiene el conjunto de datos (k), la cantidad de combinaciones de etiquetas y el valor TCS referente a la complejidad del conjunto de datos. Se observa la reducción de palabras después de aplicar cada uno de los métodos de preprocesamiento. También, se puede ver que el preprocesamiento permite tener una menor complejidad del conjunto de datos multi-etiqueta.

#### 4.2.1. Conjuntos de entrenamiento y prueba

Una vez preprocesadas todos los comentarios al estudiante se procedió a particionar el conjunto de datos multi-etiqueta en dos subconjuntos uno para entrenamiento de los clasificadores multi-etiqueta y otro para probar cuán bien realizan la tarea de ubicar nuevas retroalimentaciones en los niveles tarea, proceso, regulación, elogios y otros con respecto a lo etiquetado por los experots. Los comentarios al estudiante de ambos conjuntos fueron seleccionadas de manera aleatoria quedando el 66 % para el

Tabla 4.4: Cantidad de palabras únicas en conjunto de entrenamiento después de aplicar etapas de preprocesamiento

	Antes del proceso	Después del proceso (f)	No. de etiquetas (k)	No. conjuntos de etiquetas	Complejidad (TCS)
Documento original		18677	5	22	14.53552847
Normalización	18677	18062	5	22	14.50204593
Lowercase	18677	14939	5	22	14.31221089
Lematización	18677	14143	5	22	14.25745545
Lematización con (lowercase)	18677	11395	5	22	14.04141031
Normalización, lematización y lowercase	11395	10889	5	22	13.99598875
Ortografía	10889	8537	5	22	13.7526453
Ortografía y stop words	8537	8515	5	22	13.75006496
Ortografía, stop words, signos diacríticos, frecuencias	8515	3975	5	22	12.98826039
Stem	3905	2886	5	22	12.66810711

Tabla 4.5: Distribución de retroalimentaciones en conjuntos multi-etiqueta de entrenamiento y prueba

Clase \ Nivel	Conjunto Entrenamiento					Conjunto Prueba				
	Tarea	Proceso	Regulación	Elogios	Otros	Tarea	Proceso	Regulación	Elogios	Otros
Clase 0	4746	2742	37	3015	906	2066	1153	26	1263	389
Clase 1	2963	4967	7669	4694	6803	1238	2151	3278	2041	2915
Total	7709	7709	7709	7709	7709	3304	3304	3304	3304	3304

entrenamiento y el 34 % restante para prueba. La Tabla 4.5 muestra la distribución de las retroalimentaciones en cada conjunto

Una vez que se tuvieron los dos subconjuntos, se procedió a analizar las características de cada uno. El resultado del análisis se muestra en la Tabla 4.6, recordando que la cardinalidad cuenta el número de etiquetas promedio relevantes para cada instancia en el conjunto de datos, la densidad se refiere a la cardinalidad normalizada por el número total de todas las posibles etiquetas, MeanIR obtiene la proporción máxima de desbalance, es decir, la proporción de la etiqueta más común contra la más rara y el valor SCUMBLE mide la concurrencia entre etiquetas frecuentes y raras.

El siguiente paso consistió en la extracción de características, se utilizaron usaron diferentes n-gramas como unidades de separación de textos que incluyen un-grama, dos-grama, tres-grama, uno-dos-tres-grama, uno-dos-grama y dos-tres-grama para el

Tabla 4.6: Características de los conjuntos de datos multi-etiqueta

Métrica	Conjunto Entrenamiento	Conjunto Prueba
No. Instancias	7709	3304
No. Etiquetas	5	5
No. De conjuntos de sub-etiquetas	23	21
Cardinalidad	1.484	1.482
Densidad	0.296	0.296
meanIR	27.562	17.84
Scumble	0.021	0.026

Tabla 4.7: Hiper-parámetros y espacio de búsqueda.

Algoritmo	Parámetros explorados	Espacio de búsqueda
Máquinas de soporte de vectores	Parámetro de regularización C	[0.1, 1, 10, 100]
Bosques aleatorios	Número de árboles en el bosque	[20,50, 100, 200, 400, 500, 800]
K-vecinos más cercanos multi-etiqueta	Número de vecinos a tomar en cuenta.	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]
	Parámetro de suavizado s	[0.1, 1.0, 10, 100, 200, 500]

conjunto de entrenamiento y prueba.

Por cada n-grama, se transformaron los comentarios al estudiante en una matriz término-documento representando la frecuencia de cada palabra en cada comentario. Se calculó el valor TF-IDF que combina la frecuencia de término y la frecuencia inversa de los comentarios a través de multiplicar el peso de frecuencia local de cada comentario por el peso inverso del mismo.

### 4.3. Afinación de hiper-parámetros

La etapa de afinación de hiper-parámetros tiene la finalidad de buscar los mejores parámetros para un algoritmo de clasificación. Los mejores parámetros serán los que permitan un mejor valor para una métrica.

Para realizar la búsqueda de hiper-parámetros se seleccionaron los algoritmos a afinar, el espacio de parámetros de búsqueda, el método de búsqueda y selección de candidatos, el esquema de validación cruzada y métrica de desempeño a evaluar.

Los algoritmos considerados en este trabajo fueron las máquinas de vectores de soporte, bosques aleatorios y vecinos más cercanos multi-etiqueta al ser los que mejor desempeño lograron en la primera etapa de clasificación en la tarea de ubicar los comentarios al estudiante en los niveles tarea, proceso, regulación, elogios y otros con respecto a lo etiquetado manualmente por los expertos.

Los hiper-parámetros considerados para cada algoritmo y el espacio de parámetros utilizado se muestra en la Tabla 4.7. Se explora el parámetro de regularización  $c$  y número de árboles que puede existir en el bosque para los algoritmos de máquinas de soporte de vectores y bosques aleatorios respectivamente. Para el algoritmo de vecinos más cercanos multi-etiqueta se exploran dos hiper parámetros que son el número de vecinos a tomar en cuenta y el parámetro de suavizado  $s$ .

Se utiliza la búsqueda de malla descrita en la sección 2.8.1 al ser la de menor costo computacional. Para evitar el sobre ajuste, se utiliza validación cruzada a 5 capas sobre el conjunto de entrenamiento.

La métrica de desempeño a optimizar fue la exactitud al ser la más exigente cuando se refiere a clasificación multi-etiqueta.

Al finalizar esta etapa se obtienen los hiper-parámetros que permiten una mejor clasificación multi-etiqueta de los comentarios al estudiante del conjunto de prueba, los cuales serán utilizados en la etapa de clasificación.

Tabla 4.8: Clasificadores multi-etiqueta entrenados en primera etapa

Enfoque	Método	Algoritmo	N-gramas	Total
Transformación	Relevancia binaria	Bosques aleatorios	Un-grama Dos-grama Tres-grama Un-dos-grama Dos-tres-grama Uno-dos-tres-grama	6
		Máquinas de soporte de vectores		6
		Naive bayes		6
		Regresión logística		6
	Cadenas de clasificadores	Bosques aleatorios		6
		Máquinas de soporte de vectores		6
		Naive bayes		6
		Regresión logística		6
	Conjunto potencia	Bosques aleatorios		6
		Máquinas de soporte de vectores		6
		Naive bayes		6
		Regresión logística		6
	Rakel-d	Bosques aleatorios		6
		Máquinas de soporte de vectores		6
		Naive bayes		6
		Regresión logística		6
Adaptación	-	Máquinas de soporte de vectores multi-etiqueta	6	
	-	Vecinos más cercanos multi-etiqueta	6	
<b>Total de clasificadores</b>				<b>108</b>

## 4.4. Clasificación

Esta etapa, consistió en aplicar los algoritmos de minería de textos para entrenar clasificadores multi-etiqueta que aprendieran del conjunto de entrenamiento que contiene la clasificación manual de las retroalimentaciones realizada por los expertos.

A cada algoritmo se le pasó las matrices término-documento descritas en la etapa anterior y las etiquetas que realizaron los expertos con el fin de obtener un clasificador multi-etiqueta. El entrenamiento de los clasificadores se realizó en Python, se utilizó la librería scikit-multilearn que está basada en el ecosistema scikit-learn que cuenta con los algoritmos para llevar a cabo la clasificación multi-etiqueta.

En una etapa inicial, se entrenaron 108 clasificadores multi-etiqueta ver Tabla 4.8 (sin considerar la etapa de afinación de hiper-parámetros) para realizar los estudios de selección de enfoque, algoritmo, método de transformación, características y evaluación de afinación de hiper-parámetros.

De los experimentos de la etapa inicial, 96 se realizaron con el enfoque de transformación y 12 con el enfoque de adaptación. Se utilizaron los algoritmos de clasificación de máquinas de soporte de vectores, bosques aleatorios, regresión logística, naive bayes, vecinos más cercanos multi-etiqueta y máquinas de soporte de vectores multi-etiqueta .

Los experimentos del enfoque de transformación se probaron con cuatro métodos de transformación: relevancia binaria, conjunto potencia, cadenas de clasificadores y rakel-d descritos en el marco teórico.

Se utilizaron 6 diferentes n-gramas para la representación de las retroalimentaciones en matrices término-documento y se calculando la frecuencia de término- frecuencia inversa de documento.

En la segunda etapa se entrenaron 14 clasificadores multi-etiqueta considerando los resultado de la etapa previa, para responder el quinto cuestionamiento ¿La afinación

Tabla 4.9: Clasificadores multi-etiqueta entrenados en etapa 2

Enfoque	Método	Algoritmo	N-gramas	Hiper-parámetros	Total
Transformación	Relevancia binaria	Bosques aleatorios	Un grama	Por defecto	2
		Máquinas de soporte de vectores			2
	Cadenas de clasificadores	Bosques aleatorios			2
		Máquinas de soporte de vectores			2
	Conjunto potencia	Bosques aleatorios		2	
		Máquinas de soporte de vectores		2	
Adaptación	-	Vecinos más cercano multi-etiqueta			2
<b>Total de clasificadores multi-etiqueta</b>					<b>14</b>

de hiper-parámetros permite mejorar el desempeño de clasificadores multi-etiqueta? ver Tabla 4.9. De estos experimentos, 12 utilizaron el enfoque de transformación y dos con el enfoque de adaptación. Se utilizaron solamente los algoritmos de bosques aleatorios, máquinas de soporte de vectores y vecinos más cercanos multi-etiqueta al ser los que mejor desempeño evidenciaron en la etapa 1.

Los métodos de transformación utilizados fueron relevancia binaria, cadenas de clasificadores y conjunto potencia. Para la representación de cada retroalimentación en matriz término documento se utilizó la bolsa palabras (un-grama) al evidenciar ser el mejor en desempeño y tiempo de entrenamiento.

A diferencia de la etapa 1, previo a la etapa de clasificación propuesta en la metodología para el entrenamiento de los clasificadores multi-etiqueta, se aplica una etapa de afinación de hiper-parámetros que devuelve los hiper-parámetros que permiten tener un mejor desempeño. Después en la etapa de clasificación se realizan los experimentos utilizando los parámetros por defecto de cada algoritmo y los encontrados.

## 4.5. Análisis y Evaluación

Para evaluar el desempeño de los clasificadores multi-etiqueta se proporcionó a cada uno los comentarios al estudiante del conjunto de prueba sin las clasificaciones realizadas por los diseñadores instruccionales para que clasificaran automáticamente cada una conforme los niveles tarea, proceso, regulación, elogios y otro propuestos de Hattie y Timperley. Después, se compararon las clasificaciones automáticas realizadas por los clasificadores multi-etiqueta con las clasificadas por los diseñadores instruccionales a través de las métricas de clasificación multi-etiqueta de pérdida de Hamming, promedio ponderado de la precisión y exhaustividad a nivel micro (f1), exactitud, pérdida de ordenamiento, media de precisión, cobertura y tiempo en segundos en llevar a cabo el entrenamiento, descritas en el marco teórico.

Para el análisis se realizaron estudios en los que se compararon los clasificadores multi-etiqueta obtenidos a través de gráficas radiales que permiten comparar varias métricas de desempeño y visualizar qué clasificadores multi-etiqueta tienen desempeños similares.

Se realizaron seis estudios con el fin de determinar el enfoque de clasificación multi-etiqueta, los algoritmos de clasificación, métodos de transformación, características que

---

mejor desempeño permiten en la clasificación automática de los comentarios al estudiante en los niveles tarea, proceso, regulación, elogios y otros. El último estudio se realizó con el fin de detectar si la aplicación de una fase de afinación de hiper-parámetros permite la mejora en el desempeño de los clasificadores multi-etiqueta.

A continuación, para cada estudio se da una breve introducción del problema a tratar y los procedimientos seguidos

#### **4.5.1. Selección del enfoque de clasificación multi-etiqueta**

Cómo se describió en la sección 2.7.2 la clasificación multi-etiqueta se aborda desde dos enfoques, transformación y adaptación por lo que se analizó cuál de los permite entrenar clasificadores multi-etiqueta con mejor desempeño en la tarea de ubicar los comentarios al estudiante en los niveles tarea, proceso, regulación, elogios y otros con respecto a lo que realizaron los expertos manualmente.

Se analizó el desempeño de los clasificadores multi-etiqueta entrenados de las etapas 1 y 2, comparando el desempeño de los que fueron entrenados con el enfoque de transformación contra los que utilizaron el enfoque de adaptación. Para determinar el mejor se utilizaron gráficas radiales para visualizar los valores alcanzados en las métricas de desempeño de pérdida de Hamming, f1 a nivel micro, exactitud, pérdida de ordenamiento y media de precisión.

En estos gráficos a cada métrica de desempeño se le proporciona un eje que empieza en el centro, todos los ejes se disponen radialmente, con distancias iguales entre sí, manteniendo la misma escala entre todos los ejes, cada valor de variable se traza a lo largo de su eje individual y todas las variables en un conjunto de datos se conectan para formar un polígono. Para este trabajo, el polígono con mayor área es indicativo de tener un mejor desempeño.

#### **4.5.2. Selección de algoritmos de clasificación**

En la literatura, se proponen diferentes algoritmos para clasificar textos, entre los más populares para el enfoque de transformación se encuentran naive bayes, máquinas de soporte de vectores, bosques aleatorios y regresión logística que son probados durante la primera etapa de clasificación.

En este estudio, se hace una comparación de los algoritmos previamente mencionados a través de analizar los desempeños de cada clasificador multi-etiqueta identificando los que mejor realizan la tarea de ubicar automáticamente los comentarios al estudiante en los niveles tarea, proceso, regulación, elogios, y otros con respecto a lo etiquetado manualmente por los expertos.

De igual forma, se estudian los algoritmos que utilizan el enfoque de adaptación de máquinas de soporte de vectores multi-etiqueta y vecinos más cercanos multi-etiqueta propuesto por Zhang y Zhou et al. (2007). Para determinar el mejor algoritmo, se comparan las métricas de desempeño de pérdida de Hamming, f1 a nivel micro, exactitud, pérdida de ordenamiento y media de precisión a través de gráficas radiales.

---

### 4.5.3. Selección de características

Uno de los pasos en el análisis de textos tiene que ver con la unidad de texto a analizar y la separación del texto basado en la unidad de análisis. Esta puede ser una palabra o conjuntos de palabras, las utilizadas en este trabajo fueron los n-gramas siguientes: un-grama, dos-grama, tres-grama, uno-dos-tres-grama, uno-dos-grama y dos-tres-grama.

En este estudio se hace una comparación de las diferentes características o n-gramas mencionados previamente a través de analizar el desempeño de clasificadores multi-etiqueta para identificar los que permiten tener un mejor desempeño en la tarea de ubicar automáticamente los comentarios al estudiante en los niveles tarea, proceso, regulación, elogios y otros con respecto a lo etiquetado manualmente por los expertos.

Se utilizaron los experimentos de la primera etapa de clasificación y se compararon las métricas de desempeño, para determinar el mejor n-grama, se comparan las métricas de desempeño de pérdida de Hamming, f1 a nivel micro, exactitud, pérdida de ordenamiento y media de precisión a través de gráficas radiales.

También, se hace un análisis del tiempo que se requiere para el entrenamiento de cada clasificador multi-etiqueta a través de gráfica de barras para determinar la mejor de las características ya que existen n-gramas con un desempeño similar.

### 4.5.4. Evaluación de la afinación de hiper-parámetros

Para determinar si la afinación de hiper-parámetros mejora el desempeño de clasificadores multi-etiqueta en la tarea de ubicar comentarios al estudiante en los niveles tarea, proceso, regulación, elogios y otros con respecto a lo etiquetado manualmente por los expertos, se utilizaron los experimentos de la etapa 2.

Se comparan las métricas de desempeño de pérdida de Hamming, f1 a nivel micro, exactitud, pérdida de ordenamiento, media de precisión y tiempo.

# Capítulo 5

## Resultados

Este capítulo consta de seis secciones, cada una de ellas correspondiente a los resultados obtenidos a partir de los experimentos planteados en el capítulo anterior, con sus respectivos análisis y conclusiones preliminares. En la primera sección se muestran los resultados obtenidos del entrenamiento de 108 clasificadores multi-etiqueta utilizando los enfoques transformación y adaptación con diferentes métodos de transformación y n-gramas. En la segunda sección se hace un estudio comparativo de los enfoques de clasificación multi-etiqueta adaptación y transformación. En la tercera se hace una comparación de los algoritmos de clasificación y se determina cuáles son los que permiten realizar mejor la tarea propuesta. En la cuarta se hace un análisis de diferentes n-gramas para determinar los que permiten un mejor desempeño de los clasificadores multi-etiqueta. En la quinta sección se presenta un estudio de los diferentes métodos de transformación. Por último en la sexta sección se determina si la afinación de hiperparámetros permite mejorar el desempeño de clasificadores multi-etiqueta.

### 5.1. Resultados de los experimentos

Las Tablas 5.1, 5.2, 5.3, 5.4, 5.5 y 5.6 muestran los desempeños alcanzados por los clasificadores multi-etiqueta entrenados en la primera etapa de clasificación. Se muestran las métricas de clasificación multi-etiqueta pérdida de Hamming,  $f_1$ (micro), exactitud, pérdida de ordenamiento, cobertura, promedio de precisión y tiempo de entrenamiento obtenidas por cada clasificador al utilizar un-gram, dos-gram, tres-gram, uno-dos-tres-gram, dos-tres-gram y uno-dos-gram. Las flechas en las métricas indican si es mejor un valor alto o bajo.

Para la métrica pérdida de Hamming que mide el porcentaje de etiquetas que son predichas incorrectamente se observa que la mejor clasificación se logra al utilizar las características un-gram, uno-dos-tres-gram y uno-dos-gram. El enfoque de transformación de datos que mejor clasifica es relevancia binaria seguido de cadenas de clasificadores y raket-d con los algoritmos base máquinas de soporte de vectores y bosques aleatorios. Los clasificadores mencionados logran clasificar bien 92 de cada 100 etiquetas. Los al-

Tabla 5.1: Métricas de desempeño de clasificadores utilizando un-gram

Enfoque	Método	Algoritmo	Perdida de Hamming ↓	F1 (micro) ↑	Exactitud ↑	Perdida de ordenamiento ↓	Cobertura ↓	Media de precisión ↑	Tiempo (Segundos) ↓	
Transformación	Relevancia binaria	Máquinas de soporte de vectores	<b>0.08</b>	<b>0.858</b>	0.695	0.181	2.293	0.589	36	
		Bosques aleatorios	0.082	0.855	0.7	0.178	2.277	<b>0.6</b>	27	
		Regresión logística	0.096	0.832	0.631	0.216	2.439	0.554	5	
		Naive bayes	0.126	0.776	0.549	0.271	2.696	0.479	<b>0</b>	
	Cadenas de clasificadores	Máquinas de soporte de vectores	0.083	0.852	0.707	0.175	2.295	2.295	0.59	34
		Bosques aleatorios	0.084	0.852	<b>0.713</b>	0.173	2.271	2.271	0.599	25
		Regresión logística	0.097	0.829	0.661	0.201	2.402	2.402	0.562	7
		Naive bayes	0.127	0.77	0.556	0.276	2.728	2.728	0.478	<b>0</b>
		Máquinas de soporte de vectores	0.09	0.84	0.698	0.188	2.336	2.336	0.576	24
		Bosques aleatorios	0.092	0.837	0.696	0.192	2.346	2.346	0.581	13
Conjunto potencia	Regresión logística	0.095	0.832	0.677	0.195	2.376	2.376	0.565	11	
	Naive bayes	0.125	0.772	0.587	0.267	2.692	2.692	0.487	<b>0</b>	
	Máquinas de soporte de vectores	0.085	0.852	0.7	<b>0.17</b>	<b>2.264</b>	<b>2.264</b>	0.588	199	
	Bosques aleatorios	0.092	0.837	0.693	0.194	2.347	2.347	0.587	21	
Rakel-d	Regresión logística	0.093	0.838	0.668	0.187	2.336	2.336	0.567	11	
	Naive bayes	0.129	0.768	0.561	0.272	2.715	2.715	0.481	<b>0</b>	
	Vécinos más cercanos multi-etiqueta	0.156	0.735	0.55	0.303	2.671	2.671	0.486	144	
Adaptación	-	Máquinas de soporte de vectores multi-etiqueta	0.228	0.433	0.181	0.729	4.24	0.394	110	

Tabla 5.2: Métricas de desempeño de clasificadores utilizando dos-gram

Enfoque	Método	Algoritmo	Perdida de Hamming ↓	F1 (micro) ↑	Exactitud ↑	Perdida de ordenamiento ↓	Cobertura ↓	Media de precisión ↑	Tiempo (Segundos) ↓	
Transformación	Relevancia binaria	Máquinas de soporte de vectores	0.115	0.789	0.57	0.302	2.776	0.507	15	
		Bosques aleatorios	<b>0.109</b>	<b>0.801</b>	0.587	0.29	2.718	<b>0.524</b>	35	
		Regresión logística	0.126	0.775	0.574	0.28	2.721	0.485	9	
		Naive bayes	0.146	0.741	0.532	0.315	2.849	0.453	<b>0</b>	
	Cadenas de clasificadores	Máquinas de soporte de vectores	0.128	0.773	0.598	0.276	2.695	0.502	0.502	18
		Bosques aleatorios	0.123	0.784	0.623	0.263	2.623	0.519	0.519	37
		Regresión logística	0.127	0.773	0.586	0.275	2.713	0.484	0.484	13
		Naive bayes	0.147	0.736	0.531	0.322	2.888	0.452	0.452	<b>0</b>
		Máquinas de soporte de vectores	0.122	0.783	0.614	0.261	2.632	0.493	0.493	12
		Bosques aleatorios	0.119	0.792	0.628	0.251	2.583	0.506	0.506	17
Conjunto potencia	Regresión logística	0.127	0.773	0.597	0.272	2.688	0.485	0.485	19	
	Naive bayes	0.14	0.746	0.558	0.302	2.825	0.458	0.458	<b>0</b>	
	Máquinas de soporte de vectores	0.119	0.782	0.57	0.303	2.79	0.504	0.504	424	
	Bosques aleatorios	0.116	0.797	<b>0.63</b>	<b>0.245</b>	<b>2.557</b>	0.512	0.512	17	
Rakel-d	Regresión logística	0.123	0.781	0.589	0.265	2.66	0.492	0.492	22	
	Naive bayes	0.146	0.742	0.538	0.301	2.796	0.448	0.448	<b>0</b>	
Adaptación	-	Vecinos más cercanos multi-etiqueta	0.141	0.738	0.517	0.356	2.966	0.474	213	
	-	Máquinas de soporte de vectores multi-etiqueta	0.249	0.319	0.101	0.845	4.605	0.367	265	

Tabla 5.3: Métricas de desempeño de clasificadores utilizando tres-gram

Enfoque	Método	Algoritmo	Perdida de Hamming ↓	F1 (micro) ↑	Exactitud ↑	Perdida de ordenamiento ↓	Cobertura ↓	Media de precisión ↑	Tiempo (Segundos) ↓	
Transformación	Relevancia binaria	Máquinas de soporte de vectores	0.177	0.674	0.466	0.399	3.205	<b>0.405</b>	6	
		Bosques aleatorios	0.177	0.677	0.472	0.396	3.183	0.403	12	
		Regresión logística	0.18	0.669	0.454	0.402	3.222	0.4	6	
		Naive bayes	0.195	0.645	0.426	0.427	3.306	0.384	<b>0</b>	
	Cadenas de clasificadores	Máquinas de soporte de vectores	0.178	0.673	0.468	0.398	3.203	0.403	0.403	8
		Bosques aleatorios	0.177	0.677	0.477	0.394	3.178	0.403	0.403	13
		Regresión logística	0.181	0.666	0.453	0.404	3.232	0.398	0.398	11
		Naive bayes	0.194	0.647	0.425	0.428	3.308	0.388	0.388	<b>0</b>
	Conjunto potencia	Máquinas de soporte de vectores	0.178	0.673	0.47	0.397	3.193	0.404	0.404	3
		Bosques aleatorios	<b>0.176</b>	0.678	0.48	<b>0.393</b>	<b>3.166</b>	0.405	0.405	5
		Regresión logística	0.18	0.669	0.463	0.401	3.21	0.399	0.399	16
		Naive bayes	0.188	0.65	0.444	0.417	3.274	0.384	0.384	<b>0</b>
Rakel-d	Máquinas de soporte de vectores	0.178	0.673	0.467	0.398	3.2	0.403	0.403	359	
	Bosques aleatorios	<b>0.176</b>	<b>0.679</b>	<b>0.478</b>	<b>0.393</b>	3.167	0.404	0.404	9	
	Regresión logística	0.18	0.669	0.46	0.401	3.215	0.4	0.4	18	
	Naive bayes	0.192	0.648	0.436	0.419	3.271	0.381	0.381	<b>0</b>	
Adaptación	-	Véctinos más cercanos multi-etiqueta	0.205	0.61	0.362	0.466	3.513	0.367	268	
	-	Máquinas de soporte de vectores multi-etiqueta	0.452	0.534	0.047	0.669	3.966	0.31	245	

Tabla 5.4: Métricas de desempeño de clasificadores utilizando un-dos-tres-gram

Enfoque	Método	Algoritmo	Perdida de Hamming ↓	F1 (micro) ↑	Exactitud ↑	Perdida de ordenamiento ↓	Cobertura ↓	Media de precisión ↑	Tiempo (Segundos) ↓
Transformación	Relevancia binaria	Máquinas de soporte de vectores	<b>0.081</b>	<b>0.857</b>	0.692	0.182	2.304	0.585	48
		Bosques aleatorios	0.082	0.856	0.705	0.18	<b>0.603</b>	41	
		Regresión logística	0.093	0.835	0.644	0.208	0.556	18	
		Naive bayes	0.133	0.764	0.552	0.286	0.467	<b>0</b>	
	Cadenas de clasificadores	Máquinas de soporte de vectores	0.083	0.853	0.709	<b>0.174</b>	2.284	0.587	42
		Bosques aleatorios	0.084	0.853	<b>0.712</b>	0.176	2.278	0.6	32
		Regresión logística	0.094	0.835	0.673	0.194	2.373	0.565	29
		Naive bayes	0.134	0.759	0.553	0.292	2.77	0.467	<b>0</b>
		Máquinas de soporte de vectores	0.09	0.841	0.7	0.186	2.326	0.575	30
		Bosques aleatorios	0.093	0.834	0.696	0.197	2.364	0.582	17
Conjunto potencia	Regresión logística	0.094	0.834	0.68	0.194	2.378	0.566	41	
	Naive bayes	0.134	0.755	0.574	0.288	2.76	0.466	<b>0</b>	
	Máquinas de soporte de vectores	0.086	0.851	0.695	0.178	2.282	0.578	999	
	Bosques aleatorios	0.084	0.853	0.706	0.176	<b>2.268</b>	0.594	30	
Rakel-d	Regresión logística	0.09	0.843	0.671	0.187	2.332	0.572	67	
	Naive bayes	0.136	0.753	0.557	0.287	2.762	0.459	<b>0</b>	
Adaptación	-	Vecinos más cercanos multi-etiqueta	0.152	0.741	0.556	0.297	2.663	0.48	780
	-	Máquinas de soporte de vectores multi-etiqueta	0.226	0.424	0.172	0.747	4.307	0.401	3601

Tabla 5.5: Métricas de desempeño de clasificadores utilizando un-dos-gram

Enfoque	Método	Algoritmo	Perdida de Hamming ↓	F1 (micro) ↑	Exactitud ↑	Perdida de ordenamiento ↓	Cobertura ↓	Media de precisión ↑	Tiempo (Segundos) ↓
Transformación	Relevancia binaria	Máquinas de soporte de vectores	<b>0.081</b>	<b>0.857</b>	0.692	0.183	2.305	0.584	43
		Bosques aleatorios	0.082	0.856	0.702	0.18	2.282	<b>0.6</b>	31
		Regresión logística	0.094	0.833	0.64	0.211	2.421	0.556	11
		Naive bayes	0.127	0.776	0.566	0.272	2.676	0.475	<b>0</b>
	Cadenas de clasificadores	Máquinas de soporte de vectores	0.083	0.853	0.709	<b>0.173</b>	2.279	0.588	38
		Bosques aleatorios	0.083	0.854	<b>0.718</b>	<b>0.173</b>	2.266	0.599	28
		Regresión logística	0.094	0.834	0.671	0.195	2.378	0.565	16
		Naive bayes	0.126	0.774	0.57	0.275	2.705	0.477	<b>0</b>
	Conjunto potencia	Máquinas de soporte de vectores	0.088	0.844	0.704	0.183	2.315	0.578	27
		Bosques aleatorios	0.093	0.836	0.694	0.194	2.357	0.581	15
		Regresión logística	0.094	0.834	0.681	0.194	2.375	0.566	25
		Naive bayes	0.129	0.765	0.585	0.277	2.717	0.476	<b>0</b>
Rakel-d	Máquinas de soporte de vectores	0.082	0.855	0.71	0.174	2.276	0.59	769	
	Bosques aleatorios	0.083	0.855	0.703	0.174	<b>2.261</b>	0.596	37	
	Regresión logística	0.092	0.837	0.673	0.197	2.388	0.567	40	
	Naive bayes	0.134	0.759	0.563	0.285	2.741	0.463	<b>0</b>	
Adaptación	-	Vécinos más cercanos multi-etiqueta	0.153	0.739	0.55	0.299	2.62	0.482	513
	-	Máquinas de soporte de vectores multi-etiqueta	0.224	0.433	0.176	0.74	4.29	0.404	917

Tabla 5.6: Métricas de desempeño de clasificadores utilizando dos-tres-gram

Enfoque	Método	Algoritmo	Perdida de Hamming ↓	F1 (micro) ↑	Exactitud ↑	Perdida de ordenamiento ↓	Cobertura ↓	Media de precisión ↑	Tiempo (Segundos) ↓
Transformación	Relevancia binaria	Máquinas de soporte de vectores	0.116	0.786	0.565	0.304	2.79	0.506	23
		Bosques aleatorios	<b>0.111</b>	<b>0.797</b>	0.582	0.293	2.726	<b>0.524</b>	50
		Regresión logística	0.124	0.777	0.578	0.28	2.719	0.487	18
		Naive bayes	0.149	0.736	0.528	0.319	3.862	0.45	<b>0</b>
	Cadenas de clasificadores	Máquinas de soporte de vectores	0.129	0.771	0.597	0.278	2.702	0.5	19
		Bosques aleatorios	0.123	0.784	0.622	0.264	2.628	0.519	41
		Regresión logística	0.127	0.772	0.588	0.276	2.717	0.486	24
		Naive bayes	0.15	0.729	0.529	0.329	2.911	0.449	<b>0</b>
		Máquinas de soporte de vectores	0.123	0.782	0.613	0.263	2.643	0.493	12
		Bosques aleatorios	0.12	0.789	<b>0.624</b>	<b>0.254</b>	2.598	0.507	26
Conjunto potencia	Regresión logística	0.128	0.772	0.594	0.273	2.696	0.485	34	
	Naive bayes	0.141	0.744	0.558	0.304	2.826	0.456	<b>0</b>	
	Máquinas de soporte de vectores	0.117	0.787	0.569	0.296	2.757	0.504	761	
	Bosques aleatorios	0.116	0.796	0.619	0.249	<b>2.575</b>	0.513	28	
Rakel-d	Regresión logística	0.123	0.783	0.594	0.265	2.661	0.492	48	
	Naive bayes	0.146	0.744	0.546	0.295	2.777	0.451	<b>0</b>	
Adaptación	-	Vecinos más cercanos multi-etiqueta	0.141	0.737	0.511	0.358	2.984	0.474	456
	-	Máquinas de soporte de vectores multi-etiqueta	0.247	0.327	0.105	0.842	4.598	0.371	1149

---

goritmos adaptados muestran una mala clasificación en comparación con los enfoques de transformación.

Para la métrica F1 a nivel micro que mide un balance entre cuán preciso y exhaustivo es cada clasificador, se observa que las características que logran una mejor clasificación son un-gram, uno-dos-tres-gram y uno-dos-gram. El método de transformación de datos que mejor clasificación logra es la relevancia binaria seguido de rakel-d y cadenas de clasificadores con los algoritmos base máquinas de soporte de vectores y bosques aleatorios. Los algoritmos adaptados muestran una mala clasificación en comparación con los enfoques de transformación.

Para la métrica que mide la exactitud (Accuracy), se encuentra que las características que permiten una mejor clasificación son un-gram, uno-dos-tres-gram y uno-dos-gram al alcanzar una predicción correcta de las cinco etiquetas en el 70 % de las retroalimentaciones. El enfoque de transformación que mejor clasificación logra son las cadenas de clasificadores seguido de la relevancia binaria y rakel-d. El algoritmo base que mejor clasificación logra es máquinas de soporte de vectores seguido de bosques aleatorios. Se muestra que los algoritmos adaptados y algoritmo base naive bayes son los que peor clasifican.

Para la métrica de pérdida de ordenamiento (ranking loss), se observa que las características que mejor clasificación logran son un-gram, uno-dos-tres-gram y uno-dos-gram. Los métodos de transformación de datos que mejor clasifican con cadenas de clasificadores y rakel-d con algoritmos base bosques aleatorios y máquinas de soporte de vectores. Los que peor clasifican son los algoritmos adaptados y al utilizar el algoritmo base naive bayes.

Para la métrica de cobertura, las características que permiten una mejor clasificación son un-gram, uno-dos-tres-gram y uno-dos-gram. Los métodos de transformación de datos que mejor clasifican son relevancia binaria, cadenas de clasificadores y rakel-d cuando utilizan el algoritmo base bosques aleatorios seguido de las máquinas de soporte de vectores.

Para el promedio de precisión, las características que permiten una mejor clasificación son un-gram, uno-dos-tres-gram y uno-dos-gram. El enfoque de transformación de datos que mejor clasifica es la relevancia binaria y cadenas de clasificadores cuando tienen bosques aleatorios como algoritmo base. Los algoritmos adaptados y el uso de naive bayes como algoritmo base tienen más errores en la clasificación.

Sobre el tiempo para entrenar un clasificador se observa que son los algoritmos adaptados los que más tardan seguido del enfoque de transformación de datos rakelD cuando utiliza el algoritmo base máquinas de soporte de vectores. El algoritmo base naive bayes es el más rápido en entrenar, pero con una clasificación mala al considerar las demás métricas. Al considerar las características, se observa que es un-gram el que mejor tiempo requiere para el entrenamiento del clasificador.

A partir de los resultados obtenidos en la primera fase de clasificación se realizan estudios para determinar el enfoque, algoritmo, método de transformación y características que permiten un mejor desempeño de clasificadores multi-etiqueta que categorizan retroalimentaciones conforme los niveles propuesto de Hattie y Timperley, a

continuación se desarrolla cada uno.

## 5.2. Estudio comparativo de enfoques de clasificación multi-etiqueta

Las figuras 5.1, 5.2, 5.3 muestran las gráficas radiales de los clasificadores multi-etiqueta que mejor desempeño evidenciaron utilizando el enfoque de transformación y adaptación utilizando un-gram, dos-gram y tres-gram respectiva. Los polígonos con líneas rectas corresponden a los clasificadores multi-etiqueta que fueron entrenados con el enfoque de transformación utilizando los métodos de transformación relevancia binaria, cadenas de clasificadores, conjunto potencia y rakel-d, mientras que los de la línea punteada corresponden a los clasificadores multi-etiqueta entrenados con el enfoque adaptación.

Se evidencia que los polígonos que forman los clasificadores multi-etiqueta entrenados con el enfoque de transformación tiene una área mayor que los del enfoque de adaptación, lo cual es indicativo de tener un mejor desempeño.

Es en la métrica de exactitud en la que se observa una mayor diferencia entre los enfoques de transformación y adaptación.



Figura 5.1: Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los enfoques transformación y adaptación utilizando un-grama. Para el enfoque de transformación se usan los métodos de relevancia binaria, cadenas de clasificadores y conjunto potencia con los algoritmos máquinas de soporte de vectores y bosques aleatorios. Para el enfoque de adaptación se utilizan los algoritmos máquinas de soporte de vectores multi-etiqueta y vecinos más cercanos multi-etiqueta



Figura 5.2: Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los enfoques transformación y adaptación utilizando dos-grama. Para el enfoque de transformación se usan los métodos de relevancia binaria, cadenas de clasificadores y conjunto potencia con los algoritmos máquinas de soporte de vectores y bosques aleatorios. Para el enfoque de adaptación se utilizan los algoritmos máquinas de soporte de vectores multi-etiqueta y vecinos más cercanos multi-etiqueta



Figura 5.3: Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los enfoques transformación y adaptación utilizando tres-grama. Para el enfoque de transformación se usan los métodos de relevancia binaria, cadenas de clasificadores y conjunto potencia con los algoritmos máquinas de soporte de vectores y bosques aleatorios. Para el enfoque de adaptación se utilizan los algoritmos máquinas de soporte de vectores multi-etiqueta y vecinos más cercanos multi-etiqueta

La figura 5.4 muestra de igual forma los clasificadores multi-etiqueta entrenados con los enfoques de transformación y adaptación con afinación de hiper-parámetros. Se observa son los clasificadores multi-etiqueta que utilizan el enfoque de transformación los que tienen mejor desempeño que los de adaptación. También, se observa que aplicando la etapa de afinación de hiper-parámetros permite que el algoritmo adaptado de vecinos más cercanos multi-etiqueta alcance valores cercanos a los que utilizan el enfoque de transformación.

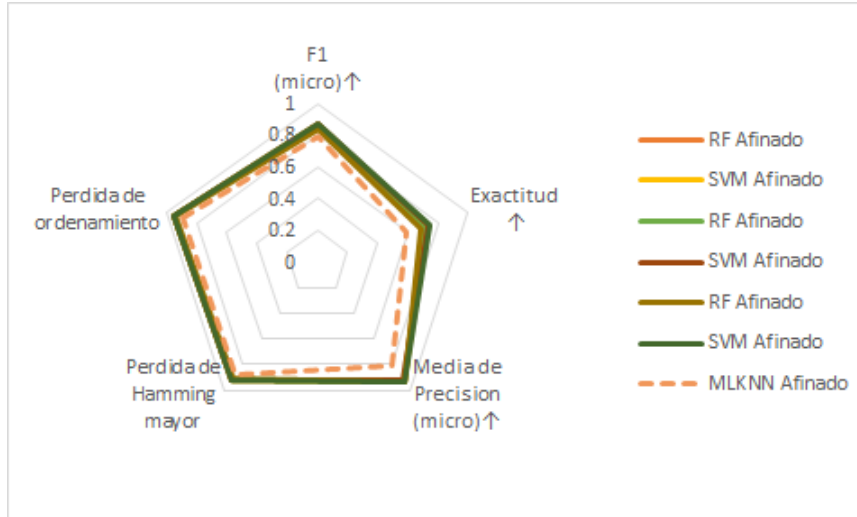


Figura 5.4: Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los enfoques transformación y adaptación utilizando un-grama con afinación de hiper-parámetros. Para el enfoque de transformación se usan los métodos de relevancia binaria, cadenas de clasificadores y conjunto potencia con los algoritmos máquinas de soporte de vectores y bosques aleatorios. Para el enfoque de adaptación se utiliza el algoritmo vecinos más cercanos multi-etiqueta

Como conclusión de este estudio se tiene que el enfoque de clasificación multi-etiqueta de transformación tendrá un mejor desempeño que la adaptación de algoritmos para el conjunto de datos utilizado en este trabajo. Cualquiera de los métodos de transformación como relevancia binaria, cadenas de clasificadores, conjunto potencia y raket-d será mejor que un algoritmo adaptado. Aplicar afinación de hiper-parámetros permitirá que los algoritmos adaptados se acerquen al desempeño de los clasificadores entrenados con el enfoque transformación pero sin superarlos.

### 5.3. Estudio comparativo de algoritmos

Las figuras 5.5, 5.6, 5.7, 5.8 muestran las gráficas radiales de los clasificadores multi-etiqueta que fueron entrenados bajo el enfoque de transformación con los métodos relevancia binaria, cadenas de clasificadores, conjunto potencia y raket-d comparando los

---

algoritmos máquinas de soporte de vectores, bosques aleatorios, regresión logística y naive bayes utilizando como características para el entrenamiento la bolsa de palabras.

Se detecta que de los polígonos formados son los de los algoritmos de máquinas de soporte de vectores y bosques aleatorios los de mayor área en comparación con regresión logística y naive bayes lo cual es indicativo de que tienen un mejor desempeño en la clasificación automática de retroalimentaciones. También se observa que el algoritmo que muestra el peor desempeño es naive bayes por lo que no debe considerarse para la realización de la tarea. El algoritmo de regresión logística muestra un buen desempeño, pero queda por debajo de las máquinas de soporte de vectores y bosques aleatorios cuando se utilizan los métodos de relevancia binaria y cadenas de clasificadores.

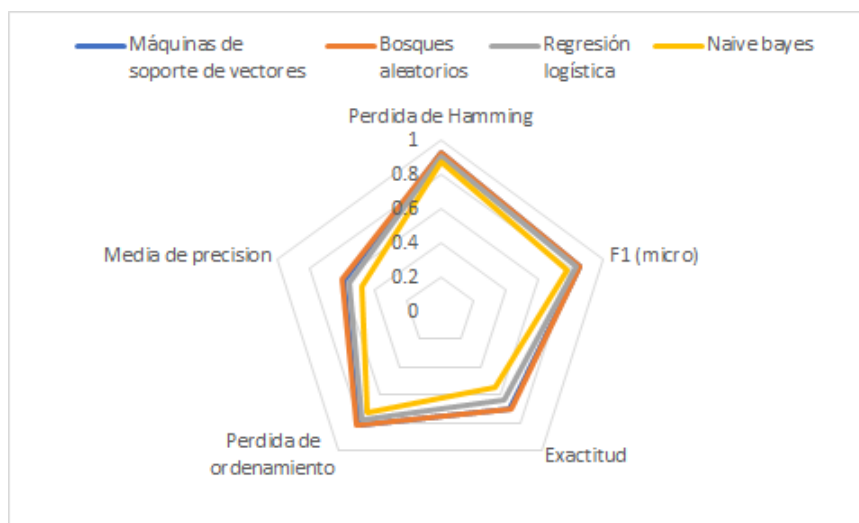


Figura 5.5: Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los algoritmos máquinas de soporte de vectores (SVM), bosques aleatorios (RF), regresión logística (LR) y naive bayes (NB) con método de transformación de relevancia binaria.

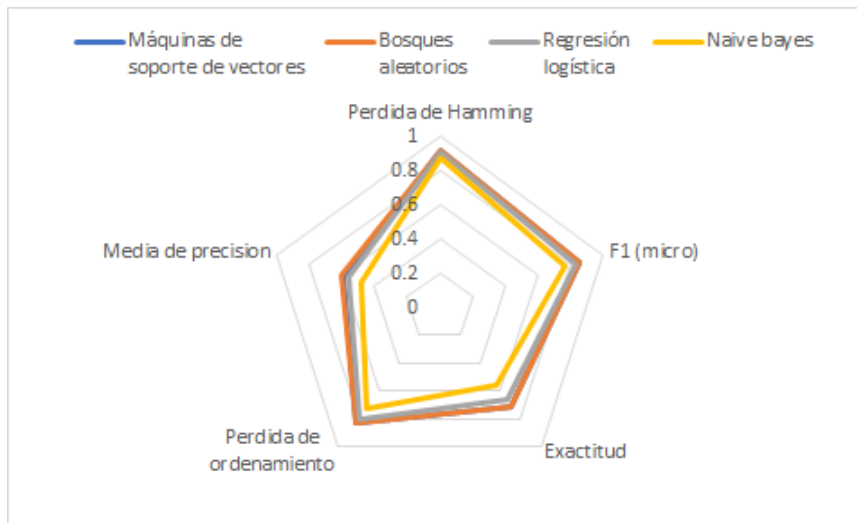


Figura 5.6: Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los algoritmos máquinas de soporte de vectores (SVM), bosques aleatorios (RF), regresión logística (LR) y naive bayes (NB) con método de transformación de cadenas de clasificadores.

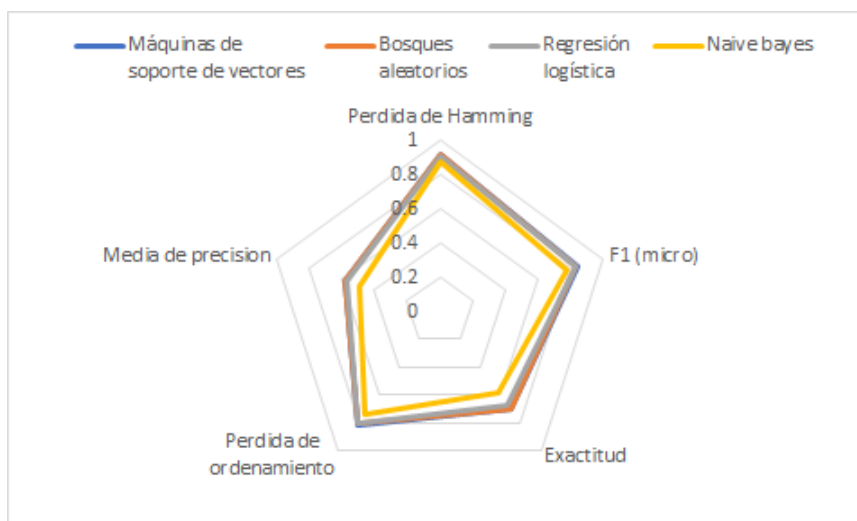


Figura 5.7: Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los algoritmos máquinas de soporte de vectores (SVM), bosques aleatorios (RF), regresión logística (LR) y naive bayes (NB) con método de transformación de conjunto potencia.

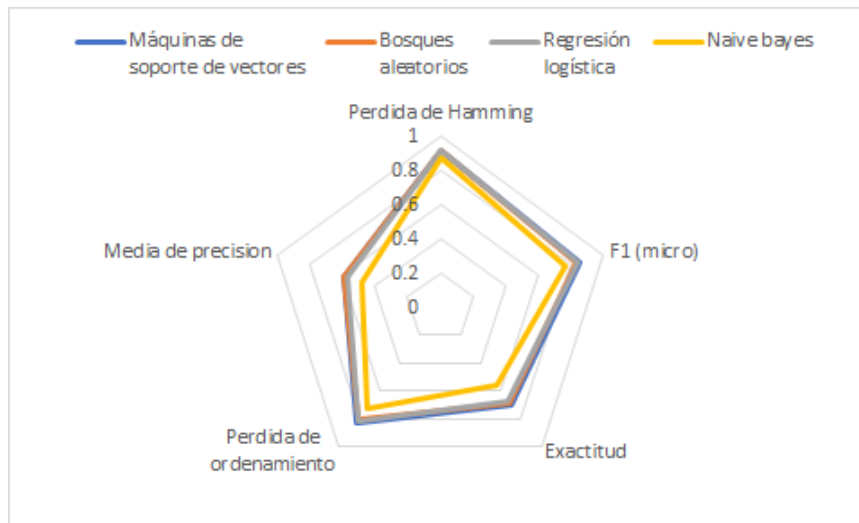


Figura 5.8: Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los algoritmos máquinas de soporte de vectores (SVM), bosques aleatorios (RF), regresión logística (LR) y naive bayes (NB) con método de transformación de raket-d.

La figura 5.9 muestra el desempeño de clasificadores multi-etiqueta entrenados con el enfoque de adaptación a través de los algoritmos vecinos más cercanos y máquinas de soporte de vectores multi-etiqueta. Se observa que es el algoritmo vecinos más cercanos multi-etiqueta el que mejor desempeño logra en contraste con máquinas de soporte de vectores multi-etiqueta al tener un polígono de mayor área.

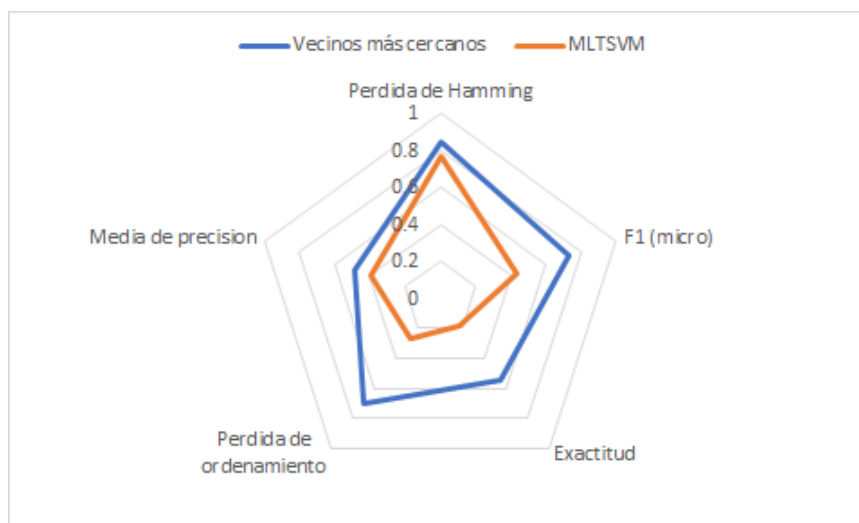


Figura 5.9: Gráfica radial de desempeño de clasificadores multi-etiqueta utilizando los algoritmos adaptados máquinas de soporte de vectores multi-etiqueta (MLTSVN) y vecinos más cercanos multi-etiqueta (MLKNN).

Las conclusiones de este estudio fueron que para el enfoque de transformación los algoritmos de máquinas de soporte de vectores y bosques aleatorios son los que permiten entrenar clasificadores multi-etiqueta con mejor desempeño en la tarea de ubicar retroalimentaciones conforme los niveles de retroalimentación de Hattie y Timperley. Por su parte, para el enfoque de adaptación es el algoritmo vecinos más cercanos multi-etiqueta el que permite tener clasificadores multi-etiqueta con un desempeño mejor.

## 5.4. Estudio comparativo de las características

Las figuras 5.10, 5.11, 5.12, 5.13, 5.14, 5.15, 5.16, 5.17 muestran el desempeño de los clasificadores multi-etiqueta utilizando el enfoque de transformación de relevancia binaria, cadenas de clasificadores, conjunto potencia y rakeld con los algoritmos máquinas de soporte de vectores y bosques aleatorios utilizando las características de un-grama, dos-grama, tres-grama, uno-dos-tres-grama, un-dos-grama y dos-tres-grama. Se detecta que para todos los métodos de transformación y algoritmos utilizados el mejor desempeño se obtiene cuando las características incluyen la división de un-grama (un-grama, uno-dos-grama y uno-dos-tres-grama) en comparación de cuando se usa dos-grama, tres-grama, dos-tres-grama.

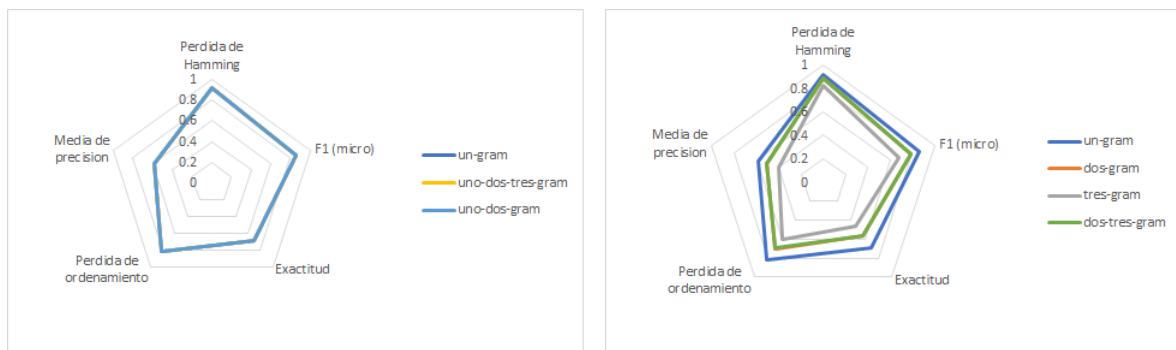


Figura 5.10: Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método de relevancia binaria y algoritmo máquinas de soporte de vectores con diferentes n-gramas

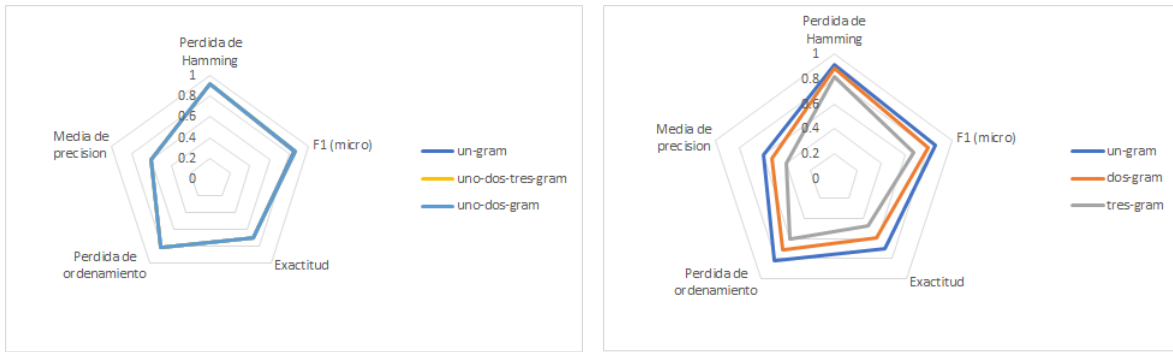


Figura 5.11: Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método de relevancia binaria y algoritmo bosques aleatorios con diferentes n-gramas



Figura 5.12: Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método de cadenas de clasificadores y algoritmo máquinas de soporte de vectores con diferentes n-gramas

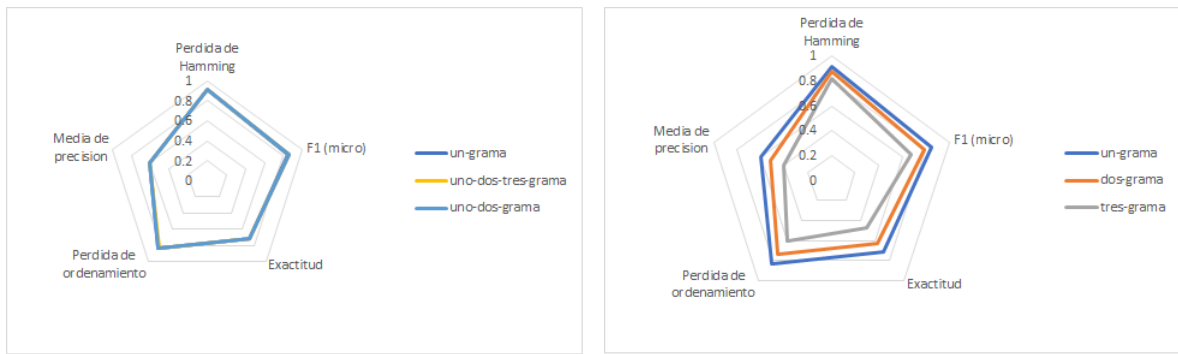


Figura 5.13: Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método de cadenas de clasificadores y algoritmo bosques aleatorios con diferentes n-gramas



Figura 5.14: Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método conjunto potencia y algoritmo máquinas de soporte de vectores con diferentes n-gramas



Figura 5.15: Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método conjunto potencia y algoritmo bosques aleatorios con diferentes n-gramas

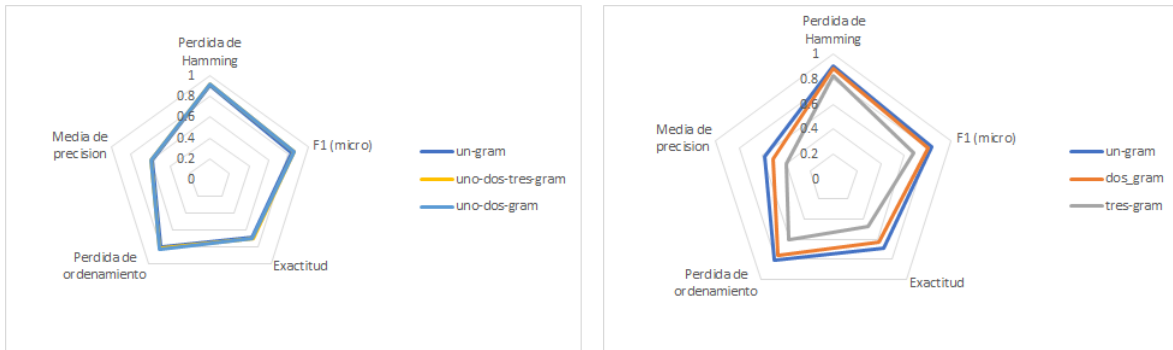


Figura 5.16: Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método raket-d y algoritmo máquinas de soporte de vectores con diferentes n-gramas

La figura 5.18 muestra el tiempo invertido en el entrenamiento de los clasificadores multi-etiqueta utilizando las características un-gran, uno-dos-grama y uno-dos-tres-grama. Se observa que cuando se utiliza un-grama se requiere menos tiempo motivo por el cual se puede considerar que el uso de un-grama permitirá tener el mejor desempeño en un menor tiempo.

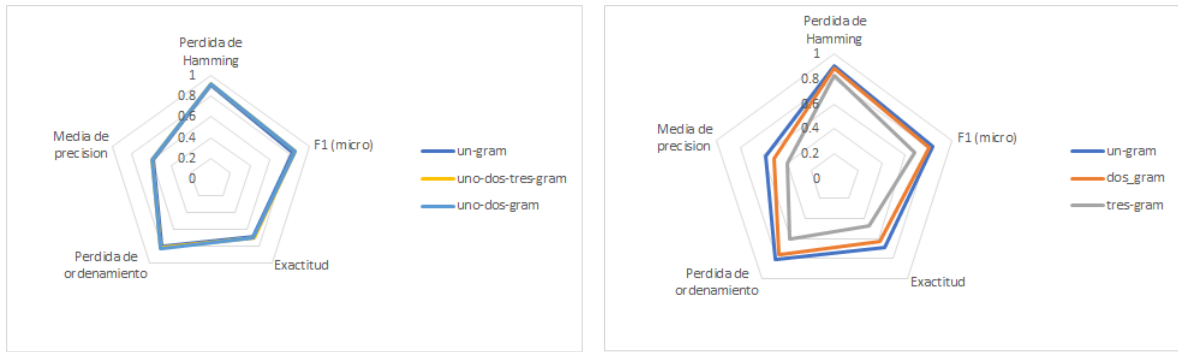


Figura 5.17: Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque transformación utilizando el método raket-d y algoritmo bosques aleatorios con diferentes n-gramas

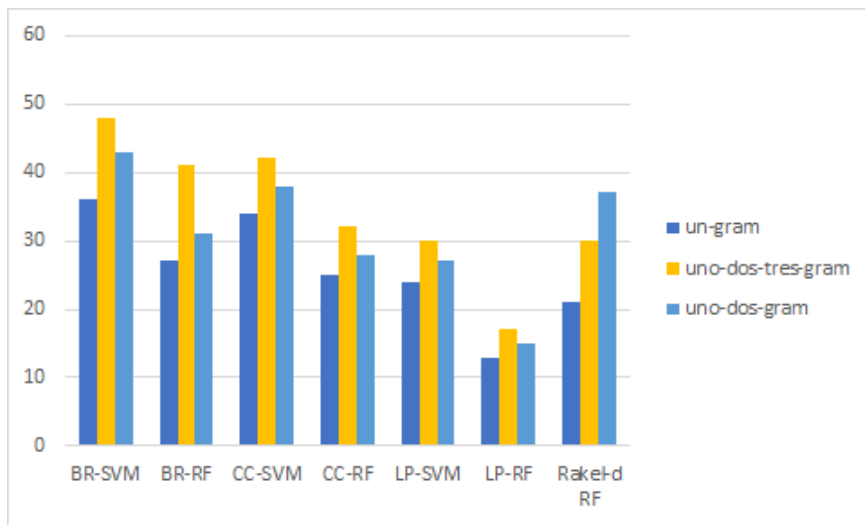


Figura 5.18: Segundos para el entrenamiento de clasificadores multi-etiqueta utilizando un-grama, uno-dos-grama y uno-dos-tres-grama.

En la figura 5.19 se muestra el desempeño de los clasificadores multi-etiqueta entrenados con el enfoque de adaptación a través del algoritmo MLKNN con las divisiones de características un-grama, dos-grama, tres-grama, uno-dos-tres-grama, un-dos-grama y dos-tres-grama. Se observa que los clasificadores con mejor desempeño son aquellos que incluyen entre sus características la división un-grama. También se detecta que el uso de dos-grama también alcanza desempeños altos.

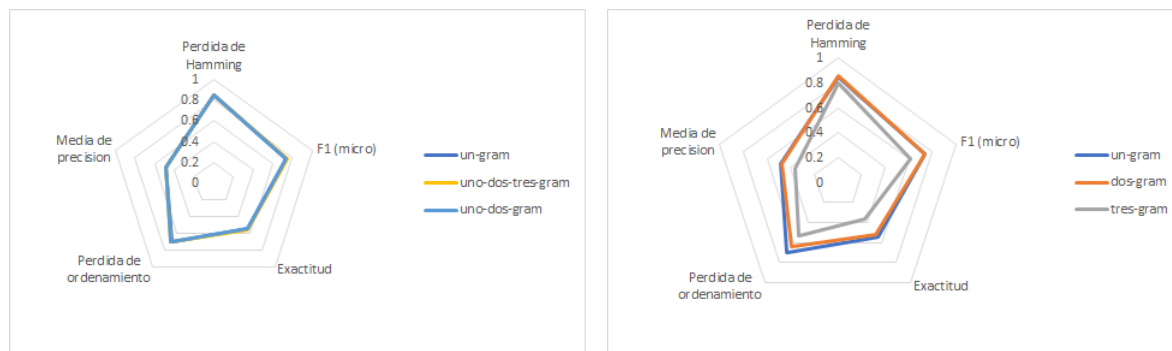


Figura 5.19: Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados con el enfoque adaptación utilizando el algoritmo vecinos más cercanos multi-etiqueta con diferentes n-gramas

Del análisis del tiempo que se requiere para el entrenamiento de clasificadores multi-etiqueta con el enfoque de adaptación se detecta que será menor cuándo se utiliza la división un-grama, razón por la que puede considerarse que la mejor característica a utilizar es un-grama, ver figura 5.20.

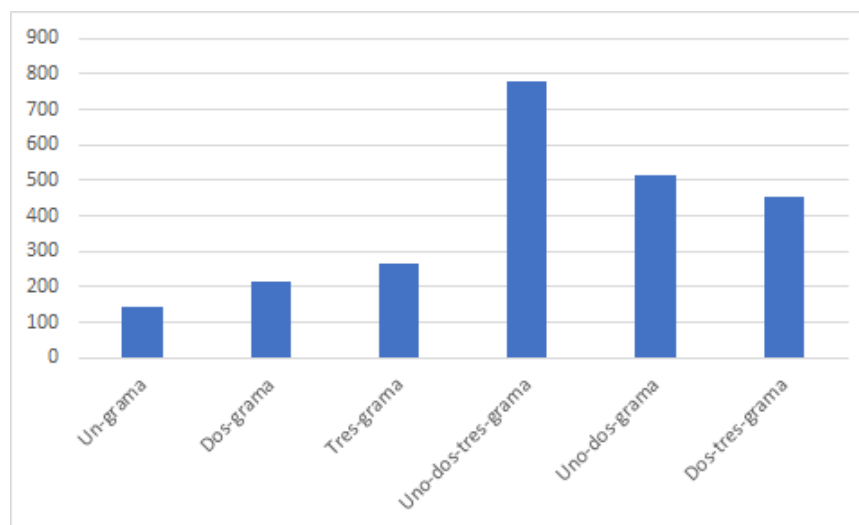


Figura 5.20: Segundos para el entrenamiento de clasificadores multi-etiqueta utilizando algoritmo MLKNN y diferentes n-gramas.

Las conclusiones de este estudio fueron que el uso de un-gram permitirá obtener clasificadores multi-etiqueta con mejor desempeño en la tarea de ubicar las retroalimentaciones en los niveles propuesto por Hattie y Timperley al tener las métricas de clasificación con valores más altos en un menor tiempo de entrenamiento.

## 5.5. Estudio comparativo de los métodos de transformación

La figura 5.21 muestran el desempeño de clasificadores multi-etiqueta entrenados con los métodos de transformación de relevancia binaria, cadenas de clasificadores, conjunto potencia y rakel-d utilizando los algoritmos de máquinas de soporte de vectores y bosques aleatorios. Se observa que cuando se utiliza el algoritmo de máquinas de soporte de vectores se obtienen valores similares para los métodos de relevancia binaria, cadenas de clasificadores y conjunto potencia. El método de transformación de rakel-d da un desempeño bajo. Para el algoritmo de bosques aleatorios se detectan desempeños similares para los métodos de transformación utilizados.

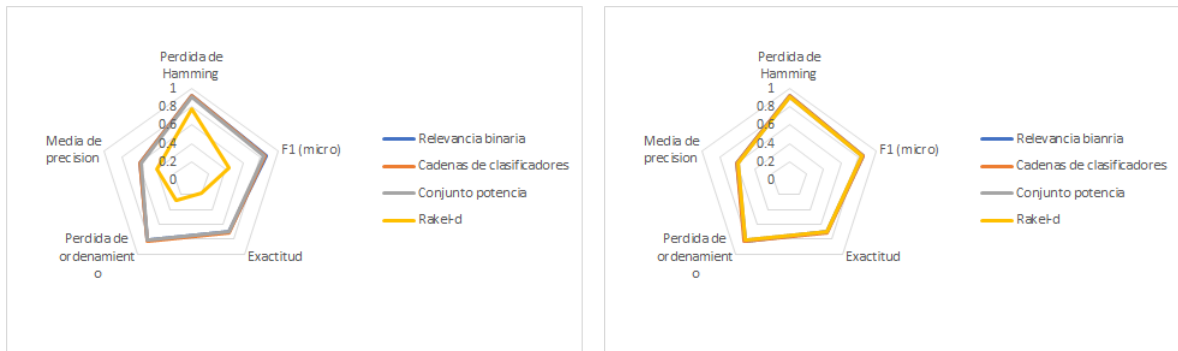


Figura 5.21: Gráficas radial de desempeño de clasificadores multi-etiqueta entrenados utilizando los métodos de transformación relevancia binaria, cadenas de clasificadores, conjunto potencia y rakel-d con algoritmos máquinas de soporte de vectores (izquierda) y bosques aleatorios(derecha)

La figura 5.22 muestra una comparativa del tiempo que se requiere para el entrenamiento de clasificadores utilizando cada método de transformación. Se evidencia que la relevancia binaria y cadenas de clasificadores son los que requieren de mayor tiempo para el entrenamiento que el método de conjunto potencia lo cual se debe a la cantidad de clasificadores binarios que deben entrenar mientras que en el segundo solo es un clasificador multiclase.

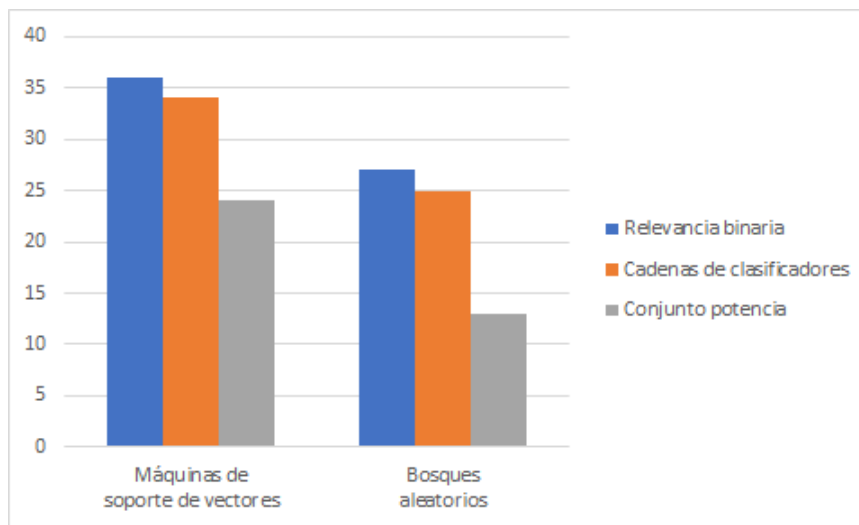


Figura 5.22: Segundos para el entrenamiento de clasificadores multi-etiqueta utilizando diferentes métodos de transformación.

Las conclusiones de este estudio fueron que cualquiera de los métodos de transformación utilizados en el estudio permite tener un buen desempeño de clasificadores multi-etiqueta que ubiquen retroalimentaciones conforme el modelo Hattie y Timperley. Aunque los métodos de conjunto potencia y raket-d alcanzan buenos desempeños debe considerarse que en el conjunto de datos no se encuentran todas la posibles combinaciones de etiquetas lo cual si una nueva retroalimentación se encuentra en una combinación no considerada se clasificará de forma errónea.

## 5.6. Evaluación de la afinación de hiper-parámetros

Los hiper-parámetros por defecto que tiene cada algoritmo y los que se obtuvieron de la etapa de búsqueda de hiper parámetros se muestran en la Tabla 5.7. Para el algoritmo de máquinas de soporte de vectores, se tiene que el mejor valor para la métrica de exactitud se logra cuando el hiper-parámetro de regularización  $C$  tiene el valor 100. Para los bosques aleatorios el número de árboles que debe tener el bosque para tener la mejor exactitud en el conjunto de entrenamiento es 400 cuando se utilizan los métodos de transformación relevancia binaria y cadenas de clasificadores; para el conjunto potencia se debe tener un valor de 500. Para el algoritmo de vecinos más cercanos multi-etiqueta se obtienen los valores  $k=8$  y  $s=1.0$  como los hiper parámetros que logran una mejor exactitud para el conjunto de datos de entrenamiento.

En la Tabla 5.8, se muestran las métricas de desempeño alcanzadas por cada clasificador utilizando valores por defecto y el uso de hiper-parámetros afinados evaluados con el conjunto de validación. Las flechas en cada una de las métricas indican si es mejor un valor alto o bajo.

Tabla 5.7: Parámetros por defecto y afinados

Algoritmo	Método de transformación	Parámetros por defecto	Parámetros afinados
Máquinas de soporte de vectores	Relevancia binaria	C=1.0	C=100
	Cadenas de clasificadores	C=1.0	C=100
	Conjunto potencia	C=1.0	C=100
Bosques aleatorios	Relevancia binaria	n_estimators = 100	n_estimators = 400
	Cadenas de clasificadores	n_estimators = 100	n_estimators = 400
	Conjunto potencia	n_estimators = 100	n_estimators = 500
Vécinos más cercanos multi-etiqueta	-	K = 10; s=1.0	K=8; s=1.0

Tabla 5.8: Desempeño de los clasificadores multi-etiqueta utilizando hiper-parámetros por defecto y afinados

Enfoque	Método	Algoritmo	Hamming Loss ↓	F1 (micro) ↑	Accuracy ↑	Ranking loss ↓	Avg Precision (micro) ↑	Coverage ↓	Tiempo (segundos) ↓	
Transformación	Relevancia binaria	Bosques aleatorios	0.083	0.854	0.698	0.05	0.935	0.221	26	
		<b>Bosques aleatorios (Afinado)</b>	0.083	0.854	0.7	0.049	0.937	0.209	104	
		Máquinas de soporte de vectores	0.081	0.856	0.691	0.044	0.931	<b>0.196</b>	193	
	Cadenas de clasificadores	<b>Máquinas de soporte de vectores (Afinado)</b>	Máquinas de soporte de vectores	0.074	<b>0.871</b>	0.72	<b>0.044</b>	0.938	0.198	393
			Bosques aleatorios	0.086	0.848	0.707	0.055	0.929	0.234	24
			<b>Bosques aleatorios (Afinado)</b>	0.084	0.852	0.712	0.05	0.934	0.218	94
		Conjunto potencia	Máquinas de soporte de vectores	0.084	0.852	0.709	0.049	0.894	0.218	179
			<b>Máquinas de soporte de vectores (Afinado)</b>	0.077	0.866	0.736	0.049	0.915	0.217	207
			Bosques aleatorios	0.115	0.798	0.64	0.077	0.902	0.335	37
	Conjunto potencia	<b>Bosques aleatorios (Afinado)</b>	0.096	0.83	0.69	0.056	0.93	0.24	48	
		Máquinas de soporte de vectores	0.118	0.789	0.628	0.08	0.884	0.345	147	
		<b>Máquinas de soporte de vectores (Afinado)</b>	0.077	0.865	<b>0.74</b>	0.044	<b>0.939</b>	0.193	190	
Adaptación	-	Vécinos más cercanos multi-etiqueta	0.123	0.789	0.589	0.115	0.792	0.496	<b>12</b>	
	-	<b>Vécinos más cercanos multi-etiqueta (Afinado)</b>	0.123	0.789	0.589	0.114	0.801	0.487	<b>12</b>	

---

Del análisis de los resultados, se detecta que el clasificador que mejor logra el objetivo de clasificar automáticamente las retroalimentaciones que genera un docente a las actividades enviadas por estudiantes en cursos en línea de la plataforma Blackboard en los niveles tarea, proceso, regulación y elogios propuestos por Hattie y Timperley (2007), a través de técnicas de minería de textos para estimar y evaluar la toma de decisiones respecto al desempeño docentes es el que utiliza el enfoque de transformación, con el algoritmo de máquinas de soporte de vectores utilizando los hiper-parámetros afinados y el método de transformación de datos de relevancia binaria. El clasificador multi-etiqueta logra tener solo el 7.4% de todas las etiquetas mal clasificadas respecto a lo etiquetado por los expertos como lo indica la métrica de Hamming loss. Se logra un buen balance de precisión y exhaustividad al obtener un valor 87.1%. Para la métrica de exactitud se logra tener 72% de las retroalimentaciones con las 5 etiquetas bien clasificadas. También, logra una precisión aceptable al lograr un 93%.

Se observa que cuando se utiliza el enfoque de transformación con el algoritmo de máquinas de soporte de vectores con hiper parámetros afinados y método de transformación de conjunto potencia se logran los mejores valores para las métricas de exactitud y precisión promedio. Sin embargo, este clasificador no se deja como primera opción ya que en el conjunto de datos multi-etiqueta de entrenamiento solo se cuenta con 23 de las 32 posibles combinaciones de etiquetas por los que si se tuviese una nueva retroalimentación que no se encuentre entre las posibles combinaciones se etiquetara de manera incorrecta.

De igual forma se nota que el uso hiper-parámetros afinados mejora las métricas de desempeño de los clasificadores multi-etiqueta independientemente del enfoque, algoritmo y método de transformación en comparación al uso de valores por defecto.

Se destaca que el tiempo de entrenamiento de los clasificadores multi-etiqueta será mayor cuando se utilizan los hiper-parámetros afinados en comparación con los parámetros por defecto.

Las conclusiones de este estudio fueron que la afinación de hiper-parámetros permite mejorar el desempeño de clasificadores multi-etiqueta que ubican retroalimentaciones conforme el modelo de retroalimentación de Hattie y Timperley. Cabe señalar que el tiempo de entrenamiento aumentará con los hiper-parámetros afinados.

# Capítulo 6

## Conclusiones

Los objetivos de esta investigación se cumplieron ya que se pudo proponer una metodología para analizar automáticamente las retroalimentaciones que genera un docente a las actividades enviadas por estudiantes en cursos en línea de la plataforma Blackboard en los niveles tarea, proceso, regulación y elogios propuestos por Hattie y Timperley (2007) a través de técnicas de clasificación multi-etiqueta para estimar y evaluar la toma de decisiones respecto al desempeño docente.

La hipótesis propuesta sobre que las técnicas de clasificación multi-etiqueta sirven para analizar automáticamente los comentarios resultantes de la revisión y el análisis del docente a las actividades que envía el estudiante en cursos en línea de la licenciatura en derecho de la plataforma Blackboard en los niveles tarea, proceso, regulación, elogios y otros se acepta al obtener clasificadores multi-etiqueta con métricas de desempeño aceptables respecto a lo etiquetado manualmente por expertos.

La construcción de herramientas que permitan analizar automáticamente los comentarios que generan los docentes a las actividades enviadas por estudiantes es una buena alternativa para las instituciones que ofertan cursos en línea para evaluar el desempeño de docente y la toma de decisiones ya que permiten identificar de manera rápida áreas de oportunidad.

La metodología propuesta es adecuada ya que permite la construcción de clasificadores multi-etiqueta para el análisis automático de los comentarios al estudiante al ubicarlos en los niveles tarea, proceso, regulación, elogios y otros.

Del estudio de enfoques de clasificación multi-etiqueta se llega a la conclusión de que el enfoque de transformación tendrá un mejor desempeño que la adaptación de algoritmos para el conjunto de datos utilizado en este trabajo. Cualquiera de los métodos de transformación como relevancia binaria, cadenas de clasificadores, conjunto potencia y raket-d tendrá un mejor desempeño que si se utilizan los algoritmos adaptados de máquinas de soporte de vectores y vecinos más cercanos multi-etiqueta.

Del estudio de los algoritmos de clasificación se concluye que para el enfoque de transformación las máquinas de soporte de vectores y bosques aleatorios son los que permiten entrenar clasificadores multi-etiqueta con mejor desempeño en la tarea de ubicar retroalimentaciones conforme los niveles de retroalimentación de Hattie y Tim-

---

perley con respecto a lo etiquetado manualmente por expertos y en contraste con los algoritmos naive bayes y regresión logística. Por su parte, para el enfoque de adaptación es el algoritmo MLKNN el que permite tener clasificadores multi-etiqueta con un desempeño mejor que el algoritmo adaptado de máquinas de soporte de vectores.

Del estudio de las características, se determina que el uso de un-gran permitirá obtener clasificadores multi-etiqueta con mejor desempeño en la tarea de ubicar las retroalimentaciones en los niveles propuesto por Hattie y Timperley al tener las métricas de clasificación con valores más altos en un menor tiempo de entrenamiento en comparación de cuando se utilizan dos-grama, tres-grama, dos-tres-grama, uno-dos-grama y uno-dos-tres-grama.

Del estudio de los métodos de transformación, se determina que tanto la relevancia binaria, cadenas de clasificadores, conjunto potencia y rakel-d permiten tener un buen desempeño de clasificadores multi-etiqueta que ubiquen retroalimentaciones conforme el modelo Hattie y Timperley con respecto a lo etiquetado manualmente por expertos. Se destaca que la relevancia binaria tendrá mejor desempeño pero no tan significativos. Aunque los métodos de conjunto potencia y rakel-d alcanzan buenos desempeños debe considerarse que en el conjunto de datos no se encuentran todas las posibles combinaciones de etiquetas lo cual si una nueva retroalimentación se encuentra en una combinación no considerada se clasificará de forma errónea.

De la evaluación de afinación hiper-parámetros se concluye que aplicar esta etapa previa al entrenamiento de algoritmos, permite mejorar el desempeño de clasificadores multi-etiqueta que ubican retroalimentaciones conforme el modelo de retroalimentación de Hattie y Timperley con respecto a lo etiquetado manualmente por expertos. Cabe señalar que la afinación de hiper-parámetros incrementará significativamente el tiempo de entrenamiento de los clasificadores multi-etiqueta.

## 6.1. Trabajo futuro

La metodología propuesta, se probó con comentarios al estudiante provenientes de cursos de la facultad de derecho por lo que se buscará probarla con otros programas educativos que se ofertan en las institución donde se realizó la investigación.

Como trabajo futuro se planea seguir investigando sobre características que permiten analizar los comentarios a los estudiantes que no sean dependientes del lenguaje, se buscará utilizar características numéricas y conteo de partes del discurso como son verbos, adjetivos y sustantivos por comentario.

También se buscará emplear palabras embebidas para generar vectores de cada comentario al estudiante y comparar el desempeño con el uso de n-gramas.

Por último, se analizará si la aplicación de técnicas de balaceo permite mejorar el desempeño de los clasificadores multi-etiqueta en la tarea de ubicar los comentarios en los niveles tarea, proceso, regulación y elogios.

# Referencias documentales

- Aguerreberre, C., Cabeza, S., Kaplan, G., Marconi, C., Cobo, C., and Bulger, M. (2018). Exploring feedback interactions in online learning environments for secondary education. *CEUR Workshop Proceedings*, 2231:1–10.
- Ajjawi, R. and Boud, D. (2017). Researching feedback dialogue: an interactional analysis approach. *Assessment and Evaluation in Higher Education*, 42(2):252–265.
- Al-Salemi, B., Ayob, M., Kendall, G., and Noah, S. A. M. (2019). Multi-label arabic text categorization: A benchmark and baseline comparison of multi-label learning algorithms. *Information Processing & Management*, 56(1):212–227.
- Almenara, J. and Graván, P. (2006). *E-actividades: un referente básico para la formación en Internet*. MAD-Eduforma, España.
- Aly, M. (2005). Survey on multiclass classification methods. *Neural Netw*, 19:1–9.
- Ardila, J. Y. and Marcela, R. E. (2015). Tres dimensiones para la evaluación de sistemas de gestión de aprendizaje (lms). *Zona próxima Revista del Instituto de Estudios en Educación Universidad del Norte*, (22):69–86.
- Asociación de Internet.mx (2019). Educación en línea en México 2018. Technical report, Asociación de Internet.mx, <https://www.asociaciondeinternet.mx/estudios/educacion-enlinea-mexico>.
- Blanco, A., Perez, O., Pérez, A., and Casillas, A. (2020). Boosting icd multi-label classification of health records with contextual embeddings and label-granularity. *Computer methods and programs in biomedicine*, 188:1–27.
- Brooks, C., Carroll, A., Gillies, R. M., and Hattie, J. (2019). A matrix of feedback for learning. *Australian Journal of Teacher Education*, 44(4):14–32.
- BUAP (2015). Reglamento de requisitos y procedimientos para la admisión, permanencia y trayectoria académica de los alumnos de la benemérita universidad autónoma de puebla en modalidades alternativas. Technical report, BUAP, <https://repositorio.buap.mx/rdocencia/public/inf-public/2019/0/Reglamento-Requisitos-Admision-Modalidades-Alternativas.pdf>.

- 
- BUAP (2016). Anuario estadístico 2015-2016. Technical report, BUAP, <https://planeacion.buap.mx/?q=menu/anuarios-estadisticos>.
- BUAP (2017). Anuario estadístico 2016-2017. Technical report, BUAP, <https://planeacion.buap.mx/?q=menu/anuarios-estadisticos>.
- BUAP (2018). Anuario estadístico 2017-2018. Technical report, BUAP, <https://planeacion.buap.mx/?q=menu/anuarios-estadisticos>.
- BUAP (2019). Anuario estadístico 2018-2019. Technical report, BUAP, <https://planeacion.buap.mx/?q=menu/anuarios-estadisticos>.
- BUAP (2020). Anuario estadístico 2019-2020. Technical report, BUAP, <https://planeacion.buap.mx/?q=menu/anuarios-estadisticos>.
- Cabero, J. (2006). Bases pedagógicas del e-learning. *Didáctica, Innovación y Multimedia*, 3(6):e0.
- Cabero, J. (2016). La educación a distancia como estrategia de inclusión social y educativa. *Revista Mexicana de Bachillerato a Distancia*, 8(15):138–147.
- Cabrera, L. A., Bessis, N., and Korkontzelos, I. (2020). Classifying emotions in stack overflow and jira using a multi-label approach. *Knowledge-Based Systems*, 195:1–12.
- Cavalcanti, A. P., Diego, A., Mello, R. F., Mangaroska, K., Nascimento, A., Freitas, F., and Gašević, D. (2020). How good is my feedback? a content analysis of written feedback. In *Proceedings of the tenth international conference on learning analytics & knowledge*, pages 428–437.
- Chen, W. J., Shao, Y. H., Li, C. N., and Deng, N. Y. (2016). Mltsvm: a novel twin support vector machine to multi-label learning. *Pattern Recognition*, 52:61–74.
- Clarenc, C. A. (2013). *Análisis comparativo de LMS*. Lulu.
- Fui, C. S. and Lian, L. H. (2018). The effect of computerized feedback on students’ misconceptions in algebraic expression. *Pertanika Journal of Social Sciences & Humanities*, 26(3):1387–1403.
- Gan, M. J. and Hattie, J. (2014). Prompting secondary students’ use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, 42(6):861–878.
- García, L. (2020). Bosque semántico: ¿educación/enseñanza/aprendizaje a distancia, virtual, en línea, digital, elearning...? *RIED. Revista Iberoamericana de Educación a Distancia*, 23(1):9–28.

- 
- García, F. J. and Seoane, A. M. (2015). Una revisión actualizada del concepto de elearning: décimo aniversario = an updated review of the concept of elearning: tenth anniversary. *Una revisión actualizada del concepto de eLearning: décimo Aniversario= An updated review of the concept of eLearning: tenth anniversary*, pages 119–144.
- García, M. A. A. (2014). Retroalimentación en educación en línea: una estrategia para la construcción del conocimiento. *RIED. Revista Iberoamericana de Educación a Distancia*, 17(2):59–73.
- Giraldo, A. F. (2015). Estrategias multi-etiqueta orientadas a la predicción funcional de proteínas en organismos embryophyta. Master’s thesis, Universidad Nacional de Colombia.
- Harris, L. R., Brown, G. T., and Harnett, J. A. (2015). Analysis of New Zealand primary and secondary student peer- and self-assessment comments: applying Hattie and Timperley’s feedback model. *Assessment in Education: Principles, Policy and Practice*, 22(2):265–281.
- Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1):81–112.
- Hernández, S. C. (2007). El constructivismo social como apoyo al aprendizaje en línea. *Apertura*, 7(7):46–62.
- Herrera, F., Charte, F., Rivera, A. J., and Del Jesus, M. J. (2016). *Multilabel classification*. Springer.
- Hutter, F., Lücke, J., and Schmidt-Thieme, L. (2015). Beyond manual tuning of hyperparameters. *KI-Künstliche Intelligenz*, 29(4):329–337.
- Jaramillo, S. and Cuasquer, V. (2013). Comparativo entre los sistemas de gestión de aprendizaje moodle y maat gknowledge. *INGE CUC*, 9(1):183–195.
- Kasim, N. N. M. and Khalid, F. (2016). Choosing the right learning management system (lms) for the higher education institution context: A systematic review. *International Journal of Emerging Technologies in Learning*, 11(6):55–61.
- Kesavaraj, G. and Sukumaran, S. (2013). A study on classification techniques in data mining. In *2013 fourth international conference on computing, communications and networking technologies (ICCCNT)*, pages 1–7. IEEE.
- Kwon, K., Park, S. J., Shin, S., and Chang, C. Y. (2019). Effects of different types of instructor comments in online discussions. *Distance Education*, 40(2):226–242.
- Mantovani, R. G., Rossi, A. L., Vanschoren, J., Bischl, B., and Carvalho, A. C. (2015). To tune or not to tune: recommending when to adjust svm hyper-parameters via meta-learning. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

- 
- Market, L. (2020). Lms market by component (solution and services), delivery mode (distance learning, instructor-led training and blended learning), deployment type, user type (academic and corporate), and region - global forecast to 2025. Technical report, MarketsandMarkets.
- Martínez, F. G. L. and Vargas, L. A. T. (2014). Retroalimentación formativa para estudiantes de educación a distancia. *RIED. Revista Iberoamericana de Educación a Distancia*, 17(2):197–221.
- Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., and Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology*, 50(1):128–138.
- Probst, P., Wright, M. N., and Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3):e1301.
- Quesada, R. (2006). Evaluación del aprendizaje en la educación a distancia “en línea”. *Revista de educación a distancia*, 5(6):1–15.
- Ramírez, G. and Valdés, D. E. (2019). El modelo de retroalimentación de Hattie y Timperley como estrategia para favorecer el cambio en las percepciones sobre la evaluación formativa en docentes y alumnos. *Revista de Investigación Educativa del Tecnológico de Monterrey*, 10(19):75–87.
- Rivera, P., Alonso, C., and Sancho, J. M. (2017). Desde la educación a distancia al e-learning: emergencia, evolución y consolidación. *Revista educación y tecnología*, 1(10):1–13.
- Salvat, B. G. (2018). La evolución del e-learning: del aula virtual a la red. *RIED. Revista Iberoamericana de Educación a Distancia*, 21(2):69–82.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1):153–189.
- Sorower, M. S. (2010). A literature survey on algorithms for multi-label learning. *Oregon State University, Corvallis*, 18:1–25.
- Uribe, S. N. and Vaughan, M. (2017). Facilitating student learning in distance education: a case study on the development and implementation of a multifaceted feedback system. *Distance Education*, 38(3):288–301.
- Van der Kleij, F. M., Feskens, R. C., and Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students’ learning outcomes: A meta-analysis. *Review of educational research*, 85(4):475–511.

- 
- Velazco, S. Y., Abuchar, A., Castilla, I., and Rivera, K. (2017). e-learning: Rompiendo fronteras. *Redes de Ingeniería*, pages 91–100.
- Vives, T. and Varela, M. (2013). Realimentación efectiva. *Investigación en educación médica*, 2(6):112–114.
- Weerts, H. J., Mueller, A. C., and Vanschoren, J. (2020). Importance of tuning hyperparameters of machine learning algorithms. *arXiv preprint*, pages 1–17.
- Wisniewski, B., Zierer, K., and Hattie, J. (2020). The power of feedback revisited: a meta-analysis of educational feedback research. *Frontiers in Psychology*, 10:e3087.
- Zhang, M. L. and Zhou, Z. H. (2007). Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition*, 40(7):2038–2048.

---

# Capítulo 7

## Anexo A

Asignatura	No. Estudiantes	Comentarios
Amparo II	14	73
Amparo II	18	80
Amparo Penal	23	238
Análisis de Casos	6	11
Análisis de Casos	16	30
Análisis de Casos	21	170
Argumentación Jurídica	30	26
Contratos Civiles	11	44
Contratos Civiles	30	292
Contratos Mercantiles	18	294
Controversias Constitucionales y Acciones de Inconstitucionalidad	5	7
Criminalística	12	61
Criminología	17	95
Delitos Especiales	15	147
Derecho Administrativo	15	47
Derecho Administrativo	26	90
Derecho Agrario	13	15
Derecho Agrario	28	24
Derecho Concursal	11	173
Derecho Contractual y Soc.	15	85
Derecho de Ejecución de las P.	15	104
Derecho de Ejecución de las P.	25	204
Derecho de la Seguridad Social	28	141
Derecho de la Seguridad Social	29	241
Derecho Económico	34	126
Derecho Financiero	17	48
Derecho Fiscal	29	105

<b>Asignatura</b>	<b>No. Estudiantes</b>	<b>Comentarios</b>
Derecho Internacional Privado	28	19
Derecho Internacional Publico	37	57
Derecho Jurisprudencial	7	51
Derecho Mercantil	33	17
Derecho Municipal	32	63
Derecho Notarial y Registral	31	130
Derecho Procesal Constitucional	13	98
Derecho Procesal del Trabajo	23	50
Derecho Procesal del Trabajo	24	71
Derecho Procesal Electoral	12	106
Derecho Procesal Mercantil	18	0
Derecho Procesal Mercantil	29	0
Derecho Procesal Penal	17	27
Derecho Procesal Penal	32	75
Derecho Romano	32	31
Derechos Humanos Cultura y Democracia	22	79
Derechos Humanos Cultura y Democracia	23	152
Derechos Humanos Cultura y Democracia	24	162
Derechos Humanos Cultura y Democracia	23	167
Derechos Humanos Cultura y Democracia	24	168
Derechos Humanos Cultura y Democracia	23	171
Derechos Humanos Cultura y Democracia	24	193
Derechos Humanos Cultura y Democracia	21	226
DHPC	8	50
DHPC	11	69
DHPC	25	128
DHTIC	13	34
DHTIC	10	42
Dogmática Jurídica Penal (P.G)	17	12
Dogmática Jurídica Penal (P.G)	29	22
Dogmática Jurídica Penal (P.G)	28	27
Familia y Sucesiones	20	77
Familia y Sucesiones	29	119
Familia y Sucesiones	31	216
Filosofía del Derecho	34	437
Formación Humana y Social	10	18
Formación Humana y Social	14	42
Formación Humana y Social	10	62
Formación Humana y Social	12	88
Formación Humana y Social	10	92

<b>Asignatura</b>	<b>No. Estudiantes</b>	<b>Comentarios</b>
Formación Humana y Social	30	168
Formación Humana y Social	30	194
Historia del Derecho	20	76
Historia del Derecho	23	88
Historia del Derecho	22	180
Historia del Derecho	24	187
Historia del Derecho	24	220
Historia del Derecho	23	296
Historia del Derecho	24	322
Historia del Derecho	23	399
Introducción a las Ciencias Sociales	23	39
Introducción a las Ciencias Sociales	20	52
Introducción a las Ciencias Sociales	24	68
Introducción a las Ciencias Sociales	23	88
Introducción a las Ciencias Sociales	23	93
Introducción a las Ciencias Sociales	24	101
Introducción a las Ciencias Sociales	21	108
Introducción a las Ciencias Sociales	24	116
Introducción al Derecho	24	26
Introducción al Derecho	22	66
Introducción al Derecho	23	69
Introducción al Derecho	23	95
Introducción al Derecho	23	98
Introducción al Derecho	24	102
Introducción al Derecho	24	120
Introducción al Derecho	20	121
Medicina Forense	28	150
Medios Alternos de Solución de Controversias	15	73
Métodos y Técnicas de Inv. en las C.S	30	18
Obligaciones	37	99
Obligaciones	17	236
Personas y Bienes	33	119
Relaciones Individuales de Trabajo	31	34
Relaciones Individuales de Trabajo	20	73
Relaciones Individuales de Trabajo	28	150
Sistema Constitucional Mexicano	32	214
Sistemas Jurídicos Contemporáneos	19	43
Sistemas Jurídicos Contemporáneos	30	62
Sistemas Jurídicos Contemporáneos	23	150

---

<b>Asignatura</b>	<b>No. Estudiantes</b>	<b>Comentarios</b>
Técnicas de Expresión Oral y Escrita	23	63
Técnicas de Expresión Oral y Escrita	20	86
Técnicas de Expresión Oral y Escrita	24	163
Técnicas de Expresión Oral y Escrita	23	170
Técnicas de Expresión Oral y Escrita	23	172
Técnicas de Expresión Oral y Escrita	24	195
Técnicas de Expresión Oral y Escrita	22	197
Técnicas de Expresión Oral y Escrita	24	245
Teoría de la Constitución	30	105
Teoría de la Constitución	30	114
Teoría de la Constitución	29	191
Teoría del Estado	32	36
Teoría del Proceso	21	4
Teoría del Proceso	23	5
Victimología	21	151
<b>Total</b>		<b>13429</b>
<b>Promedio</b>	<b>22.35</b>	<b>110.98</b>