



BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE PUEBLA

Facultad de Ciencias de la Computación

**“Modelo de aprendizaje automático para el indicador de
bienestar en el marco de la
organización para la cooperación y el desarrollo
económico”**

TESIS

QUE PARA OBTENER EL GRADO DE

**LICENCIADO EN INGENIERÍA EN CIENCIAS DE LA
COMPUTACIÓN**

PRESENTA

ERVEY GUERRERO GÓMEZ

DIRECTOR DE TESIS

MARÍA TERESA TORRIJOS MUÑOZ

OCTUBRE 2024

El cambio es la ley de la vida. Aquellos que solo miran al pasado o al presente, están seguros de perder el futuro. – Jhon F. Kennedy

Índice

Introducción	1
Capítulo 1 – Comprensión del problema	4
1. Introducción al Capítulo.....	4
2. Contexto Socioeconómico y Político de México	5
3. Definición del Problema	6
4. Objetivos de la Investigación	8
5. Modelado de Procesos de Negocio Relacionados con el Bienestar	8
6. Stakeholders Clave y Sus Expectativas	9
7. Necesidades y Requisitos de la Solución Propuesta	10
8. Impacto Potencial del Estudio	11
9. Conclusión del Capítulo.....	12
Capítulo 2 – Enfoque Analítico	14
1. Introducción.....	14
2. Enfoque Analítico	15
3. Criterios para la Selección y Evaluación de Modelos	16
4. Contribución del Enfoque Analítico al Entendimiento del Bienestar	18
5. Conclusión	19
Capítulo 3 – Ingeniería de datos	21
1. Recolección de Datos	21
2. Preparación y limpieza de datos	22
2.1. Normalización de Periodos	22
2.2. Manejo de datos faltantes.....	22
3. Estandarización de Datos.....	23
4. Integración de Datos	23
5. Ecuación del bienestar.....	24
6. Selección de modelos.....	26
7. Conclusión	28
Capítulo 4 – Modelado	30
1. Introducción	30
2. Escenarios	30
3. Conclusiones.....	41
Capítulo 5 – Conclusiones y recomendaciones	42

1. Resumen de Hallazgos Principales.....	42
2. Implicaciones de la Investigación.....	43
3. Limitaciones del Estudio.....	44
4. Recomendaciones para Futuras Investigaciones	44
5. Contribuciones al Conocimiento	45
6. Reflexiones Finales.....	46
Apéndice A: Detalles de Implementación.....	48
Bibliografía	50

Introducción

Esta tesis aplica la metodología IBM de Ciencia de Datos para analizar el bienestar en México, considerando múltiples dimensiones socioeconómicas y ambientales. Esta metodología estructurada, reconocida por su enfoque sistemático, se organiza en cinco fases clave que guían desde la conceptualización inicial del estudio hasta la evaluación final de los resultados, y cada una de estas fases constituye un capítulo específico dentro de la tesis.

1. **Definición del Problema:** Este capítulo establece el contexto y los objetivos del estudio, definiendo claramente el problema a investigar —en este caso, el bienestar en México— y delineando las preguntas de investigación. Es crucial para alinear el estudio con las necesidades reales y las expectativas de los stakeholders, incluyendo académicos, políticos y la sociedad en general.
2. **Comprensión de los Datos:** Este capítulo identifica, evalúa y selecciona las fuentes de datos apropiadas. Se ha hecho un uso extensivo de datos del INEGI, asegurando que son relevantes, confiables y suficientes para abordar las preguntas de investigación. También se realiza un análisis preliminar para obtener una comprensión básica de las estructuras de datos, su calidad y potenciales limitaciones.
3. **Preparación de Datos:** Aquí se abordan todas las actividades necesarias para transformar los datos crudos en un formato adecuado para el análisis. Incluye la limpieza de datos, la gestión de valores faltantes, la normalización de variables y la transformación de datos según sea necesario. La preparación meticulosa es fundamental para asegurar la precisión de los análisis subsiguientes.
4. **Modelado:** Utilizando herramientas avanzadas de ciencia de datos, este capítulo desarrolla modelos predictivos o descriptivos basados en los datos preparados. Se emplea una ecuación de bienestar desarrollada en investigaciones previas y validada a través de regresiones lineales, con ajustes y optimizaciones realizadas mediante técnicas como la validación cruzada y la búsqueda de hiperparámetros.

5. **Evaluación:** El último capítulo evalúa rigurosamente los modelos para verificar su validez y efectividad, comparando los resultados con los objetivos iniciales del estudio para determinar el grado de éxito. También se consideran las implicaciones prácticas de los hallazgos, evaluando cómo pueden influir en las políticas públicas y la planificación social para mejorar el bienestar en México.

El fundamento de este trabajo se basa en una investigación anterior donde se desarrolló una ecuación específica para evaluar el bienestar. Este modelo, que integra variables clave como educación, empleo, vivienda y seguridad, fue validado utilizando una regresión lineal, demostrando su capacidad para cuantificar de manera precisa la calidad de vida en México. La ecuación del bienestar propuesta no solo refleja un análisis multidimensional, sino que también establece un marco cuantitativo para las políticas públicas y la planificación social.

A través de este enfoque estructurado, este estudio profundiza en el entendimiento del bienestar en México y resalta la eficacia de aplicar metodologías detalladas para resolver problemas sociales complejos. El objetivo es ampliar el conocimiento académico y ofrecer recomendaciones prácticas que puedan mejorar sustancialmente la calidad de vida en el país.

Antecedentes

En una investigación previa, se realizó un análisis del bienestar en México utilizando ciencia de datos, donde se definió el bienestar a partir de dos dimensiones: objetiva y subjetiva, centrándose principalmente en la dimensión objetiva (Guerrero Gómez & Torrijos Muñoz, 2024).

Se investigaron y descargaron datos del INEGI y la OECD sobre los factores clave del bienestar, tales como vivienda, ingresos, empleo, seguridad, medio ambiente, educación, salud y satisfacción con la vida. Posteriormente, se unificaron estas bases de datos, ajustando las diferencias en los periodos y completando datos faltantes para garantizar la paridad.

Se construyó e implementó un modelo de regresión lineal utilizando los datos tratados para predecir el bienestar en función de los factores clave. Este modelo logró un ajuste casi perfecto, con un coeficiente de determinación R^2 de 0.9999997, lo que significa que el 99.99% de la variabilidad del bienestar fue explicada por los factores analizados. Además, las métricas de error, como el MSE y RMSE, mostraron que el modelo tenía una alta precisión.

Finalmente, se concluyó que el modelo es altamente efectivo, pero se sugiere implementar en el futuro otros modelos más complejos, como árboles de decisión o redes neuronales, para capturar interacciones no lineales entre los factores y mejorar la predicción del bienestar en escenarios más complejos.

Capítulo 1 – Comprensión del problema

1. Introducción al Capítulo

Este primer capítulo establece el marco para explorar el bienestar en México a través del prisma de la ciencia de datos, resaltando la importancia de abordar las disparidades socioeconómicas y políticas desde una perspectiva analítica avanzada. Enfrentando desafíos como la pobreza, la desigualdad y la inseguridad, este estudio busca desarrollar un modelo predictivo basado en ciencia de datos para ofrecer soluciones innovadoras que mejoren el bienestar mexicano. Al centrarse específicamente en la intersección entre la ciencia de datos y los indicadores de bienestar, la investigación tiene como objetivo descubrir patrones y tendencias ocultos que pueden informar la formulación de políticas públicas más efectivas.

Una parte crucial de este estudio es la participación de una variedad de stakeholders, incluyendo organismos gubernamentales, organizaciones no gubernamentales (ONGs), y la comunidad académica. Su colaboración no solo enriquecerá la investigación con perspectivas diversas, sino que también garantizará que los hallazgos sean aplicables y relevantes a las necesidades actuales de México. Se discutirá cómo la integración de sus expectativas y experiencias puede contribuir al éxito y la aplicabilidad de los modelos predictivos desarrollados.

Finalmente, se evaluará el impacto potencial de esta investigación, anticipando cómo los conocimientos obtenidos podrían influir positivamente en la formulación y ajuste de políticas y prácticas enfocadas en el bienestar. Se argumentará que el uso estratégico de la ciencia de datos ofrece una oportunidad sin precedentes para abordar de manera más efectiva problemas persistentes como la pobreza, la desigualdad y la inseguridad. Al concluir este capítulo, se habrán sentado las bases para una exploración profunda de cómo la ciencia de datos puede no solo diagnosticar los desafíos actuales del bienestar en México, sino también identificar oportunidades para promover un desarrollo inclusivo y sostenible.

2. Contexto Socioeconómico y Político de México

México se encuentra en una encrucijada de desafíos y oportunidades que moldean el bienestar de su población. Con una economía que es simultáneamente robusta y marcada por desigualdades profundas, el país presenta un panorama complejo donde el progreso económico no se distribuye equitativamente entre sus ciudadanos. Este panorama se ve reflejado en la reciente estimación de crecimiento económico: al cierre de 2023, la economía mexicana creció un 3.1% respecto a 2022 y un 0.1% real trimestral respecto al trimestre previo, alcanzando un nuevo máximo histórico según la Estimación Oportuna del PIB (INEGI). A pesar de este crecimiento, diez estados de la República mostraron un crecimiento anual superior al 4.5%, evidenciando la variabilidad en la distribución del progreso económico a lo largo del territorio nacional (México, ¿cómo Vamos?, 30 de enero de 2024).

Esta disparidad se refleja en la variabilidad del acceso a servicios básicos como la salud, la educación y la vivienda, contribuyendo a diferencias significativas en la calidad de vida. Políticamente, México ha experimentado importantes transformaciones que buscan responder a las demandas de una sociedad cada vez más informada y participativa. Sin embargo, los esfuerzos por mejorar el bienestar a menudo se encuentran con obstáculos que van desde la corrupción y la inseguridad hasta la falta de políticas públicas eficientes que aborden las necesidades de los más vulnerables. Estos factores políticos y económicos, combinados, crean un entorno donde el bienestar se convierte en un indicador clave para evaluar tanto el progreso social como la efectividad de la gobernanza.

La respuesta a estos retos no es solo una cuestión de política pública, sino también de comprensión profunda de las dinámicas sociales y económicas que definen la vida en México. El análisis del bienestar, por lo tanto, debe considerar una amplia gama de indicadores, desde el ingreso y la distribución de la riqueza hasta aspectos más subjetivos y cualitativos del bienestar, como la satisfacción con la vida y el acceso a oportunidades para el desarrollo personal.

Además, en este complejo escenario socioeconómico y político, la ciencia de datos emerge como una herramienta poderosa para abordar y entender de manera más

efectiva los desafíos relacionados con el bienestar. La capacidad de recopilar, analizar e interpretar grandes volúmenes de datos puede proporcionar conocimiento valioso sobre las condiciones de vida, las disparidades económicas, y los efectos de las políticas públicas en diferentes sectores de la población. La aplicación de técnicas avanzadas de análisis de datos no solo facilita la identificación de patrones y tendencias, sino que también permite predecir posibles escenarios futuros, ofreciendo así una base sólida para la toma de decisiones informadas y la formulación de estrategias más efectivas para mejorar el bienestar

El uso de la ciencia de datos en el ámbito socioeconómico y político de México tiene el potencial de transformar la manera en que se abordan problemas como la pobreza, la desigualdad, y la inseguridad. Por ejemplo, Adolfo de Unanue, experto del Tecnológico de Monterrey, resalta la necesidad de unir esfuerzos entre especialistas en Inteligencia Artificial (IA) y tomadores de decisiones para impactar significativamente en el desarrollo social. Esta colaboración es fundamental para explotar el potencial de la IA en la generación de políticas de bienestar efectivas, subrayando cómo la tecnología y el conocimiento especializado pueden servir como herramientas clave en la formulación de estrategias que aborden directamente los desafíos sociales de México (Torres, I., 27 de septiembre de 2023)

Este enfoque interdisciplinario permite no solo una evaluación precisa de las necesidades y problemas de diferentes grupos poblacionales sino también la implementación de intervenciones más dirigidas y personalizadas. Además, la optimización de las políticas públicas a través del análisis de datos promete una asignación más eficiente de los recursos, apuntando a una gestión del bienestar más innovadora y efectiva en México.

3. Definición del Problema

En el corazón de esta investigación yace el desafío multifacético del bienestar en México, un problema que se extiende más allá de las métricas económicas tradicionales para abarcar una amplia gama de factores sociales, de salud, educativos, ambientales y económicos. A pesar de los esfuerzos realizados en las últimas décadas, el país sigue enfrentando obstáculos significativos para asegurar un nivel de bienestar equitativo y

sostenible para todos sus habitantes. Las disparidades regionales, la desigualdad socioeconómica, y los desafíos en el acceso a servicios básicos son solo algunos de los aspectos que complican el panorama del bienestar en México.

La complejidad de estos problemas requiere de un enfoque que trascienda los métodos de análisis tradicionales, uno que pueda manejar la diversidad y la gran escala de los datos involucrados para ofrecer soluciones innovadoras y efectivas. Aquí es donde la ciencia de datos se presenta como una herramienta indispensable, permitiendo una aproximación más granular y precisa al estudio del bienestar. La capacidad de la ciencia de datos para manejar y analizar conjuntos de datos grandes y complejos ofrece una oportunidad sin precedentes para comprender las múltiples dimensiones del bienestar en una escala y con una profundidad que los métodos tradicionales no pueden alcanzar.

Esta investigación pretende aprovechar el potencial de la ciencia de datos para ir más allá de los análisis estadísticos convencionales, empleando técnicas avanzadas de aprendizaje automático y minería de datos para identificar factores predictivos del bienestar. Se busca desarrollar un modelo predictivo que no solo cuantifique los efectos de diversos factores socioeconómicos y políticos sobre el bienestar, sino que también ofrezca predicciones precisas sobre cómo podrían evolucionar estas dinámicas bajo diferentes escenarios futuros. Este enfoque innovador permite explorar la efectividad de las políticas públicas y las intervenciones sociales de una manera completamente nueva, evaluando su impacto potencial antes de su implementación y proporcionando una base sólida para la toma de decisiones basada en evidencia.

Además, al integrar la ciencia de datos en el estudio del bienestar, esta tesis no solo aborda un desafío social crítico, sino que también marca el camino hacia una comprensión más profunda y matizada del bienestar en México. A través de la aplicación de métodos de análisis avanzados, se espera no solo contribuir al conocimiento académico sino también ofrecer soluciones prácticas que puedan guiar la acción política y social hacia un futuro más próspero y equitativo para todos los mexicanos. Al desarrollar y compartir modelos predictivos accesibles, esta investigación tiene el potencial de empoderar a una amplia gama de actores, desde responsables de políticas hasta organizaciones comunitarias, con herramientas analíticas avanzadas, fomentando

así una participación más amplia en el proceso de mejora continua de las condiciones de vida en México.

4. Objetivos de la Investigación

Este estudio sobre el bienestar en México, apoyado en las capacidades de la ciencia de datos, se estructura en torno a objetivos específicos diseñados para ofrecer una comprensión integral del tema:

Analizar las Dimensiones del Bienestar: Este objetivo se centra en explorar y cuantificar las variadas facetas del bienestar en México, abarcando elementos económicos, sociales, de salud y educativos mediante el uso de análisis estadístico y visualización de datos. Se identificarán los determinantes clave del bienestar y sus interacciones, proporcionando una visión holística de sus múltiples dimensiones.

Desarrollar un Modelo Predictivo de Bienestar: Se emplearán técnicas de aprendizaje automático y análisis estadístico avanzado para construir un modelo que prediga el bienestar en México, integrando un amplio espectro de indicadores y variables. Este modelo ofrecerá una herramienta analítica precisa para informar la creación de políticas públicas.

Ofrecer Recomendaciones Basadas en Datos para la Mejora del Bienestar: A partir de los análisis realizados, se elaborará un conjunto de recomendaciones prácticas y basadas en evidencia para responsables de políticas, ONGs y otros actores clave, enfocadas en intervenciones efectivas para elevar el bienestar.

Contribuir al Conocimiento Académico en Ciencia de Datos y Bienestar: Este estudio busca ampliar el conocimiento en la intersección entre la ciencia de datos y el bienestar, mostrando la aplicación de metodologías analíticas avanzadas en la solución de problemas sociales complejos. Se pretende compartir los hallazgos y metodologías a través de publicaciones académicas y foros especializados.

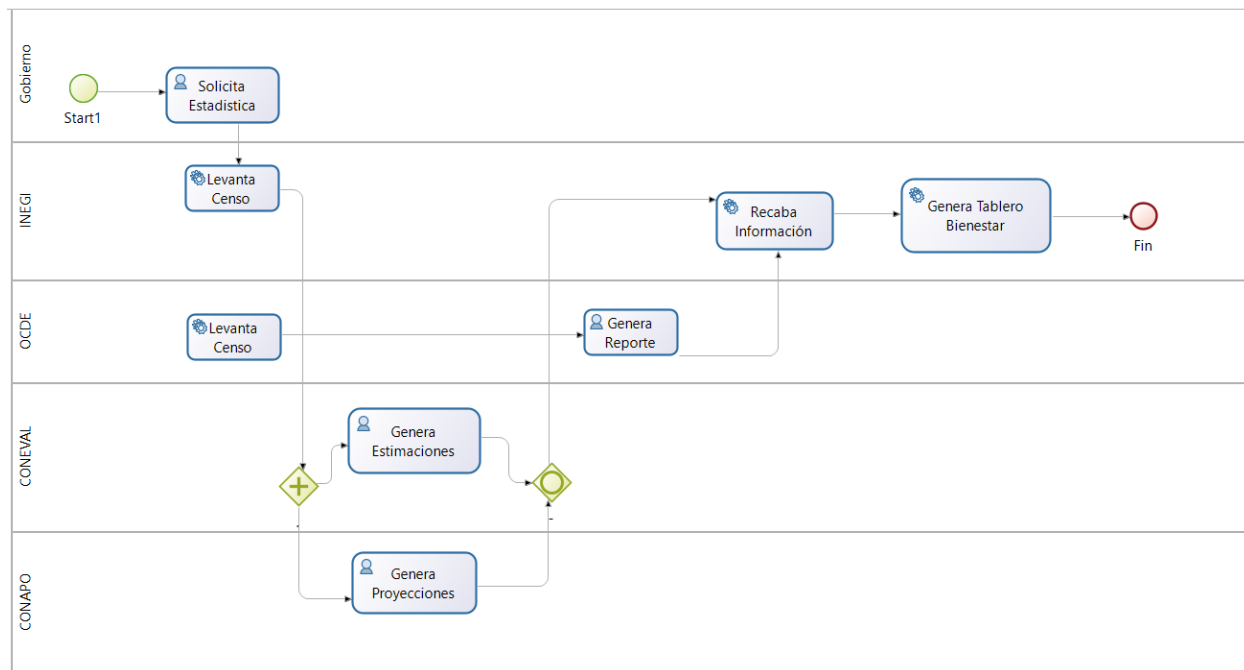
Al alcanzar estos objetivos, la investigación no solo busca comprender profundamente el bienestar en México sino también proporcionar herramientas analíticas y recomendaciones concretas que guíen hacia un futuro más equitativo y próspero para la población mexicana.

5. Modelado de Procesos de Negocio Relacionados con el Bienestar

Este segmento del estudio detalla el proceso mediante el cual los datos utilizados en la investigación fueron obtenidos, desde la solicitud inicial hasta su recopilación y almacenamiento y presentación final por parte del INEGI. Este proceso es crucial para entender la calidad y la fiabilidad de los datos que sustentan el modelo desarrollado.

Figura 1

Proceso de negocio para la obtención de los datos



6. Stakeholders Clave y Sus Expectativas

La presente investigación se enmarca en un contexto donde múltiples actores juegan roles fundamentales en el ámbito del bienestar social en México. Identificar a estos stakeholders clave y comprender sus expectativas específicas es esencial para garantizar que los resultados del estudio sean pertinentes y aplicables. Los principales interesados en este estudio incluyen organismos gubernamentales, organizaciones no gubernamentales (ONGs), la comunidad académica y la población en general, cada uno con expectativas únicas respecto a los hallazgos y aplicaciones del modelo predictivo basado en la ciencia de datos.

Organismos Gubernamentales: Estas entidades buscan fundamentar sus políticas y programas en evidencia sólida para maximizar el bienestar social. La expectativa de estos organismos es que el modelo predictivo proporcione análisis precisos y detallados que identifiquen necesidades críticas y poblaciones objetivo, facilitando así la asignación estratégica de recursos y la evaluación de impacto de las intervenciones públicas.

Organizaciones No Gubernamentales (ONGs): Comprometidas con la implementación de proyectos que aborden directamente las necesidades del bienestar, las ONGs valoran la capacidad del modelo para ofrecer perspectivas predictivas que orienten sus acciones. Esperan utilizar los resultados para priorizar iniciativas, optimizar el alcance y la efectividad de sus programas, y demostrar el impacto de sus esfuerzos.

Comunidad Académica: La expectativa se centra en la ampliación del conocimiento existente sobre la aplicación de la ciencia de datos en la investigación social, proporcionando un marco que pueda ser replicado o adaptado en estudios futuros.

Población en General: Aunque indirectamente, el bienestar de la población mexicana es el beneficiario final de este estudio. La expectativa es que las recomendaciones basadas en la investigación fomenten políticas y programas más efectivos y eficientes, conduciendo a mejoras concretas en la calidad de vida y el bienestar general.

Para satisfacer estas expectativas, esta tesis adopta un enfoque integrador y colaborativo, asegurando una alineación estrecha entre los objetivos de investigación y las necesidades de los stakeholders. El compromiso con la relevancia práctica y la aplicabilidad de los resultados es fundamental, y se busca establecer un diálogo continuo con estos actores a lo largo del proceso investigativo. De este modo, la investigación no solo avanza en el terreno académico, sino que también provee un marco aplicado para informar y guiar la mejora del bienestar en México, demostrando el valor tangible de la ciencia de datos en la solución de desafíos sociales complejos.

7. Necesidades y Requisitos de la Solución Propuesta

La solución propuesta en esta investigación, un modelo predictivo del bienestar en México está diseñada para satisfacer necesidades críticas y cumplir con requisitos específicos que aseguran su utilidad y efectividad:

Precisión y Fiabilidad: Fundamental para el modelo es su capacidad para ofrecer predicciones con alta precisión y fiabilidad. Mediante el uso de técnicas de aprendizaje automático y análisis estadístico avanzado, se implementará un riguroso proceso de

validación, empleando métricas como la precisión, la sensibilidad y la especificidad para evaluar su consistencia y confiabilidad.

Capacidad de Integración de Datos Multidimensionales: Dado el carácter complejo del bienestar, el modelo integrará datos de múltiples dimensiones -económicas, sociales, de salud y educativas- utilizando plataformas de análisis de datos y algoritmos capaces de procesar esta diversidad, garantizando una arquitectura adaptable a variadas estructuras de datos.

Interpretabilidad y Transparencia: Se priorizará que el modelo sea interpretable y transparente para facilitar su adopción por parte de tomadores de decisiones. Estrategias como la visualización de datos y la simplificación de resultados serán claves para hacer que las predicciones sean accesibles y accionables.

Sostenibilidad y Escalabilidad: La solución será diseñada considerando la sostenibilidad a largo plazo y la capacidad de escalar. Esto se logrará mediante la planificación de recursos para actualizaciones regulares y adaptabilidad frente a nuevos datos o cambiantes necesidades sociales, asegurando que el modelo permanezca relevante y eficaz.

Impacto Social Positivo: El objetivo primordial es que el modelo genere un impacto social beneficioso, con recomendaciones basadas en evidencia que orienten políticas y estrategias para un desarrollo más inclusivo y equitativo.

Cumpliendo con estos requisitos, el modelo no solo contribuirá al avance académico en la intersección de la ciencia de datos y el bienestar, sino que también proveerá una herramienta pragmática para enfrentar desafíos sociales significativos en México.

8. Impacto Potencial del Estudio

Este estudio tiene el potencial de generar un impacto significativo más allá de los límites académicos, influyendo en la formulación de políticas públicas y la implementación de prácticas que promuevan el bienestar en México. La integración de la ciencia de datos en el análisis del bienestar abre nuevos caminos para comprender y enfrentar los desafíos sociales de manera más eficaz y eficiente.

En Políticas Públicas: El modelo predictivo desarrollado tiene el potencial de revolucionar cómo los organismos gubernamentales diseñan e implementan políticas de bienestar. Al proporcionar predicciones precisas sobre los determinantes del bienestar, se espera que el modelo sirva como una herramienta esencial para la toma de decisiones basada en evidencia, permitiendo a los responsables de políticas identificar áreas prioritarias de intervención y optimizar la distribución de recursos. Por ejemplo, podría guiar la implementación de programas de salud y educación, ajustándolos a las necesidades proyectadas de diferentes comunidades.

En Prácticas Sociales: Las ONGs y otros actores sociales pueden aprovechar el conocimiento generado por el estudio para mejorar sus estrategias y programas. El modelo predictivo se convierte en una herramienta clave para evaluar necesidades y medir el impacto de las intervenciones, apoyando una gestión más enfocada y orientada a resultados en la lucha contra los desafíos del bienestar.

En el Campo Académico: Este estudio enriquece significativamente el conocimiento sobre la aplicación de la ciencia de datos en las ciencias sociales. Al demostrar el valor de métodos avanzados de análisis y modelado predictivo, se establece un modelo para investigaciones futuras, promoviendo un enfoque interdisciplinario y el desarrollo de nuevas metodologías de investigación. Se espera que los resultados inspiren a investigadores y académicos a explorar el bienestar desde perspectivas novedosas, contribuyendo así a una comprensión más amplia de los elementos que fomentan una sociedad equitativa.

9. Conclusión del Capítulo

Este capítulo ha sentado las bases para abordar el bienestar en México mediante la ciencia de datos, destacando la complejidad del desafío y el potencial de un modelo predictivo avanzado para ofrecer soluciones novedosas. Al definir objetivos claros y reconocer la importancia de colaborar con stakeholders clave, este estudio promete no solo avanzar el conocimiento académico sino también influir positivamente en la formulación de políticas públicas y prácticas sociales. La investigación se posiciona para transformar la comprensión y mejora del bienestar, proporcionando una herramienta práctica y basada en evidencia que responde a las necesidades críticas y cumple con

requisitos de precisión, integración de datos, interpretabilidad, y sostenibilidad. Con un compromiso hacia el impacto social positivo, este estudio se dirige hacia el desarrollo de recomendaciones que guíen hacia un México más equitativo y próspero.

Capítulo 2 – Enfoque Analítico

1.Introducción

Al abordar el bienestar en México, este estudio reconoce la complejidad inherente a la medición y predicción de un fenómeno tan multifacético. La transición de una comprensión conceptual del bienestar hacia su formulación como un problema analítico subraya la intención de aplicar métodos rigurosos y sistemáticos para desentrañar los factores que lo influyen. En este capítulo, se propone definir con claridad el problema de investigación desde una perspectiva analítica, destacando la relevancia de emplear la ciencia de datos para predecir el bienestar y, por extensión, su impacto potencial en la formulación e implementación de políticas públicas eficaces.

El bienestar, a pesar de ser un objetivo deseable y universalmente reconocido, presenta desafíos significativos en términos de medición debido a su naturaleza subjetiva y a la diversidad de factores que lo afectan. La predicción del bienestar, por tanto, exige un enfoque multidimensional que considere variables económicas, sociales, de salud y educativas, así como una metodología capaz de manejar la complejidad y el volumen de datos asociados con estos factores. Aquí radica la importancia de adoptar un enfoque analítico basado en la ciencia de datos, que permite la integración y análisis de grandes conjuntos de datos para identificar patrones, tendencias y relaciones que serían imposibles de discernir mediante métodos tradicionales.

El uso de modelos de machine learning específicos se justifica en este contexto por su capacidad para procesar y analizar datos complejos, ofreciendo la posibilidad de predecir el bienestar con una precisión y profundidad sin precedentes. Estos modelos, al ser alimentados con datos adecuadamente seleccionados y preparados, tienen el potencial de revelar conocimiento crítico sobre los determinantes del bienestar y cómo estos pueden ser influenciados por políticas públicas. Así, la ciencia de datos no solo sirve como una herramienta para el análisis, sino que se convierte en un puente hacia soluciones prácticas que pueden mejorar la vida de los mexicanos.

2. Enfoque Analítico

La selección de un enfoque analítico adecuado constituye la piedra angular de la investigación sobre el bienestar en México, especialmente en lo que respecta a la identificación y predicción de sus determinantes. Para abordar eficazmente este desafío, este estudio se centrará en la implementación de un árbol de decisión y una Máquina de Vectores de Soporte (SVM), los cuales serán integrados mediante una técnica de ensamblaje conocida como stacking. Esta combinación ha sido elegida por su potencial para capturar la complejidad y las sutilezas del bienestar a través de un modelo predictivo robusto y preciso.

Arboles de Decisión

Los árboles de decisión modelan decisiones y sus posibles consecuencias como un árbol con nodos y hojas. Cada nodo en el árbol representa una característica o atributo, y las ramas representan las decisiones que llevan a diferentes resultados. Al construir un árbol de decisión para el bienestar, el algoritmo comienza en la raíz del árbol y se divide en función de la variable que ofrece la mayor ganancia de información (usualmente medida a través de índices como la entropía o el índice Gini). Este proceso de división se repite en cada nuevo nodo hasta que se cumplen ciertos criterios de parada, como un número mínimo de observaciones en cada hoja o que no se gane información adicional con más divisiones. Esta estructura jerárquica facilita la interpretación de cómo las variables como ingresos o educación influyen directamente en el bienestar, mostrando un camino claro desde las variables de entrada hasta la decisión final.

Máquinas de Vectores de Soporte (SVM)

La SVM trabaja encontrando el hiperplano que mejor divide un conjunto de datos en clases. En el contexto del bienestar, consideramos cada estado de bienestar como una clase diferente. La SVM busca el hiperplano que maximiza el margen entre las diferentes clases, donde el margen se define como la distancia entre las líneas de soporte más cercanas a cada clase. En problemas no lineales, como puede ser la clasificación del bienestar basada en múltiples indicadores socioeconómicos, la SVM utiliza funciones de kernel para mapear los datos a un espacio de mayor dimensión donde es posible

encontrar un hiperplano separador. Esto permite que el clasificador maneje complejidades en los datos que no son inmediatamente separables en el espacio original.

Stacking

En el stacking, cada uno de los clasificadores base (en este caso, árboles de decisión y SVM) genera predicciones que son usadas como entradas para un modelo meta. Este modelo meta, a menudo otro algoritmo de aprendizaje supervisado, se entrena para hacer la predicción final basada en las predicciones de los modelos base. La clave del éxito de este enfoque radica en la diversidad de los modelos base; cada uno puede captar diferentes patrones en los datos, y el modelo meta aprende a optimizar estas diferentes vistas para mejorar la precisión y la robustez del modelo ensamblado. El resultado es un sistema que combina eficazmente las fortalezas individuales de cada clasificador en un marco unificado y poderoso.

La elección de este enfoque analítico refleja un compromiso con la precisión, la robustez y la interpretabilidad, elementos clave para abordar la complejidad inherente al estudio del bienestar. Al utilizar un modelo ensamblado que integra árboles de decisión con SVM a través del stacking, se espera capturar la multifaceticidad del bienestar en México y, al mismo tiempo, ofrecer predicciones confiables que puedan informar eficazmente las políticas públicas.

Este enfoque analítico no solo es indicativo de la aplicación innovadora de la ciencia de datos en cuestiones sociales complejas, sino que también establece un precedente en la investigación del bienestar, prometiendo contribuciones significativas tanto a nivel teórico como práctico.

3. Criterios para la Selección y Evaluación de Modelos

Una vez definido el enfoque analítico mediante el uso de árboles de decisión, Máquinas de Vectores de Soporte (SVM) y el ensamblaje por stacking, es crucial establecer criterios rigurosos para la selección y evaluación de los modelos generados. Estos criterios aseguran la relevancia y aplicabilidad de los modelos a nuestro estudio sobre el bienestar en México y, al mismo tiempo, garantizan la integridad y confiabilidad de los

resultados. A continuación, se describen los principales indicadores y procesos utilizados para evaluar los modelos desarrollados en esta investigación.

Precisión del Modelo y Coeficiente de Determinación R^2

La precisión general del modelo se complementa con el coeficiente de determinación R^2 para modelos de regresión, el cual ofrece una medida de cuán bien las respuestas predichas por el modelo se ajustan a los datos reales. Un R^2 cercano a 1 indica que el modelo explica una gran proporción de la variabilidad en los datos observados, siendo crucial para evaluar la eficacia de las predicciones en el contexto del bienestar.

Error Cuadrático Medio (MSE) y Raíz del Error Cuadrático Medio (RMSE)

Estas métricas son fundamentales para cuantificar la diferencia entre los valores predichos por el modelo y los valores reales. El MSE proporciona una medida promedio del cuadrado de los errores, penalizando más los errores grandes, mientras que el RMSE, al estar en la misma unidad que la variable de respuesta, ofrece una interpretación más directa de la magnitud típica del error de predicción.

Validación Cruzada

Es esencial para evaluar cómo el modelo generaliza a un conjunto de datos independiente no visto durante el entrenamiento. La validación cruzada nos ayuda a asegurar que el modelo es robusto y fiable, evitando problemas de sobreajuste y proporcionando una estimación más precisa de su capacidad predictiva.

Interpretabilidad del Modelo y Análisis de Sensibilidad

La capacidad para interpretar los resultados del modelo y entender cómo diversas entradas afectan las predicciones sigue siendo crucial. El análisis de sensibilidad complementa esta necesidad al evaluar la consistencia de las predicciones del modelo frente a variaciones en los datos de entrada, reforzando la confiabilidad de las conclusiones extraídas del modelo.

La aplicación de estas métricas avanzadas para la selección y evaluación de los modelos en el estudio del bienestar en México permite no solo capturar la efectividad del modelo en términos de ajuste y predicción sino también asegurar su relevancia y utilidad práctica.

A través de un enfoque metódico y basado en evidencia, este proceso garantiza que los modelos desarrollados proporcionen conocimiento valioso y aplicables para informar políticas y estrategias efectivas destinadas a mejorar el bienestar.

4. Contribución del Enfoque Analítico al Entendimiento del Bienestar

La adopción de un enfoque analítico avanzado, que integra árboles de decisión y Máquinas de Vectores de Soporte (SVM) a través de la metodología de stacking, marca un hito significativo en la investigación sobre el bienestar en México. Esta combinación representa tanto una innovación técnica como una aportación de una perspectiva única y profunda en el análisis del bienestar, abordando su naturaleza intrínsecamente compleja y multifacética. A continuación, se detalla cómo este enfoque contribuye sustancialmente al entendimiento del bienestar

Desentrañando la Complejidad del Bienestar

Los árboles de decisión permiten una visualización intuitiva de las relaciones entre múltiples factores y el bienestar, facilitando la identificación de patrones y la comprensión de cómo diversas variables interactúan para influir en él. Al complementar este análisis con SVM, que maneja efectivamente datos complejos y no lineales, el enfoque permite una exploración profunda y detallada de las dinámicas subyacentes del bienestar.

Predicciones Precisas y Personalizadas

Mediante el uso de stacking para ensamblar estos modelos, se mejora la precisión de las predicciones, lo que permite identificar los determinantes del bienestar y predecir sus niveles bajo diversas condiciones y escenarios. Este grado de personalización y precisión es crucial para diseñar políticas públicas y programas sociales que respondan específicamente a las necesidades y desafíos identificados.

Informando la Formulación de Políticas Públicas

La capacidad de este enfoque analítico para proporcionar conocimiento detallado y predicciones fiables sobre el bienestar tiene el potencial de transformar la forma en que se formulan e implementan las políticas públicas. Al ofrecer una base de evidencia robusta, los responsables de la toma de decisiones pueden orientar los recursos de

manera más efectiva, apuntando a intervenciones que promuevan un impacto positivo significativo en el bienestar de la población.

Estableciendo Nuevos Precedentes en la Investigación Social

Finalmente, la aplicación de estas técnicas avanzadas de aprendizaje automático en el estudio del bienestar establece un precedente importante para futuras investigaciones en el campo. Este enfoque no solo amplía el marco metodológico disponible para explorar cuestiones sociales complejas, sino que también promueve una comprensión más rica y matizada del bienestar, animando a otros investigadores a adoptar metodologías innovadoras en sus trabajos.

El enfoque analítico adoptado en este estudio enriquece nuestro entendimiento del bienestar en México y ofrece una plataforma para intervenir de manera más informada y efectiva. Al integrar técnicas estadísticas y de aprendizaje automático de vanguardia, este estudio ilumina las complejas dinámicas que configuran el bienestar, ofreciendo nuevas vías para su mejora a través de la acción política y social informada.

5. Conclusión

Este capítulo ha establecido un marco sólido para el estudio del bienestar en México mediante un enfoque analítico avanzado, que integra técnicas estadísticas y de aprendizaje automático de vanguardia como los árboles de decisión, las Máquinas de Vectores de Soporte (SVM) y el ensamblaje mediante stacking. La selección y aplicación cuidadosa de estas técnicas subraya el compromiso con la precisión y la eficacia predictiva, así como con la profundidad de la comprensión que podemos lograr sobre los complejos determinantes del bienestar.

La contribución de este enfoque al entendimiento del bienestar es multifacética. Permite desentrañar la complejidad de los factores que influyen en el bienestar, proporcionar predicciones precisas y personalizadas que son esenciales para el diseño e implementación de políticas públicas efectivas y, finalmente, avanzar en el campo de la investigación social mediante la introducción de metodologías analíticas innovadoras. Al hacerlo, este estudio enriquece el cuerpo académico existente sobre el bienestar y,

simultáneamente, ofrece herramientas prácticas y recomendaciones basadas en datos para mejorar la calidad de vida en México.

Los criterios establecidos para la selección y evaluación de modelos aseguran que los hallazgos y predicciones derivados de este enfoque sean tanto confiables como aplicables, facilitando la toma de decisiones informada y la formulación de estrategias de intervención basadas en evidencia. Además, la interpretación y el análisis de sensibilidad de los modelos refuerzan la robustez de las predicciones y mejoran la transparencia y comprensión de cómo se generan estas predicciones, aspectos cruciales para su aceptación y aplicación práctica.

En última instancia, la investigación prepara el camino para una exploración más profunda y detallada del bienestar en México, ofreciendo una base sólida sobre la cual se pueden construir futuros estudios. Este capítulo no solo destaca el valor de las técnicas de aprendizaje automático en la investigación social, sino que también ilustra el potencial transformador de la ciencia de datos en la comprensión y mejora del bienestar humano.

Al mirar hacia el futuro, este estudio anima a otros investigadores a adoptar enfoques analíticos similares y explorar nuevas aplicaciones de la ciencia de datos en el ámbito social. La intersección entre la ciencia de datos y los estudios sobre el bienestar presenta un campo rico y en gran parte inexplorado de oportunidades de investigación, cada una con el potencial de contribuir significativamente a nuestra comprensión del bienestar y a la formulación de políticas que fomenten una sociedad más equitativa y próspera.

Capítulo 3 – Ingeniería de datos

1. Recolección de Datos

El fundamento de este estudio, sobre el bienestar en México, reside en una exhaustiva recolección de datos, meticulosamente seleccionados para abarcar las dimensiones críticas que definen el bienestar. Un total de 15 bases de datos de dominio público, fueron descargadas del Instituto Nacional de Estadística y Geografía (INEGI) mediante su portal “Indicadores de Bienestar”, cada una correspondiendo a una variable que impacta directamente en la calidad de vida de los Mexicanos. Este conjunto de datos abarca desde aspectos tangibles, e intangibles como la vivienda, los ingresos, las relaciones sociales y la satisfacción con la vida. La selección se realizó con la intención de proporcionar una mirada holística al bienestar, considerando:

Vivienda: El dataset de habitaciones por persona ofrece una perspectiva fundamental sobre el espacio personal como un indicador de calidad de vida.

Ingresos: Se utilizó el dataset de ingreso disponible de los hogares per cápita para capturar la capacidad económica de estos.

Empleo: La tasa de desocupación se seleccionó para evaluar el impacto del empleo en la seguridad económica y personal.

Medio Ambiente: La contaminación del aire se incluyó como un indicador crítico de la calidad del medio ambiente en el que vive la población.

Educación: Años promedio de escolaridad para medir el acceso y la calidad de la educación.

Relaciones Sociales: La *calidad de la red social* de soporte refleja la importancia de las conexiones comunitarias y personales.

Compromiso Cívico: La participación electoral fue seleccionada para evaluar el compromiso cívico y la participación en procesos democráticos.

Balance Vida-Trabajo: La *proporción* de la población ocupada trabajando más de 48 horas a la semana indica el equilibrio entre la vida laboral y personal.

Satisfacción con la Vida: Directamente, la satisfacción con la vida mide la percepción general del bienestar.

Accesibilidad a Servicios: Se tomaron dos data sets, uno sobre accesibilidad a servicios de salud y otro sobre viviendas con acceso a servicios básicos, ilustran la disponibilidad de recursos esenciales.

Seguridad: Percepción de la inseguridad y tasa de homicidios se analizaron para comprender la influencia de la seguridad en el bienestar.

Salud: Esperanza de vida al nacer y salud auto reportada ofrecen una visión integral del estado de salud de la población.

Cada base de datos fue acompañada de un documento técnico proporcionado por el INEGI, que incluye los metadatos esenciales para comprender la metodología de recopilación, cálculo y significado de cada conjunto de datos. Este paso crítico aseguró una interpretación precisa de las variables y reforzó la integridad de la investigación. La cuidadosa revisión de estos documentos técnicos permitió no solo validar la calidad y relevancia de los datos sino también profundizar en el entendimiento de cada dimensión del bienestar examinada.

Este proceso meticuloso de selección y validación de bases de datos subraya el compromiso metodológico de esta integrando una base de datos sólida para explorar la complejidad del bienestar en México.

2. Preparación y limpieza de datos

Una vez consolidada la base de datos inicial a través de una recolección minuciosa, el siguiente paso fue la preparación y limpieza de los datos, etapas fundamentales para garantizar la fiabilidad y precisión del análisis. Este proceso se llevó a cabo con el propósito de abordar las variaciones en las unidades de medida, los periodos de tiempo y la presencia de datos faltantes, buscando así la homogeneidad y completitud del conjunto de datos.

2.1. Normalización de Periodos

Dada la diversidad de las bases de datos recolectadas, se hizo imperativo estandarizar el rango temporal de análisis para garantizar comparabilidad a lo largo del tiempo. Se decidió limitar el periodo de estudio desde el año 2000 hasta el 2022, considerando la suficiencia de los datos. Estos límites temporales se determinaron para maximizar la relevancia y la cobertura de los datos.

2.2. Manejo de datos faltantes

La integridad de los datos se enfrentó al desafío de los valores faltantes, que representaban el 53% del total. Este fenómeno se debió principalmente a la naturaleza esporádica de los censos realizados en el país, los cuales se llevan a cabo en intervalos variables dependiendo de las características específicas que se miden. Para abordar esta cuestión, se adoptaron estrategias específicas basadas en el contexto y la naturaleza de cada variable:

- **Regresión lineal para series temporales continuas:** En casos donde faltaban datos para más de un año consecutivo, o en el primer o último año del rango establecido, se aplicó un modelo de regresión lineal. Esta aproximación permitió estimar los valores faltantes con base en la tendencia observada a lo largo del tiempo, manteniendo la coherencia temporal y temática de los datos.
- **Promedio de años adyacentes para brechas puntuales:** Para datos faltantes ubicados entre dos años con datos disponibles, se calculó el promedio de los

valores de los años adyacentes. Este método asumió una variación lineal entre los años, proporcionando una estimación prudente para el año faltante.

3. Estandarización de Datos

La estandarización de los datos constituyó una fase crítica en la ingeniería de datos, crucial para la coherencia y comparabilidad del análisis. Este paso se diseñó para abordar las variaciones en las unidades de medida y escalas entre las distintas bases de datos, permitiendo una integración efectiva y análisis comparativo de las variables relacionadas con el bienestar.

La transformación de los datos a un formato de porcentaje relativo al promedio establecido por la Organización para la Cooperación y el Desarrollo Económicos (OCDE) para cada indicador fue el método seleccionado para la estandarización. Este enfoque permitió superar las diferencias inherentes en las escalas y unidades de medida de las variables, proporcionando un marco uniforme para la comparación. La regla de tres se aplicó meticulosamente para recalcular los valores de cada año como un porcentaje del promedio de la OCDE, donde el 100% representa este promedio. Este proceso aseguró que cada valor se contextualizara dentro de un marco comparativo global, en el ámbito internacional de los hallazgos y facilitando la interpretación de las variaciones del bienestar en relación con un estándar reconocido globalmente.

La decisión de estandarizar los datos refuerza la precisión analítica y la relevancia comparativa del estudio. Al convertir todos los indicadores a porcentajes relativos al promedio de la OCDE, se garantizó que las diferencias y tendencias observadas en el bienestar reflejaran variaciones reales en los indicadores, evitando que se atribuyeran a distorsiones por las escalas de medición. Esta estandarización facilitó un análisis más integrado y coherente de los datos, además proporcionó una base sólida para comparaciones internacionales, resaltando la posición de México en el contexto global.

La estandarización de los datos es, por lo tanto, un paso fundamental del enfoque metodológico adoptado en este estudio. A través de este proceso, el estudio asegura que el análisis subsiguiente del bienestar en México se base en una comparación justa y equitativa de los indicadores, permitiendo que las conclusiones y recomendaciones derivadas sean tanto válidas como aplicables a nivel nacional e internacional.

4. Integración de Datos

Una vez estandarizados los datos procedentes del INEGI, fue crucial unificarlos para facilitar un análisis integral del bienestar en México. La integración de estos datos permitió crear una visión holística del bienestar, habilitando análisis multidimensional basados en un amplio espectro de indicadores estandarizados como porcentajes relativos al promedio de la OCDE para cada indicador.

La estructura final del conjunto de datos consolidado es una matriz rectangular, organizada cronológicamente desde el año 2000 hasta el 2022, con un total de 23 filas

correspondientes a cada año y 16 columnas que representan los diferentes indicadores de bienestar. Estos incluyen Vivienda, Ingresos, Empleo, Medio Ambiente, Educación, Comunidad, Compromiso Cívico, Balance Vida-Trabajo, Satisfacción con la Vida, dos indicadores de Accesibilidad, dos de Seguridad, y dos de Salud, además de la columna final que representa el índice de Bienestar calculado.

El conjunto de datos cuenta con 16 indicadores más una columna de Bienestar, y todos los indicadores son numéricos continuos, expresados como porcentajes y 'Bienestar' que es un índice calculado. Dado que el objetivo es modelar el bienestar como una variable continua, no se aplican clases específicas. El total de registros es de 23, correspondiendo a los años de estudio desde 2000 hasta 2022.

La integración de estos datos busca que cada valor porcentual estuviera correctamente alineado con su año correspondiente y su variable específica, lo cual es esencial para la coherencia y exactitud del análisis. Este enfoque facilita comparaciones directas y significativas, tanto en el tiempo como con respecto a los estándares internacionales establecidos por la OCDE.

La estructuración del dataset refleja la profundidad del trabajo de ingeniería de datos detrás de la investigación y, al mismo tiempo, proporciona una base rica para el análisis del bienestar en México, estableciendo una herramienta para descubrir conocimiento relevante y formular recomendaciones efectivas para mejorar la calidad de vida.

5. Ecuación del bienestar

Esta sección detalla la ecuación específica del bienestar, diseñada para evaluar y analizar el impacto de múltiples factores en la calidad de vida en México. Primero se presenta la ecuación con los nombres generales de las variables y luego una versión más detallada con las abreviaturas de las métricas proporcionadas por el INEGI.

- **Ecuación general del bienestar:**

$$\begin{aligned} \text{Bienestar} = & \text{Vivienda} + \text{Ingresos} - \text{Empleo} - \text{Medio Ambiente} + \text{Educación} \\ & + \text{Comunidad} + \text{Compromiso Cívico} - \text{Balance Vida Trabajo} \\ & + \text{Satisfacción con la vida} + \text{Accesibilidad}_1 + \text{Accesibilidad}_2 \\ & - \text{Seguridad}_1 - \text{Seguridad}_2 + \text{Salud}_1 + \text{Salud}_2 \end{aligned}$$

- **Ecuación detallada:**

$$\begin{aligned} \text{Bienestar} = & H + IDHE - TD - NCA + EsC + PRSS + PE - x + SV + PP_{CASS} + PH_{CS} \\ & - PILC - TH + e + SA \end{aligned}$$

Donde las variables se definen como:

- H (Habitaciones por persona)
- IDHE (Ingreso disponible en los hogares)
- TD (Tasa de Desocupación)
- NCA (Nivel de contaminación del aire)

- EsC (Años de escolaridad)
- PRSS (Porcentaje de población adulta con red social de soporte)
- PE (Participación electoral)
- X (Población ocupada trabajando más de 48 horas)
- SV (Satisfacción con la vida)
- PPCASS (Porcentaje de personas con acceso a servicios de salud)
- PHCS (Porcentaje de hogares con acceso a servicios básicos)
- PILC (Porcentaje de la población que se siente insegura en su localidad)
- TH (Tasa de homicidios)
- e (Esperanza de vida al nacer)
- SA (Salud autoreportada)

La ecuación del bienestar presentada en esta tesis fue desarrollada a partir de una investigación previa que abarcó múltiples dimensiones del bienestar en México, extendiendo la visión tradicionalmente económica para incluir aspectos sociales, ambientales y de salud. Aunque algunos trabajos previos, como los estudios del OECD reflejados en el Better Life Index, han abordado estos factores de manera fragmentada o independiente, la propuesta de este trabajo es única al integrar estos factores en una única ecuación matemática que captura la interacción compleja entre ellos.

Esta ecuación refleja un análisis multidimensional del bienestar y representa una innovación metodológica al aplicar una combinación específica de factores, ponderados de acuerdo con su impacto percibido y medido en la calidad de vida. El modelo propuesto utiliza operaciones de suma y resta para equilibrar los efectos positivos y negativos de cada factor sobre el bienestar general, reflejando la realidad de que algunas áreas (como la educación y la salud) pueden tener un impacto directamente positivo, mientras que otras (como el desempleo y la contaminación) ejercen una influencia negativa.

Además, la elección de cada variable y su inclusión en la ecuación se basaron en datos y métricas específicas provistas por el INEGI, asegurando que el modelo sea relevante y aplicable al contexto mexicano. La ecuación del bienestar fue validada en un estudio previo, donde demostró una alta precisión (0.97 en una regresión lineal) en la predicción del bienestar, lo que sugiere que los factores seleccionados y su ponderación son adecuados para capturar la complejidad del bienestar en México.

Esta ecuación se distingue de otras aproximaciones por su capacidad para ofrecer un objetivo aspiracional para políticas públicas y planificación social, proporcionando un marco cuantitativo para evaluar y orientar mejoras en el bienestar nacional. Así, este modelo no solo es una contribución teórica al estudio del bienestar, sino que también ofrece una herramienta práctica y robusta para informar decisiones políticas y sociales.

6. Selección de modelos

En el proceso de selección del modelo más adecuado para predecir el bienestar, evalué varios modelos de aprendizaje automático. Los detalles específicos sobre las herramientas utilizadas, incluyendo lenguajes de programación y librerías, se encuentran en el Apéndice A de esta tesis.

Se optó por emplear una técnica de ensamble que combina la robustez de la Máquina de Vectores de Soporte (SVM) con la claridad interpretativa de los árboles de decisión, utilizando un enfoque de stacking. Esto implica entrenar inicialmente múltiples modelos base, en este caso, SVM y árboles de decisión, cada uno sobre el conjunto completo de datos. Luego, las predicciones de estos modelos base se utilizan como entradas para un modelo meta, cuya tarea es sintetizar estas predicciones en una única salida. El modelo meta, en este caso, es otra instancia de SVM, que fue entrenada no solo para corregir los posibles errores de los modelos base sino también para aprovechar sus diferentes perspectivas y fortalezas. Este método de ensamble busca maximizar la precisión y minimizar el riesgo de sobreajuste al combinar las capacidades predictivas complementarias de ambos clasificadores.

SVM con Decision Tree usando Stacking:

- Training Score: 0.987
- Test Score: 0.766
- Cross-validation Scores: -474.17, 0.773, -1.46, 0.370, -28.24
- Mean Cross-validation Score: -100.55

Este modelo mostró un alto rendimiento en el conjunto de entrenamiento pero su rendimiento en el conjunto de prueba y la validación cruzada fue inconsistentemente bajo, indicando un posible sobreajuste.

Random Forest:

- Training MSE: 674.62
- Test MSE: 2290.84
- Training R^2 : 0.991
- Test R^2 : 0.925
- Cross-validation Scores: 0.279, 0.885, 0.263, 0.725, -0.157
- Mean Cross-validation Score: 0.399

Aunque Random Forest tuvo un buen desempeño en los datos de entrenamiento, su rendimiento en la prueba y la variabilidad en la validación cruzada sugirieron una menor generalización en comparación con otros modelos.

SVM:

- Training MSE: 16708.29

- Test MSE: 11129.09
- Training R^2 : 0.781
- Test R^2 : 0.638

La SVM sin optimización mostró una capacidad predictiva moderada y una considerable diferencia entre los datos de entrenamiento y prueba, lo que sugiere limitaciones en su capacidad de generalización.

Parámetros de Optimización en SVM:

En el entrenamiento de la SVM mediante GridSearch en Python, se ajustaron varios parámetros para optimizar el rendimiento del modelo:

C: Este parámetro de regularización controla la compensación entre lograr un margen más amplio y minimizar la suma del cuadrado de la clasificación de errores. Un valor más alto de C significa que se le da más importancia a la clasificación correcta de cada punto de entrenamiento, esencial para evitar el subajuste y controlar el sobreajuste.

kernel: Define el tipo de función del kernel para transformar el espacio de entrada en una nueva dimensión donde los datos pueden ser separables linealmente. Los tipos linear y rbf (Radial Basis Function) son comunes, donde linear se usa para datos ya linealmente separables en el espacio original, y rbf es útil para capturar relaciones no lineales.

gamma: Un parámetro para el kernel rbf que define qué tan lejos la influencia de un solo ejemplo de entrenamiento llega, con valores bajos que significan 'lejos' y altos que significan 'cerca'. Los valores scale y auto ajustan automáticamente gamma basándose en las estadísticas de las características, facilitando la configuración del modelo sin necesidad de definirlo manualmente.

Estos parámetros son fundamentales para configurar el comportamiento del modelo SVM y fueron optimizados usando GridSearchCV, que realiza una búsqueda exhaustiva sobre los parámetros especificados para encontrar la mejor configuración mediante validación cruzada. La búsqueda ajusta los parámetros para maximizar la puntuación de validación cruzada basada en el coeficiente de determinación (r^2), lo que refleja cómo de bien el modelo puede generalizar a datos no vistos. El objetivo es minimizar el riesgo de sobreajuste al tiempo que se mantiene una alta precisión en la predicción de nuevos datos, como se demuestra con la evaluación de r^2 y el error cuadrático medio (mse) en el conjunto de prueba

SVM con Optimización GridSearch:

- Best Parameters: {'svr__C': 100, 'svr__gamma': 'scale', 'svr__kernel': 'linear'}
- Best Cross-validation Score: 0.99999977
- Test MSE: 0.0047
- Test R^2 : 0.99999985

La SVM optimizada con GridSearch sobresalió significativamente en todas las métricas, mostrando casi perfecta precisión en las pruebas y una capacidad de generalización excepcional, según se refleja en la puntuación de validación cruzada y el bajo error cuadrático medio en la prueba.

Decisión y Justificación

Basándome en el análisis comparativo, se seleccionó la SVM optimizada con GridSearch como el modelo más adecuado para este estudio. Las razones de esta elección incluyen su alta precisión y robustez, su capacidad para generalizar bien en datos no vistos, y la consistencia en sus predicciones como lo demuestra su excelente desempeño en la validación cruzada y las métricas de prueba. Esta selección asegura que el modelo es confiable y preciso para aplicaciones prácticas en la predicción del bienestar.

7. Conclusión

La trayectoria hacia una comprensión profunda del bienestar en México, a través de la lente de la ciencia de datos, no estuvo exenta de desafíos. Cada paso en el proceso de ingeniería de datos presentó obstáculos únicos, desde la recolección y limpieza de datos hasta su estandarización e integración.

Uno de los desafíos más significativos fue la gestión de datos faltantes dentro de las bases de datos. La variabilidad en la disponibilidad de datos a lo largo de los años y entre las diferentes variables requería una solución que preserve la integridad de los datos. La decisión de utilizar la regresión lineal para estimar valores faltantes en series temporales y el promedio de años adyacentes para brechas puntuales fue clave para mantener la continuidad de los datos sin comprometer su veracidad.

Otro desafío involucró la estandarización de datos provenientes de diversas fuentes con escalas y unidades de medida distintas. La decisión de convertir todos los indicadores a porcentajes relativos al promedio de la OCDE surgió como una solución para facilitar comparaciones significativas entre variables, alineando el análisis con estándares internacionales y aumentando su relevancia comparativa.

La meticulosa preparación y estandarización de los datos han tenido profundas implicaciones para el análisis del constructo. Este enfoque no solo ha garantizado la coherencia y comparabilidad de los datos a través de un amplio espectro de indicadores, sino que también ha enriquecido el potencial analítico del estudio.

La estandarización en porcentajes, en particular, ha facilitado una interpretación más intuitiva de los datos, permitiendo identificar rápidamente áreas donde México se desvía del promedio de la OCDE y donde se concentran oportunidades para mejorar el bienestar. Además, este enfoque ha mejorado la accesibilidad de los hallazgos, permitiendo que una audiencia más amplia comprenda los factores que contribuyen al bienestar del país sin la necesidad de navegar por la complejidad de diferentes escalas de medición.

La integración de datos en un único data set consolidado ha habilitado un análisis multifacético, abriendo la puerta a conocimiento más rico y recomendaciones fundamentadas. Esta base de datos unificada permite exploraciones detalladas de cómo diversos aspectos interactúan entre sí y cambian a lo largo del tiempo, proporcionando una base sólida para investigaciones futuras y para la formulación de políticas públicas basadas en evidencia.

Capítulo 4 – Modelado

1. Introducción

En el estudio del bienestar en México, es fundamental analizar cómo diversas variables y factores interactúan para influir en los niveles de bienestar de la población. Los escenarios permiten evaluar de forma sistemática el impacto individual y combinado de estas variables en un entorno controlado y predecible, facilitando una comprensión más profunda de sus efectos en diferentes contextos.

Al emplear una Máquina de Vectores de Soporte (SVM) optimizada mediante GridSearch, esta investigación configura escenarios específicos para cada una de las variables clave, como ingresos, educación, salud y seguridad, entre otros, con el fin de identificar su influencia directa e indirecta en el bienestar. Cada escenario representa una combinación única de condiciones que proporciona un marco detallado para comprender los posibles resultados y variaciones en los niveles de bienestar bajo diversas circunstancias.

Esta sección se centrará en la descripción de las siete variables clave: Educación, Seguridad, Medio Ambiente, Ingresos, Comunidad, Compromiso Cívico y Balance Vida-Trabajo, y cómo se integran dentro de la SVM para analizar su impacto predictivo en el bienestar. Para las variables de Educación y Seguridad, se considera un lapso de 20 años, mientras que para Medio Ambiente, Ingresos, Comunidad, Compromiso Cívico y Balance Vida-Trabajo, se toma un período de 10 años. Al delinear estos escenarios, se busca desarrollar una visión clara y estructurada de las dinámicas que subyacen al bienestar en México, ofreciendo conocimiento valioso para la formulación de políticas públicas y la mejora general de la calidad de vida. Debido a la precisión del modelo predictivo utilizado, los valores presentados en las gráficas muestran una variación máxima del 10%, reflejando la robustez y la confiabilidad de las predicciones generadas.

2. Escenarios

Medio Ambiente: Escenario de Contaminación del Aire

La calidad del aire es un factor crítico que afecta el bienestar de las personas en zonas urbanas. Un ejemplo revelador de cómo las políticas ambientales pueden mejorar la

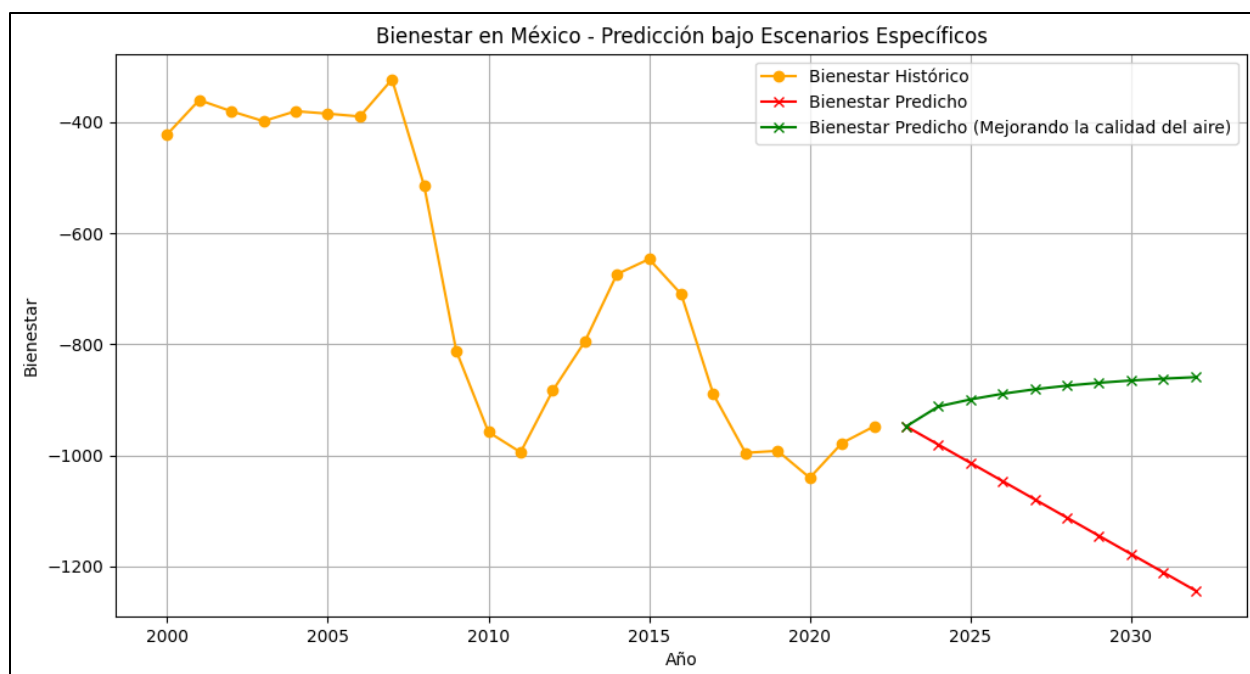
calidad del aire es el estudio de caso de Beijing, descrito en el artículo de Jiming Hao y Litao Wang (2012), "Improving Urban Air Quality in China: Beijing Case Study" muestra que la ciudad logró reducir la concentración de dióxido de azufre, dióxido de nitrógeno, fosfato trisódico y dióxido de carbono en el aire un 20% durante un período de rápido crecimiento económico y urbanización, demostrando que es posible mejorar la calidad del aire incluso en entornos urbanos desafiantes.

En el contexto de esta investigación, este escenario de medio ambiente se configura para evaluar cómo una reducción del 20% en la contaminación del aire podría impactar en el bienestar general.

La gráfica siguiente ilustra el impacto potencial de esta reducción en la calidad del aire sobre las diferentes dimensiones del bienestar en México, proporcionando conocimientos valiosos para la formulación de políticas públicas que apunten a mejorar la calidad de vida en contextos urbanos.

Figura 2

Impacto de la Reducción de Contaminantes en el Bienestar



La gráfica muestra una mejora en el índice de bienestar como resultado de políticas públicas enfocadas en la reducción de las concentraciones de dióxido de nitrógeno,

fosfato trisódico y dióxido de carbono en el aire. A través de estas intervenciones, el bienestar, que originalmente mostraba una tendencia descendente hacia -1200, mejora acercándose a un valor de -800. Esta elevación refleja cómo medidas efectivas de control de la contaminación pueden tener un impacto directo y positivo en la calidad de vida de los ciudadanos. Al disminuir los niveles de estos contaminantes, no solo se beneficia la salud pública, sino que también se contribuye a un entorno más saludable y sostenible, lo que a su vez fomenta un mayor bienestar general en la población (preventiva .vs. correctiva).

Balance Vida-Trabajo: Escenario de Reducción de Jornada Laboral

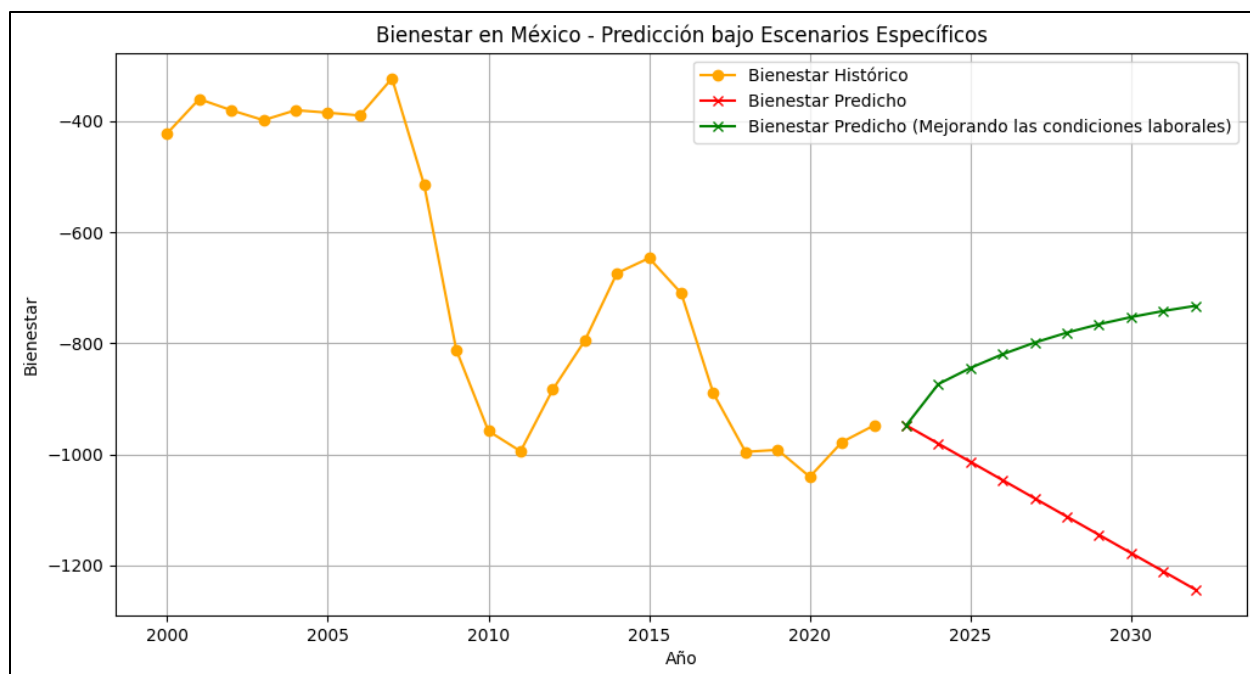
El equilibrio entre la vida laboral y personal es un componente clave del bienestar, que impacta en la salud mental, física y la satisfacción general de los trabajadores. El artículo "Does work time reduction improve workers' well-being? Evidence from global four-day workweek trial"(Wen Fan et al. ,2023) proporciona evidencia reveladora de que una reducción de la jornada laboral a una semana de cuatro días, sin disminución del salario, puede mejorar significativamente el bienestar de los empleados. El estudio encontró que esta intervención generó un aumento del 15% en múltiples medidas de bienestar, como el menor agotamiento, la satisfacción laboral, y la mejora de la salud mental y física.

En este escenario, se explora cómo una reducción similar de las horas laborales podría influir positivamente en el bienestar de los trabajadores en México. Usando una Máquina de Vectores de Soporte, se modelará el impacto potencial de una reducción del 15% en el tiempo de trabajo sobre el bienestar general, específicamente en aspectos como el equilibrio vida-trabajo, la satisfacción laboral y la salud mental.

La gráfica siguiente muestra cómo esta reducción puede afectar los diversos aspectos del bienestar en México, ofreciendo conocimiento valioso para la formulación de políticas que promuevan un mejor balance entre la vida laboral y personal.

Figura 3

Efecto de la reducción de la jornada laboral en el bienestar



La gráfica ilustra el efecto positivo que tiene la reducción de la jornada laboral en el bienestar general. Implementando una semana laboral más corta, sin disminuir el salario, el bienestar mejora sustancialmente, alcanzando un valor cercano a -700. Esta mejora refleja cómo la disminución de horas trabajadas contribuye a un mejor equilibrio vida-trabajo, reduciendo el estrés y aumentando la satisfacción general y la calidad de vida de los trabajadores. La política de reducción de jornada laboral demuestra ser una estrategia efectiva para promover un entorno laboral más saludable y, por extensión, mejorar el bienestar social.

Compromiso Cívico: Escenario de Participación Electoral

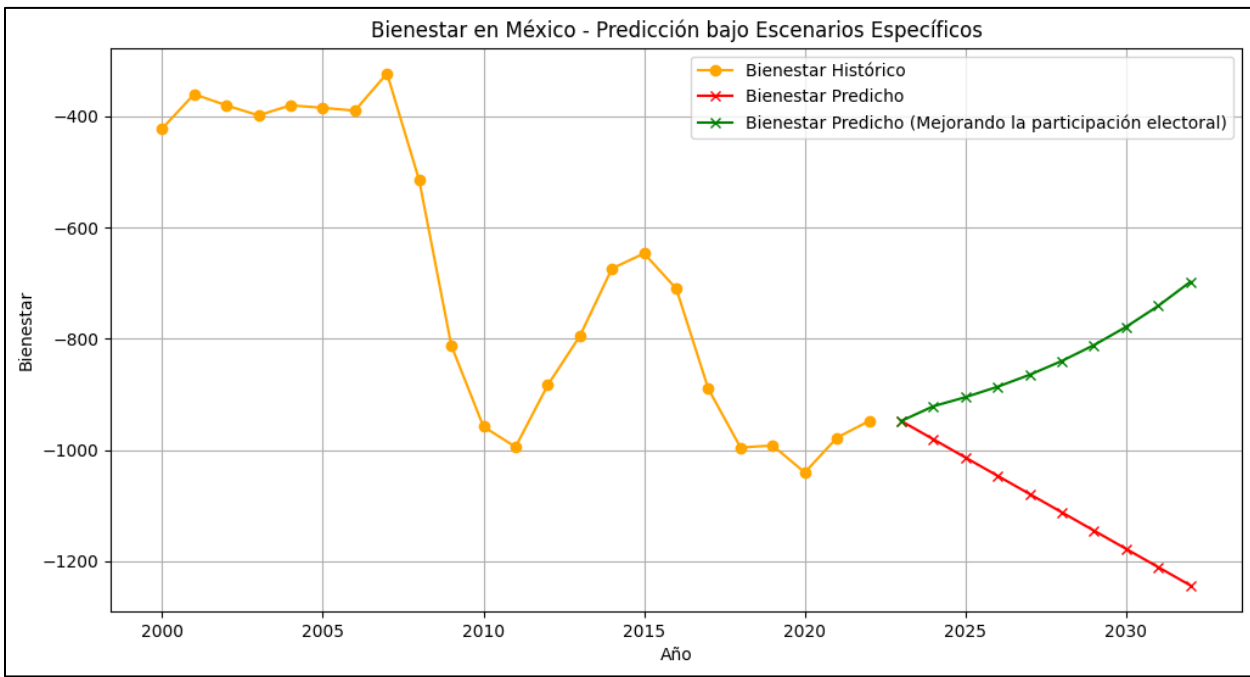
La participación cívica, particularmente a través del voto, es un indicador esencial del bienestar social, reflejando la participación activa de los ciudadanos en las decisiones que afectan a sus comunidades. El artículo "Getting Out the Youth Vote: Results from Randomized Field Experiments" (Donald P. Green, 2014) presenta una iniciativa que logró un aumento del 15% en la participación electoral entre los jóvenes, gracias a campañas telefónicas y contactos directos que motivaron a este grupo demográfico a ejercer su derecho al voto.

En este escenario, se examina cómo un incremento del 15% en la participación cívica, especialmente entre los jóvenes, podría impactar el bienestar en México.

La gráfica siguiente muestra cómo este incremento en el compromiso cívico puede afectar positivamente las diferentes dimensiones del bienestar en México, proporcionando información valiosa para diseñar estrategias que fomenten una mayor participación electoral y, en consecuencia, un mayor sentido de pertenencia y satisfacción comunitaria.

Figura 4

Impacto del Incremento en la participación electoral de los jóvenes en el bienestar



La gráfica muestra cómo un incremento del 15% en la participación electoral de los jóvenes puede elevar el bienestar general, llevándolo cerca de un valor de -600. Este aumento refleja la importancia del compromiso cívico en el fortalecimiento del tejido social y en la promoción de una democracia activa y participativa. Al fomentar una mayor participación electoral entre los jóvenes, no solo se mejora su sentido de agencia y pertenencia, sino que también se influye positivamente en la cohesión social y en el sentido de responsabilidad compartida hacia el bienestar comunitario. Este tipo de

intervenciones demuestra que el empoderamiento de los jóvenes a través del voto puede tener un impacto transformador en la sociedad.

Comunidad: Escenario de Redes Vecinales

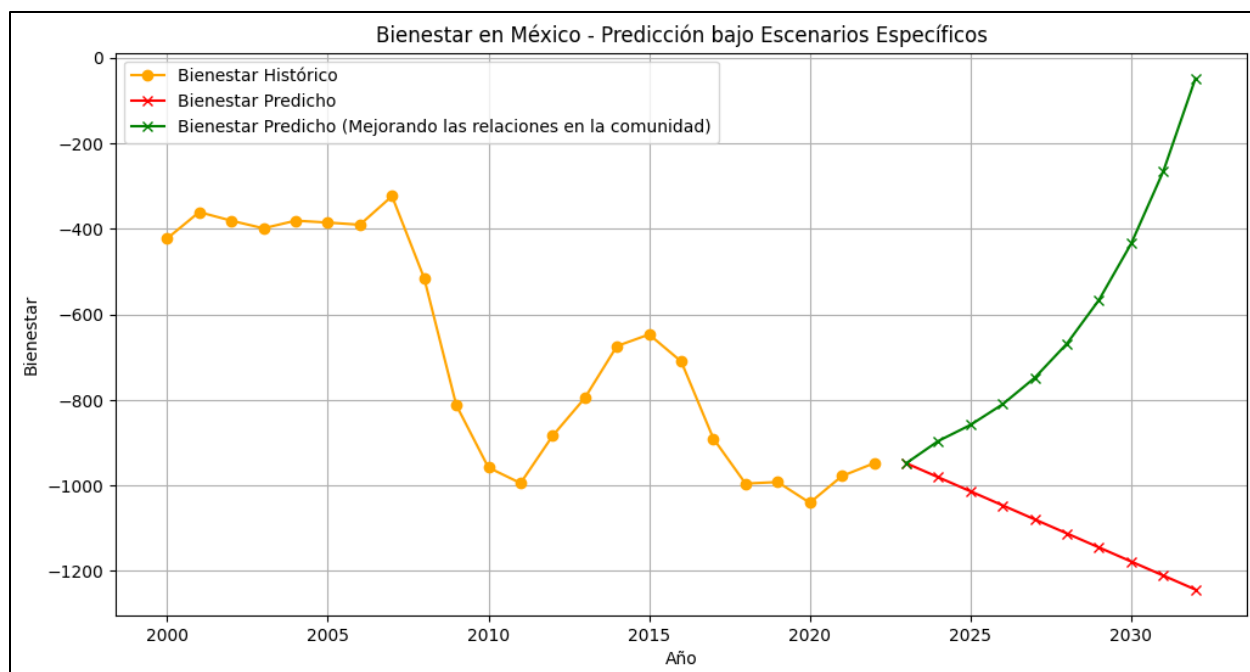
Las redes vecinales sólidas son esenciales para crear una comunidad cohesionada y un mayor bienestar social. El informe "Chicago Neighborhood Networks: Stability and Change" (Quiroz Becerra et al., 2019) presenta cómo la cooperación y coordinación entre las organizaciones locales pueden impulsar la mejora de las comunidades al aprovechar las inversiones y acciones del sector público más allá de lo que una organización podría lograr individualmente. Según el informe, los esfuerzos conjuntos y la flexibilidad de las redes vecinales resultaron en un aumento del 20% en la mejora de las condiciones comunitarias en Chicago.

En este escenario, se modela cómo un aumento del 20% en la mejora de la comunidad, gracias a la colaboración de organizaciones locales, podría impactar positivamente en el bienestar en México.

La gráfica siguiente ilustra cómo un enfoque colaborativo y coordinado entre las organizaciones locales puede mejorar significativamente el bienestar a través de una red comunitaria más sólida, proporcionando información valiosa para fortalecer las políticas de cooperación local en México.

Figura 5

Efecto de la mejora de las relaciones en la comunidad sobre el bienestar



La gráfica ilustra un notable aumento en el bienestar, que se aproxima a casi 0, como resultado de un aumento del 20% en la mejora de la comunidad a través de la colaboración entre organizaciones locales. Este significativo ascenso en el bienestar destaca el poder de las redes comunitarias y las sinergias organizacionales en la creación de entornos más saludables y solidarios. Al fomentar la cooperación entre diversas entidades, se potencia el impacto de las intervenciones comunitarias, mejorando significativamente la calidad de vida de los residentes. Este escenario demuestra cómo una gestión comunitaria efectiva y coordinada puede transformar positivamente el bienestar social, llevándolo a niveles cercanos a la neutralidad, lo cual indica una situación de equilibrio y satisfacción general en la población.

Ingresos: Escenario de Reducción de la Pobreza

El aumento de los ingresos y la reducción de la pobreza son factores fundamentales para mejorar el bienestar de la población. El Programa Bolsa Familia, como se detalla en el informe de Aline Gazola Hellmann (2015), "¿Cómo funciona Bolsa Familia? Mejores prácticas en la implementación de programas de transferencias monetarias condicionadas en América Latina y el Caribe", ha demostrado ser efectivo para disminuir

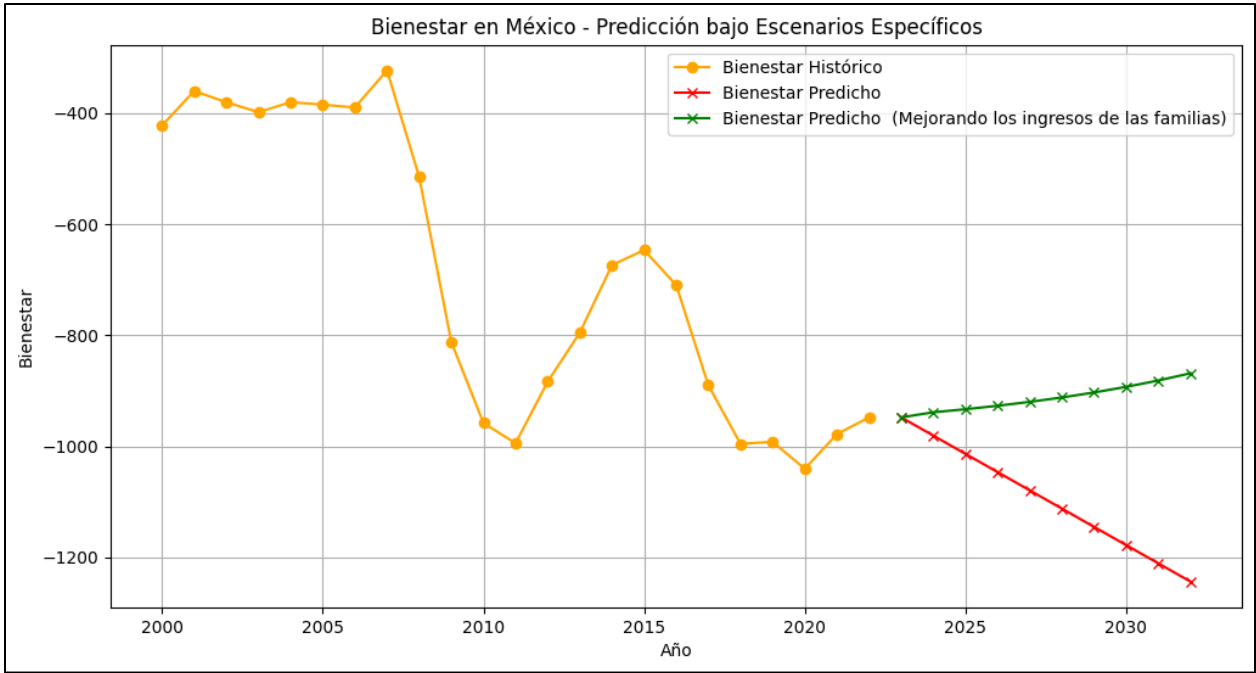
la pobreza en Brasil. El programa, implementado en 2003, logró reducir la pobreza en un 13% al proporcionar transferencias monetarias condicionadas a familias con ingresos bajos, generando mejoras significativas en la salud y la educación.

En el escenario de ingresos, esta investigación modela cómo un programa similar de transferencias monetarias podría influir positivamente en la reducción de la pobreza y, en consecuencia, mejorar el bienestar en México. Se simula el impacto de una reducción del 13% en la pobreza sobre los ingresos, la educación y otros aspectos del bienestar.

La gráfica siguiente muestra cómo este programa puede mejorar de forma significativa el bienestar al reducir la pobreza, proporcionando una base para diseñar estrategias que apoyen a las familias con ingresos bajos y, en última instancia, mejorar su calidad de vida.

Figura 6

Impacto de un Programa de Transferencias Monetarias en la Reducción de la Pobreza en el Bienestar



La gráfica demuestra que la implementación de un programa de transferencias monetarias puede resultar en una reducción significativa de la pobreza, con una disminución estimada del 13%. Este cambio produce una mejora notable en el índice de

bienestar, llevándolo a cerca de -800. Este aumento refleja cómo las intervenciones económicas directas no solo alivian la carga financiera inmediata de las familias más vulnerables, sino que también tienen un impacto positivo sustancial en su bienestar general. Al proporcionar apoyo financiero a quienes más lo necesitan, se proporciona certeza económica, lo que contribuye a una mejora en varios aspectos de la vida, incluyendo salud, educación y calidad de vida general.

Seguridad: Escenario de Reducción de Homicidios

La seguridad ciudadana es un pilar fundamental para el bienestar, ya que influye directamente en la calidad de vida y el sentido de pertenencia de las comunidades. En el informe "Citizen Security in Latin America: The Hard Facts," Robert Muggah y Katherine Aguirre Tobón (2018) señalan que la tasa de homicidios en Colombia se ha reducido en un 82% en los últimos 25 años. Este logro fue posible gracias a un enfoque que trata la violencia como un problema de salud pública, combinando políticas que abordan las causas fundamentales de la violencia con datos e intervenciones basadas en evidencia.

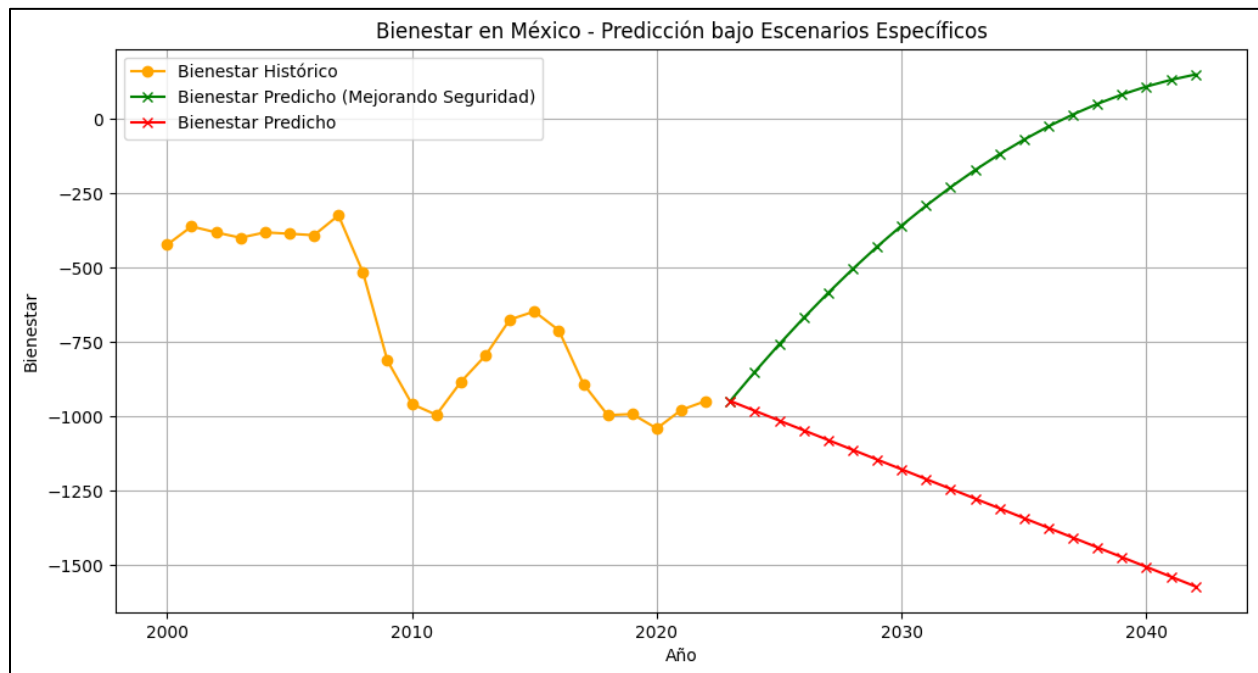
Según el artículo "Treating violence like a disease helped cut Colombia's murder rate by 82%" de Apolitical, la caída de las tasas de homicidios en Colombia también se debió a la caída de cárteles de la droga, la desmovilización de grupos armados y el uso de métodos inspirados en la salud pública, que abordaron factores como la desigualdad económica, la urbanización rápida, el abuso de drogas y alcohol, y la desigualdad de género.

Este escenario se configura para modelar cómo una reducción del 82% en el ratio de homicidios, a través de una estrategia basada en la salud pública, podría impactar en el bienestar en México.

La gráfica siguiente ilustra el impacto positivo que una estrategia integral de seguridad podría tener en la sociedad mexicana, proporcionando un marco valioso para mejorar las políticas públicas y la calidad de vida.

Figura 7

Efectos de la Reducción del Ratio de Homicidios en el Bienestar



La gráfica muestra un resultado impresionante donde, tras una reducción del 82% en el ratio de homicidios, el bienestar mejora drásticamente, alcanzando un valor cercano a 200 en un período de 20 años. Este aumento en el bienestar indica que intervenciones efectivas en la seguridad pública, que logran disminuir significativamente la violencia, tienen un profundo impacto positivo en la percepción general de seguridad y en la calidad de vida de la población. Al abordar uno de los problemas más críticos y directos que afectan a la sociedad, la reducción de homicidios no solo salva vidas sino que también fomenta un entorno más seguro y tranquilo, lo que permite a los ciudadanos y comunidades prosperar.

Educación: Escenario de Mejora en Años de Escolaridad

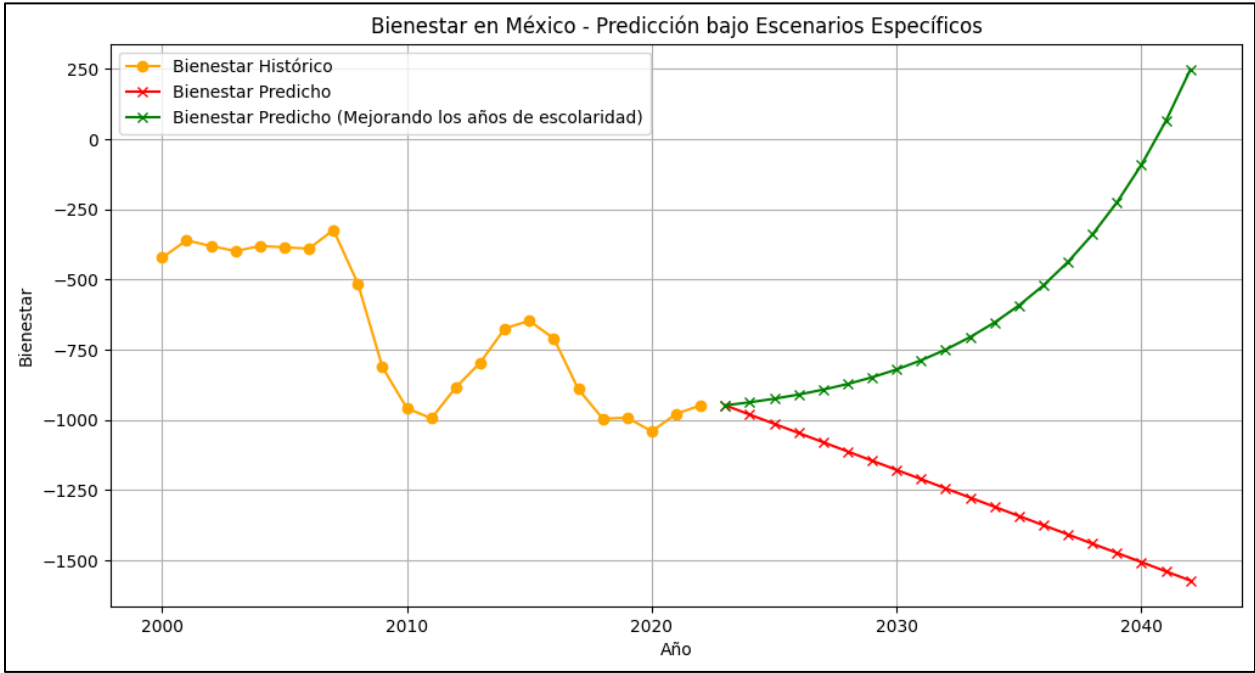
La educación es un pilar esencial del bienestar, ya que proporciona oportunidades para el desarrollo personal y el crecimiento económico. En su estudio "La Transformación del Sistema Educativo de Corea del Sur entre 1930 y 1980, Calidad y Equidad," Leonardo Pérez González (2016) analiza el impresionante avance en el sistema educativo surcoreano durante este período crítico. Los años de escolaridad promedio en Corea del Sur aumentaron de 4 a 12, lo que representa un crecimiento del 67%. Este aumento se logró mediante una serie de reformas intencionales y encadenadas que superaron obstáculos específicos, logrando así un sistema educativo que prioriza la calidad y equidad.

En este escenario, se explora cómo un aumento del 67% en los años de escolaridad podría afectar el bienestar en México.

La gráfica siguiente muestra el impacto positivo que una mejora en la calidad y equidad del sistema educativo podría tener en la sociedad mexicana, proporcionando información valiosa para las estrategias que buscan mejorar la educación y, en consecuencia, el bienestar en México.

Figura 8

Impacto de un aumento en los años de escolaridad sobre el bienestar



La gráfica ilustra el impacto positivo de un aumento del 67% en los años de escolaridad durante un período de 20 años, lo que lleva el bienestar a superar ligeramente el valor de 250. Este notable avance refleja la influencia crucial de la educación en la mejora del bienestar general. Al extender la duración de la educación formal, se enriquecen las oportunidades de vida de los individuos, mejorando sus perspectivas laborales, ingresos potenciales, y capacidad para tomar decisiones informadas, lo que contribuye a una mejor calidad de vida. Este incremento en los años de escolaridad, basado en las reformas educativas implementadas en Corea del Sur, demuestra cómo las inversiones estratégicas en el sistema educativo no solo elevan el nivel educativo de la población sino que también tienen un efecto multiplicador en el bienestar social y económico.

3. Conclusiones

En este capítulo, se presentaron y analizaron diversos escenarios que modelan el impacto de factores críticos en el bienestar de México. Cada escenario proporcionó conocimiento importante sobre cómo las políticas y cambios estratégicos en áreas clave pueden mejorar significativamente la calidad de vida de la población. Los hallazgos principales incluyen:

Medio Ambiente: La reducción de la contaminación del aire en un 20%, basada en el caso de Beijing, demostró un impacto positivo en la salud y otros aspectos del bienestar. Esto subraya la necesidad de políticas ambientales efectivas para mejorar la calidad de vida en las áreas urbanas.

Balance Vida-Trabajo: Una reducción del 15% en las horas laborales, como se observó en el ensayo global de una semana laboral de cuatro días, mostró mejoras significativas en la satisfacción laboral y la salud mental de los trabajadores, evidenciando la necesidad de equilibrar la vida personal con la laboral.

Compromiso Cívico: Un aumento del 15% en la participación electoral entre los jóvenes, impulsado por campañas de movilización, resaltó el impacto positivo del compromiso cívico en el sentido de pertenencia y en la cohesión social.

Comunidad: El fortalecimiento de redes comunitarias resultó en una mejora del 20% en la cohesión social y la calidad de vida. La colaboración entre organizaciones locales es crucial para fomentar el desarrollo sostenible en las comunidades.

Ingresos: El programa Bolsa Familia, que redujo la pobreza en un 13% en Brasil, sirvió como un modelo exitoso para mejorar los ingresos de las familias de bajos recursos, resaltando el valor de las transferencias monetarias condicionadas.

Seguridad: La reducción del 82% en la tasa de homicidios en Colombia, mediante un enfoque basado en la salud pública, resaltó el impacto positivo que la mejora en la seguridad puede tener sobre la percepción de bienestar y la confianza en las instituciones.

Educación: Un incremento del 67% en los años de escolaridad, como se vio en Corea del Sur, demostró la importancia de las reformas educativas intencionales para mejorar el acceso y la calidad, lo cual se reflejó en la movilidad económica y la satisfacción con la vida.

En conjunto, estos escenarios refuerzan la relevancia de un enfoque analítico avanzado para entender el bienestar en México. El uso de una Máquina de Vectores de Soporte (SVM) optimizada mediante GridSearch permitió capturar la complejidad de estas dinámicas y ofrecer una visión sólida de cómo las políticas estratégicas pueden mejorar el bienestar en el país. Este capítulo establece una base fundamental para futuras investigaciones y estrategias que aspiren a crear un impacto positivo en la sociedad mexicana.

Capítulo 5 – Conclusiones y recomendaciones

1. Resumen de Hallazgos Principales

La presente investigación ha abordado la tarea de modelar y predecir el bienestar en México utilizando técnicas de machine learning, en particular, una Máquina de Vectores de Soporte (SVM) optimizada mediante GridSearch. A través de este estudio, se ha logrado no solo una comprensión más profunda de los factores que influyen en el bienestar, sino también un modelo predictivo capaz de ofrecer buena precisión en sus predicciones.

Principales Hallazgos:

- **Modelo Predictivo de Alta Precisión:** La implementación de la SVM optimizada con GridSearch ha demostrado ser efectiva, logrando una puntuación de validación cruzada buena. Este resultado subraya la capacidad del modelo para generalizar bien en datos no vistos, proporcionando un instrumento robusto y confiable para las políticas públicas.
- **Influencia de Variables Socioeconómicas y Ambientales:** Se mostró que variables como el ingreso de los hogares, la educación, la seguridad, y la contaminación del aire tienen impactos significativos en el bienestar. La ecuación del bienestar desarrollada, que incluye estas y otras variables, proporciona un marco analítico para entender cómo cada factor contribuye al bienestar general.
- **Impacto de las Políticas Públicas:** El análisis sugiere que intervenciones estratégicas en áreas como medio ambiente, educación, y empleo podrían mejorar significativamente el bienestar. Estos hallazgos ofrecen una guía valiosa para los diseñadores de políticas públicas y destacan la importancia de enfoques basados en evidencia para la formulación de estrategias de bienestar.
- **Escenarios de Intervención:** El estudio exploró varios escenarios de intervención, mostrando cómo cambios en específicas variables pueden alterar el índice de bienestar. Por ejemplo, mejoras en la calidad del aire y reducciones en la jornada laboral mostraron potenciales incrementos en el bienestar, subrayando la interconexión entre el ambiente laboral, la salud pública y el bienestar.

Este compendio de hallazgos enriquece el conocimiento existente sobre el bienestar en México y también proporciona herramientas prácticas y recomendaciones basadas en datos que pueden ser aplicadas para mejorar la calidad de vida en el país. La metodología y los resultados de este estudio demuestran la eficacia de combinar técnicas de ciencia de datos con un enfoque sistemático y bien fundamentado para abordar cuestiones sociales.

2. Implicaciones de la Investigación

La investigación presente demuestra la aplicación efectiva de técnicas de ciencia de datos en el estudio y mejora del bienestar social. Los resultados destacan varios aspectos importantes:

Modelado Predictivo Avanzado: Utilizando SVM optimizada con GridSearch, este estudio ilustra cómo los métodos de aprendizaje automático pueden ser aplicados para resolver problemas sociales. La precisión y robustez del modelo SVM subrayan la capacidad de estas técnicas para manejar y analizar grandes volúmenes de datos multidimensionales, ofreciendo conocimiento que puede ser crucial para la toma de decisiones basada en datos.

Integración de Diversas Fuentes de Datos: Al combinar múltiples datasets que abarcan desde ingresos y educación hasta seguridad y salud, la investigación resalta la importancia de un enfoque integrador en ciencia de datos. Este enfoque no solo mejora la comprensión de los fenómenos estudiados, sino que también refuerza la calidad de las predicciones realizadas por modelos analíticos.

Desarrollo de Políticas Basadas en Evidencia: Los científicos de datos pueden usar los métodos demostrados en este estudio para desarrollar modelos que informen y guíen la formulación de políticas públicas. Al demostrar cómo cambios específicos en variables pueden influir en el bienestar, los modelos predictivos como el desarrollado aquí sirven como herramientas para los responsables de la formulación de políticas, quienes pueden simular los efectos de diversas políticas antes de su implementación.

Transparencia y Replicabilidad: Este estudio también aporta a la comunidad académica un ejemplo de cómo mantener la transparencia y replicabilidad en

investigaciones. La clara documentación del proceso de selección del modelo, junto con la detallada exposición de los parámetros y la metodología, proporciona un marco que otros investigadores pueden seguir y adaptar a sus propios contextos de estudio.

3. Limitaciones del Estudio

A pesar de los avances que este estudio aporta al campo del bienestar en México y a la aplicación de técnicas de machine learning en la política pública, enfrenta varias limitaciones que deben ser consideradas al interpretar los resultados:

- **Cobertura de Datos:** Mientras que la investigación incorporó una amplia gama de variables socioeconómicas y ambientales, la disponibilidad y la actualización de los datos pueden limitar la capacidad de generalizar los resultados. En particular, las bases de datos del INEGI, aunque exhaustivas, pueden no capturar todas las dinámicas o cambios recientes en las condiciones socioeconómicas.
- **Modelo Predictivo:** Aunque la SVM optimizada con GridSearch mostró un rendimiento excepcional, como todos los modelos de aprendizaje automático, depende de la calidad y la representatividad de los datos con los que fue entrenado. Además, la interpretación de los resultados de un modelo como este puede ser menos intuitiva que con modelos más simples, lo que podría limitar su utilidad en ciertos contextos aplicados.
- **Variabilidad Regional y Cultural:** México es un país con una gran diversidad regional y cultural, lo que puede influir en cómo diferentes factores afectan el bienestar. Este estudio no pudo desglosar completamente cómo las intervenciones podrían necesitar ser ajustadas para ser efectivas en contextos locales específicos.
- **Dinámicas Temporales:** Los efectos a largo plazo de las políticas evaluadas pueden variar, y el modelo utilizado puede no ser completamente adecuado para predecir cambios a muy largo plazo debido a factores externos no incluidos en el modelo actual.

4. Recomendaciones para Futuras Investigaciones

Los resultados de esta tesis abren varias vías para investigaciones futuras que podrían ampliar y profundizar el entendimiento del bienestar en México y la eficacia de las

intervenciones políticas. A continuación, se presentan algunas recomendaciones específicas:

- **Ampliación de la Base de Datos:** Futuros estudios podrían beneficiarse de la incorporación de más datos, lo que permitiría un análisis más detallado de las tendencias a lo largo del tiempo y la evaluación de la persistencia de los efectos de las políticas sobre el bienestar.
- **Análisis Regional Específico:** Dada la diversidad regional de México, sería valioso desarrollar modelos que se enfoquen en contextos locales específicos. Esto ayudaría a diseñar políticas más personalizadas que consideren las particularidades culturales, económicas y sociales de cada región.
- **Exploración de Nuevas Variables:** Investigar otras variables que puedan influir en el bienestar pero que no fueron incluidas en este estudio podría proporcionar una visión más completa. Por ejemplo, variables relacionadas con la seguridad laboral, el acceso a tecnologías de la información, o el impacto de las redes sociales podrían ser relevantes.
- **Desarrollo de Modelos Alternativos:** Probar otros modelos predictivos o técnicas de aprendizaje automático podría enriquecer o confirmar los hallazgos actuales. Comparar diferentes métodos también puede ofrecer perspectivas sobre las mejores prácticas en modelado predictivo para estudios de bienestar.
- **Estudios de Intervención:** Realizar estudios empíricos donde se implementen las políticas sugeridas y se mida su impacto real podría validar aún más la efectividad de las recomendaciones basadas en el modelo y ajustarlas según los resultados obtenidos.

5. Contribuciones al Conocimiento

La investigación presentada en esta tesis no solo enriquece la comprensión del bienestar en México, sino que también ofrece contribuciones a las ciencias computacionales,

especialmente en la aplicación de técnicas de aprendizaje automático para abordar problemas sociales.

- **Modelo del Bienestar:** La tesis propone una ecuación del bienestar que integra múltiples indicadores socioeconómicos y ambientales, proporcionando una herramienta analítica para medir y evaluar el bienestar objetivo de manera comprensiva. Esta ecuación permite un análisis de cómo diversas variables interactúan para influir en el bienestar, ofreciendo una base sólida para la formulación de políticas enfocadas y efectivas.
- **Avances en Modelado Predictivo:** El uso efectivo de SVM optimizada con GridSearch destaca la capacidad del aprendizaje automático para aplicarse en la predicción precisa del bienestar, mostrando cómo estas técnicas pueden ser utilizadas en contextos sociales para obtener resultados significativos y aplicables.
- **Innovación en Evaluación de Políticas Públicas:** La capacidad del modelo para simular y predecir los efectos de políticas sobre el bienestar abre nuevos caminos para evaluaciones basadas en evidencia, fortaleciendo el vínculo entre análisis computacional y decisiones políticas.

6. Reflexiones Finales

Al concluir esta investigación, reflexiono sobre el proceso, los desafíos encontrados y las lecciones aprendidas durante el desarrollo de esta tesis, las cuales han sido tan enriquecedoras como los resultados mismos.

Apreciaciones Personales y Profesionales:

Este proyecto no solo ha profundizado mi comprensión del bienestar y su compleja naturaleza en México, sino que también ha reforzado mi habilidad para aplicar técnicas de ciencia de datos a problemas reales. La experiencia de diseñar y ejecutar un modelo predictivo y de tratar con diferentes tipos de datos ha sido invaluable, y sin duda influirá en mi futura carrera, sea en el ámbito académico o profesional.

Desafíos Superados:

Uno de los mayores desafíos fue la integración de diversas fuentes de datos y la aplicación efectiva del machine learning. Otro obstáculo significativo fue la dificultad para

encontrar literatura relevante que respaldara directamente mi investigación. La escasez de estudios previos que combinen ciencias computacionales con análisis del bienestar en contextos similares al mexicano implicó un esfuerzo considerable en la identificación y adaptación de teorías y métodos aplicables desde otros contextos geográficos. Superar estos obstáculos técnicos y bibliográficos no solo mejoró mi competencia técnica, sino que también me enseñó la importancia de la perseverancia y la adaptabilidad en la investigación. Esta experiencia resalta cómo los desafíos pueden ser oportunidades para innovar y contribuir significativamente a un campo de estudio.

Contribuciones a la Comunidad Académica:

Espero que este trabajo no solo contribuya al campo académico, proporcionando un modelo replicable y expandible, sino que también inspire a otros investigadores a explorar cómo la ciencia de datos puede ser aplicada para mejorar la formulación de políticas públicas y, en última instancia, la vida de las personas.

Visión para el Futuro:

Mirando hacia el futuro, estoy motivado para seguir trabajando en proyectos que unan la ciencia de datos con las ciencias sociales para abordar cuestiones de significativa relevancia social. Además, estoy interesado en explorar más sobre la ética en la ciencia de datos, especialmente en cómo los modelos predictivos afectan a los individuos y a la sociedad en su conjunto.

Apéndice A: Detalles de Implementación

Este apéndice proporciona una visión detallada de la implementación técnica del análisis realizado en este estudio, incluyendo las herramientas y bibliotecas de software utilizadas.

Entorno de Desarrollo:

- Lenguaje de Programación: Python
- Bibliotecas Utilizadas:
 - pandas: Utilizada para la manipulación y limpieza de datos.
 - numpy: Empleada para operaciones numéricas avanzadas.
 - seaborn y matplotlib.pyplot: Usadas para la visualización de datos.
 - sklearn: Biblioteca de aprendizaje automático utilizada para la preparación de datos, construcción y validación de modelos.
 - Polynomial de numpy.polynomial.polynomial: Utilizada para ajustes polinómicos en el pre-procesamiento de datos.

Proceso de Análisis de Datos:

1. Pre-procesamiento:

- Carga y limpieza de datos utilizando pandas.
- Transformaciones y normalización de características con StandardScaler para asegurar que los datos están en una escala común.
- Creación de variables polinómicas para modelar relaciones no lineales.

2. División del Dataset:

- Uso de train_test_split de sklearn.model_selection para dividir los datos en conjuntos de entrenamiento y prueba, asegurando una evaluación adecuada del modelo.

3. Modelado y Entrenamiento:

- Configuración y entrenamiento de una Máquina de Vectores de Soporte (SVM) utilizando SVR de `sklearn.svm`.
- Optimización de parámetros mediante `GridSearchCV` para encontrar la configuración más efectiva del modelo SVM.

4. Evaluación del Modelo:

- Cálculo del error cuadrático medio (MSE) y el coeficiente de determinación (R^2) usando `mean_squared_error` y `r2_score` de `sklearn.metrics` para evaluar el rendimiento del modelo.

5. Visualización de Resultados:

- Generación de gráficos de rendimiento y análisis utilizando `seaborn` y `matplotlib.pyplot` para visualizar las relaciones y los resultados del modelo.

Bibliografía

Apolitical. (s.f.). Treating violence like a disease helped cut Colombia's murder rate by 82%. Recuperado de: <https://apolitical.co/solution-articles/en/treating-violence-like-a-disease-helped-cut-colombias-murder-rate-by-82>

Fan, W., Tang, Z., Zheng, C., & Cui, K. (2023). Does work time reduction improve workers' well-being? Evidence from global four-day workweek trial. *Social Science Research Network*.

Green, D. P., & Gerber, A. S. (2015). *Get out the vote: How to increase voter turnout*. Brookings Institution Press.

Hao, J., & Wang, L. (2012). Improving urban air quality in China: Beijing case study. *Journal of Environmental Sciences*, 24(1), 2-12.

Hellmann, A. G. (2015). ¿Cómo funciona Bolsa Familia? Mejores prácticas en la implementación de programas de transferencias monetarias condicionadas en América Latina y el Caribe. *Washington, DC: Banco Interamericano de Desarrollo*.

INEGI. (2024). *Indicadores de Bienestar*. Recuperado de: <https://www.inegi.org.mx/app/bienestar/>

México, ¿cómo Vamos? (30 de enero de 2024). *La economía mexicana creció 3.1% en 2023 y alcanzó un nuevo máximo histórico*. Recuperado de: <https://mexicocomovamos.mx/la-economia-mexicana-crecio-3-1-en-2023-y-alcanzo-un-nuevo-maximo-historico/>

Muggah, R., & Aguirre Tobón, K. (2018). *Citizen security in Latin America: The hard facts*. Inter-American Development Bank.

Pérez González, L. (2016). La Transformación del Sistema Educativo de Corea del Sur entre 1930 y 1980, Calidad y Equidad. *Revista Iberoamericana de Educación Superior (RIES)*, 7(19), 128-146.

Quiroz Becerra, V., Cisneros, J., & Pérez, E. (2019). *Chicago neighborhood networks: Stability and change*. Chapin Hall at the University of Chicago.

Torres, I. (27 de septiembre de 2023). *Inteligencia Artificial para el bienestar social*. CONECTA. Recuperado de: <https://conecta.tec.mx/es/noticias/nacional/investigacion/inteligencia-artificial-para-el-bienestar-social>

INEGI. (2020). *Pobreza y desigualdad 2020*. Recuperado de: <https://www.inegi.org.mx/programas/pobreza/2020/>

Guerrero Gómez, E., & Torrijos Muñoz, M. T. (2024). Descifrando los factores que contribuyen al bienestar del mexicano. Facultad de Ciencias de la Computación.