



**BENEMÉRITA UNIVERSIDAD AUTÓNOMA  
DE PUEBLA**

---

---

**FACULTAD DE CIENCIAS DE LA COMPUTACIÓN**

**MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

**MODELO DE INTELIGENCIA DE NEGOCIOS PARA LA  
PROMOCIÓN DE LA PARTICIPACIÓN CIUDADANA EN LAS  
ELECCIONES FEDERALES DE MÉXICO**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**MAESTRO EN CIENCIAS DE LA COMPUTACIÓN**

**P R E S E N T A:**

**LIC. TOMÁS DÍAZ TAPIA**

**DIRECTORA DE TESIS:**

**MARÍA TERESA TORRIJOS MUÑOZ**

**CO-ASESOR:**

**CARLOS PALOMINO JIMENEZ**

**ENERO 2017**



# Contenido

<b>1. ANTECEDENTES</b> .....	5
<b>1.1. Antecedentes de la Participación ciudadana.</b> ....	5
1.1.1. Participación ciudadana en el ámbito internacional.....	5
1.1.2. Participación electoral en el ámbito nacional.....	5
1.1.3. Participación ciudadana en el ámbito estatal .....	6
1.1.4. Sistemas de Inteligencia.....	7
1.1.5. Participación ciudadana en el ámbito de inteligencia de negocios .....	9
<b>1.2. Justificación y propósito de la investigación</b> .....	9
<b>1.3. Objetivos</b> .....	10
<b>1.4. Alcances y limitaciones del proyecto de investigación</b> .....	11
<b>1.5. Metodología de la investigación</b> .....	11
<b>1.6. Impacto social esperado</b> .....	12
<b>2. MARCO TEORICO</b> .....	14
<b>2.1. Dominio de aplicación.</b> .....	14
<b>2.2. Inteligencia de Negocios</b> .....	15
<b>2.3. Redes Neuronales Artificiales</b> .....	17
<b>2.4. Algoritmo de Red Neuronal de Microsoft</b> .....	21
2.4.1. Combinación y Activación .....	22
2.4.2. Backpropagation (Retro propagación), Función de Error y Gradiente Conjugado .....	22
2.4.3. Topología de la Red.....	24
2.4.4. Entrenamiento de la Red Neuronal.....	25
2.4.5. Interpretación del Algoritmo de Red Neuronal .....	25
<b>2.5. Modelo de análisis multidimensional.</b> .....	26
<b>3. TRATAMIENTO DE LOS DATOS.</b> .....	30
<b>3.1. Selección de datos.</b> .....	30
3.1.1. Fuente de datos INE .....	30
3.1.2. Fuente de datos INEGI.....	31
<b>3.2. Extracción de datos</b> .....	32
<b>3.3. Transformación y depuración de los datos.</b> .....	33
<b>3.4. Carga de datos</b> .....	35

<b>3.5. Indicadores clave</b> .....	36
<b>3.6. Cruce de datos y última depuración</b> .....	44
<b>4. MULTIDIMENSIONALIDAD: INTELIGENCIA DE NEGOCIOS.</b> .....	48
4.1. Modelo multidimensional en el dominio de aplicación.....	48
<b>5. INTELIGENCIA DE NEGOCIOS</b> .....	54
<b>5.1. Análisis de la Estructura de minería de datos</b> .....	54
5.1.1. Crear Estructura de minería de datos .....	54
5.1.2. Algoritmo de minería de datos .....	55
5.1.3. Procesar Estructura de minería de datos.....	56
5.1.4. Predicción de la Estructura de minería de datos .....	58
<b>5.2. Análisis de datos</b> .....	60
<b>5.3. Resultados de la predicción</b> .....	62
<b>5.4. Rangos para Predicción</b> .....	68
<b>6. RESULTADOS Y CONCLUSIONES</b> .....	72
<b>Bibliografía</b> .....	78

# Tabla de ilustraciones

FIG. 1. EJEMPLO DE RED NEURONAL.....	21
FIG. 2. VISOR DE CONTENIDO DE RED NEURONAL.....	26
FIG. 3. EJEMPLO DE MODELO DE ESTRELLA.....	27
FIG. 4. EJEMPLO DE MODELO DE CASCADA.....	28
FIG. 5. VISOR PARA ANALIZAR DATOS DEL CUBO.....	28
FIG. 6. SISTEMA DE CONSULTA DE LA ESTADÍSTICA DE LAS ELECCIONES FEDERALES.....	31
FIG. 7. DATOS QUE ARROJA EL INE DESPUÉS DE APLICAR LOS FILTROS SELECCIONADOS.....	31
FIG. 8. PORTAL DEL INEGI DONDE ESTÁN LOS DATOS A ESCALAS GEOELECTORALES.....	32
FIG. 9. RESULTADO DE UNIR LAS 256 TABLAS CON LOS RESULTADOS ELECTORALES.....	33
FIG. 10. RESULTADO DE UNIR LAS TABLAS DEL CENSO DE POBLACIÓN 2010.....	34
FIG. 11. EJEMPLO DE FILAS QUE SE ELIMINARON.....	35
FIG. 12. INTERFAZ DEL ASISTENTE PARA LA IMPORTACIÓN DE TABLAS EN SQL SERVER.....	35
FIG. 13. TABLA CON LOS ÍNDICES SOCIOECONÓMICO.....	39
FIG. 14. ÍNDICE DE PARTICIPACIÓN CIUDADANA EN ELECCIONES.....	40
FIG. 15. NIVELES QUE SE CONSIDERAN PARA LA DISCRETIZACIÓN DE LOS ÍNDICES.....	41
FIG. 16. ESTRUCTURA DEL CUBO DE INFORMACIÓN.....	48
FIG. 17. EJEMPLO DE DISTRIBUCIÓN DE TOTALES.....	49
FIG. 18. MEDIDA [RES_ELECCIONES].....	50
FIG. 19. COMBINACIÓN DE LOS 7 ÍNDICES.....	51
FIG. 20. CREAR ESTRUCTURA DE MINERÍA DE DATOS.....	55
FIG. 21. MODELO DE MINERÍA DE DATOS Y EL ALGORITMO A EMPLEAR.....	56
FIG. 22. PROCESAMIENTO DE LA ESTRUCTURA DE MINERÍA DE DATOS.....	58
FIG. 23. PREDICCIÓN DE LA PARTICIPACIÓN CIUDADANA EN ELECCIONES.....	60
FIG. 24. ESTRUCTURA DE MINERÍA DE DATOS CON DATOS DISCRETOS.....	60
FIG. 25. ESTRUCTURA DE MINERÍA DE DATOS CON DATOS CONTINUOS.....	61
FIG. 26. GRÁFICO DE ELEVACIÓN CON MODELO DE DATOS CONTINUOS.....	61
FIG. 27. GRÁFICO DE ELEVACIÓN CON MODELO DE DATOS CONTINUOS.....	62
FIG. 28. TABLA CON LA PREDICCIÓN DEL ALGORITMO DE RED NEURONAL.....	63
FIG. 29. RED DE DEPENDENCIA.....	64
FIG. 30. NUEVA RED DE DEPENDENCIA.....	65
FIG. 31. ENTRENAMIENTO DE RED NEURONAL DE LA TOOLBOX DE MATLAB.....	67
FIG. 32. RANGOS DE PREDICCIÓN 1.....	68
FIG. 33. RANGOS DE PREDICCIÓN 2.....	68
FIG. 34. SALIDAS DE LA RED NEURONAL.....	69
FIG. 35. COMPARACIÓN DE LOS DOS METODOS DE ANÁLISIS.....	75
FIG. 36. SEGUNDO MODELO ENCONTRADO.....	76

# **CAPITULO 1**

## 1. ANTECEDENTES

Dentro de la literatura especializada existe un acuerdo de que la participación ciudadana es una pieza clave en la vida de una democracia representativa. En las últimas décadas diversas democracias han experimentado una disminución en sus tasas de participación electoral. Los bajos índices de participación han llamado la atención de las instituciones políticas que consideran este hecho como un elemento negativo para la estabilidad y legitimidad de la democracia (IFE, 2013).

Los antecedentes de esta investigación están basados en una búsqueda de referencias en revistas de interés científico. La búsqueda se hizo en el acervo bibliográfico que proporciona la Benemérita Universidad Autónoma de Puebla a través de los recursos electrónicos que la Dirección General de Bibliotecas tiene a su disposición.

Las revistas de carácter científico en las cuales se investigó fueron: ACM, IEEE, Science Direct, Springer Link principalmente. Se buscó la información relacionada con el proyecto de investigación ingresando el constructo en el buscador de cada revista, filtrando por tipo de publicación, por área de publicación y que se mostraran únicamente artículos referentes a temas específicos

### 1.1. Antecedentes de la Participación ciudadana.

#### 1.1.1. Participación ciudadana en el ámbito internacional

La democracia tuvo tiempos turbulentos en América Latina. Durante años la región tuvo diversos regímenes militares y la democracia fue vista como frágil, temporal y superficial en contenido, pero en los últimos 25 años ha ido creciendo de tal modo que se considera ese desarrollo como una señal de madurez política, donde los ciudadanos de la región han evolucionado. Algunos piensan que este desarrollo de la democracia se debió a la liberalización económica y del libre comercio así como a la influencia de Estados Unidos (Smith, 2004).

En América Latina las tendencias han mostrado que en los últimos sesenta años, la participación ciudadana creció notablemente en la región. Al comparar a México con el resto de los países de la región, los niveles de participación no llegan al promedio (IFE, 2013).

La no legitimidad de los gobiernos y de liderazgo político en América Latina, llevó a situaciones económicas, políticas y sociales insostenibles, lo que ha provocado la falta de credibilidad en los procesos electorales ciudadanos, lo que se traduce en un mayor abstencionismo en las elecciones. En 2006 el porcentaje de participación ciudadana en las elecciones disminuyó treinta puntos en América Latina a comparación del 70% entre 1986 y 1996. Entre 2007 y 2010 la participación aumento veinte puntos a comparación del 2006.

Esos cambios tan bruscos en la participación político electoral en América Latina podría estar sustentado en la falta de confianza (Mendieta, 2012).

En esta misma línea, en un estudio de participación ciudadana en el Reino Unido, Kolovos & Harris (2005) menciona como conclusión, que los dos principales factores que influyen en la participación en Reino Unido son la edad y la identificación partidista. Algunos otros factores como el estatus social, el nivel educativo y el género no tienen una influencia fuerte, pero la intensa competencia política y las divisiones ideológicas entre los partidos políticos tienden a una mayor participación.

La situación de Suecia es interesante también, en las elecciones de 2002, 2006 y 2010 el 80% de la población emitió su derecho a votar. Dichos años electorales sirvieron para hacer un estudio de la participación ciudadana en donde las variables que incluyeron en el modelo a estudiar fueron: las variables demográficas que incluían sexo, estado civil, edad lineal, municipio en el que viven, variables de socio economía, escolaridad y variables de migración que incluían país de origen, años de inmigración y ciudadanía. En el estudio encontraron que el nivel de escolaridad de los inmigrantes influye en la baja participación electoral, en general, entre más bajo el nivel es escolaridad menos probabilidad hay participen en las elecciones (Bevelander, 2015).

Otro caso documentado, es el de Estados Unidos, donde Radcliff (2001) hace un análisis de las elecciones en el periodo de 1950 a 1994. Encontrando que la principal influencia en la disminución de la participación ciudadana es el cambio generacional, las estrategias partidistas y la disminución de los niveles de eficacia del partidismo e interés de los votantes. El modelo empleado para el análisis, combina un conjunto normal de predictores de participación demográficos y actitudinales en donde las variables que se consideran son el estado civil, situación financiera, edad, situación laboral, género, raza, ingresos, educación, ocupación, movilidad residencial, fuerza partidista, eficacia política, que tan frecuente se informa sobre campañas políticas en los periódicos, si vivió en algún estado del sur y cuáles fueron sus tendencias a la hora de votar en diferentes años electorales.

### 1.1.2. Participación electoral en el ámbito nacional

INE es el organismo encargado de organizar las elecciones federales en México, y a través del estudio censal de la participación ciudadana en las elecciones Federales de 2012 disponible creado por IFE, (2013) fueron basados las secciones 1.1.2 y 1.1.3.

Se informa que la participación ciudadana llegó al 62.08%; 3.58% más alto que en el 2006, pero aun así este porcentaje se encuentra muy por debajo del máximo histórico alcanzado en la primera elección presidencial de la cual el IFE fue la encargada en 1994 de 77.16%.

El INE concuerda en que para explicar el comportamiento de los ciudadanos a la hora de emitir su voto no debe limitarse al estudio de las instituciones involucradas, sino también se tiene que analizar las variables demográficas como son sexo, edad, región, etc.; las cuales fueron identificadas como posibles determinantes de la participación electoral.

En el estudio hecho por el IFE en 2013 se reporta que la participación nacional en las elecciones federales, respecto a el sector femenino llegó al 66.08% superando al sector masculino el cual obtuvo un total de 57.77%. En este se considera que la diferencia entre los sexos se debe al aumento de la población femenina en el mercado laboral, a un aumento en el porcentaje de mujeres que están en puestos de toma de decisiones o incluso a los discursos de candidatos dirigidos a los intereses de las mujeres. El estudio termina haciendo una comparación con otros países principalmente con Estados Unidos en donde la tendencia en la que el porcentaje de participación electoral es mayor en el sector femenino que con su contraparte alcanzando hasta 9 puntos de diferencia.

En el análisis que hizo el IFE de la participación ciudadana según la edad, encontró que los niveles más altos de participación se concentran en la población de 40 a 79 años siendo los adultos mayores a 80 años y los jóvenes en edad de 20 a 30 años los que menos participan. Las personas en edad de votar que por primera vez emiten su voto alcanzan una tasa de participación alrededor del 60%.

Al revisar la participación a nivel nacional por tipo de sección se observa que la población urbana y la mixta tienen un porcentaje similar ya que la sección urbana supera a la mixta por sólo un .30% pero la población de la sección rural supera en participación a la urbana.

Se plantea que al hacer un análisis de los procesos electorales en México de 1961 a 2003 se muestra que no hay una relación entre la marginación y la participación electoral, esta última se puede encontrar en mayor o menor medida en zonas urbanizadas como en zonas rurales pero en periodos de transiciones democráticas las entidades con menor marginación tiende a ser mayor la participación electoral.

La parte de participación a nivel nacional termina concluyendo que las mujeres en la edad de 18 a 49 años de poblaciones rurales participan más que las urbanas entre un 3.80% y 5.18% por arriba. Comparando la zona rural y la zona mixta se observa que la zona rural está por arriba de la mixta entre 3.33 y 4.29%. Cuando se habla del sector de hombres, en todos los rubros las zonas rurales tienen una mayor participación que las demás.

### 1.1.3. Participación ciudadana en el ámbito estatal

En el estudio que hizo el IFE acerca de la participación ciudadana en las elecciones federales del 2012, se hace la observación que en 13 entidades de México la participación fue mayor a la media nacional de 62.08%. Se destaca que en las entidades en donde hubo elecciones simultáneas, es decir hubo elección de presidente de la república, Gobernador del Estado y en algunos casos de presidentes municipales, la participación ciudadana en los comicios fue mayor que en las que únicamente se eligió Presidente de México, siendo que en las primeras la tasa de participación del 63.50% superando por 4.15 puntos porcentuales a las demás que tuvieron una tasa de 59.35%.

En el análisis detallado que hizo el IFE (2013) de los Estados donde hubo elecciones concurrentes y en las que no se expone que el 22% de ellos presentó una participación alta,

más de la mitad de los estados de la República tuvieron una participación media y el último 22% de los estados fue baja tomando como intervalos de [51%-60%] para el nivel de participación bajo, de [60%-68%] como un nivel medio y por último de [68%-77%] como un rango de participación alto. Se resalta que los estados del norte son los que tienen menos participación y los estados de oriente son los que muestran los niveles más altos.

En todas las entidades del país se observa que la mayor participación fue la de las mujeres, como por ejemplo en Guanajuato donde la diferencia entre la participación femenina y la masculina fue de 10.54 puntos.

En cuestión de edad, la tendencia que hubo a nivel nacional siguió a nivel estatal siendo que la participación de los jóvenes fue la menor en los rangos de edad de 20 a 39 años y a comparación con el rango de edad de 40 a 79 años, este último estuvo 15.24 puntos porcentuales arriba en participación.

Teniendo en cuenta las secciones de cada estado las zonas rurales son las que siguen teniendo el índice más alto de participación, siendo que se obtuvo un índice de 63.80% seguidas de las zonas urbanas con un 60.86% y las zonas mixtas con el 60.75%. Se destaca que en el Distrito Federal, Guanajuato y Querétaro, el porcentaje de votación de las zonas rurales fue menor a la de las demás zonas.

En general se puede observar que las tendencias que hubo en las elecciones a nivel nacional se replicaron a nivel estatal a excepción de algunos especiales.

Con base en estas investigaciones se propone estudiar la participación ciudadana considerando como variables que la define: resultados electorales de 1991 a 2015, índices de marginación, tipos de zonas, edades de los votantes, escolaridad, situación laboral.

Podrían definirse más variables, pero eso se verá al continuar desarrollando el proyecto de investigación.

#### 1.1.4. Sistemas de Inteligencia

López, Fernández, Machado, Agustín y Zabala (2006) mencionan en Pérez y Valdez (2014) que en la actualidad, tener un sistema de información es muy importante para las organizaciones, ya que les permite estar en constante innovación y desarrollo. También comentan que las herramientas de gestión empresarial son las que definen los sistemas de innovación y desarrollo, tales como la gestión de conocimiento, el diseño industrial y la inteligencia competitiva.

De acuerdo a Gibbons y Prescott (1996) *“la inteligencia competitiva se ha desarrollado en el ámbito de la dirección estratégica de las organizaciones como el proceso de la obtención, análisis, interpretación y difusión de información de valor estratégico sobre la industria y los competidores, que se transmite a los responsables de la toma de decisiones en el momento oportuno”*.

En este mismo contexto se incluye la vigilancia de los competidores, el seguimiento sobre la evolución de los mercados electorales, la vigilancia tecnológica y la vigilancia de su entorno. La vigilancia tecnológica contempla las nuevas tecnologías disponibles y recién surgidas que existen y se puedan usar en las campañas electorales. (Valdez, 2012).

En el ámbito de tecnología y como instrumento, “*los sistemas de información permiten generar ventajas competitivas sostenibles*” (Pérez y Amador, 2009). Por lo tanto la inteligencia competitiva logra campañas electorales exitosas, en la que la información, el conocimiento y la tecnología sirven para generar estrategias para la toma de decisiones (Valdez, 2012).

Según Pérez y Valdez, (2014) no existe un modelo o sistema de información de inteligencia competitiva para los partidos políticos. La inteligencia competitiva empresarial puede servir para ayudar a construir un modelo para los partidos políticos y campañas electorales convirtiéndose en un factor que determine las ventajas competitivas logrando el triunfo en una campaña electoral.

En esta línea y respecto a la revisión de la literatura, se han encontrado trabajos de investigación en el área de Ciencias Sociales, enfocados principalmente a la importancia de la participación ciudadana y la forma en que afecta la vida democrática y estabilidad de un país. Sin embargo no se encontró evidencia del uso de modelos matemáticos enfocados a la inteligencia de negocio.

Buscando referencias acerca de la inclusión de Business Intelligence a sectores públicos, se encontró el artículo de Hartley y Seymour, (2011), que expresaba que el Gobierno de Sudáfrica cuenta con un Sistema de Gestión de Desempeño (Performance Management System) capaz de la prestación de servicios. La investigación se centra en poder hacer que el sector público haga uso de BI para entender a los ciudadanos de Sudáfrica y hacer un uso efectivo de los recursos, además que los planes y decisiones del Gobierno puedan ser analizados con ayuda de análisis multidimensionales de todos los datos relevantes. Se describen los beneficios de BI en el sector Gobierno:

- La facilidad de obtener información descifrable y completa sin la necesidad de herramientas sofisticadas.
- Análisis extensos de datos que proporciona respuestas a búsquedas exhaustivas.
- Ayuda a formular estrategias efectivas y políticas para la facilitación a los ciudadanos.

También se comenta que el uso de sistemas de BI o sistemas de gestión de desempeño no garantiza el mejoramiento de la prestación de servicios. Sin embargo Maila, (2009), expone que una investigación está necesitada de descubrir como una organización del sector público pueda aplicar efectivamente sistemas BI en conjunto de políticas funcionales, sistemas e instrumentos para mejorar la prestación de servicios. El artículo se centra en encontrar los factores que obstaculizan el que las organizaciones del sector público adopten los sistemas

BI y son la *información, tecnología, procesos* (sólo se puede aplicar BI cuando hay procesos estructurados), *objetivos y valores, tener personal capacitado*.

#### 1.1.5. Participación ciudadana en el ámbito de inteligencia de negocios

Gill, en 2008 presenta un trabajo de investigación en donde considera que la previsión electoral es un desafío para las ciencias políticas y en su artículo se describe un modelo de redes neuronales que es el perceptrón multicapa para el pronóstico de resultados electorales. El autor realizó una encuesta basado a como se comportó el gobierno y los avances que tuvo, y las preguntas estuvieron diseñadas de tal forma se respondiera sí o no. Los datos fueron entrenados en un perceptrón de dos capas dándole como resultado las tendencias de elección para los comicios próximos en la India.

También se encuentra el caso de Irán, donde Khaze, Masdari, & Hojjatkah (2013) tratan de estimar la participación en las elecciones. Los datos incluían datos y perspectivas de una muestra de 100 actitudes individuales de los ciudadanos. Realizaron una red neuronal en matlab con 10 capas ocultas, 10 datos de entrada y 3 de salida. Al final pudieron predecir que el índice de participación de los ciudadanos en las elecciones de Irán sería del 91%.

Akinwonmi, Kuboye, & Thompson (2013), presentaron un trabajo para seleccionar al candidato adecuado de un partido político para las elecciones generales en Nigeria utilizando redes neuronales. Los atributos que se utilizaron para analizar datos fueron características con las cuales partidos políticos seleccionaban a sus candidatos como son: educación, experiencia previa, promesas de campaña, ideas sobre la educación, planes sobre agua potable, electricidad, generación de empleo y salud. Al final los datos que sacó la red neuronal mostraba la persona que sería más apta para ser el candidato para las siguientes elecciones.

Para concluir, Hayashi, Hsieh & Setiono (2010) describen una aplicación de BI utilizando redes neuronales para analizar la heterogeneidad de los consumidores cuando comen fuera de su casa en Taiwan. El *data set* que utilizaron fue una colección de datos de 800 consumidores taiwaneses el cual contenía información demográfica como el género, la edad y los ingresos que obtenían las personas, así como sus tendencias psicológicas para comer. Como resultado, pudieron encontrar los distintos segmentos en los que se dividen los consumidores y predecir el comportamiento de estos.

## 1.2. Justificación y propósito de la investigación

Los tipos de gobierno, federal, estatal y municipal así como los partidos políticos requieren de instrumentos que permitan mejorar su quehacer institucional en lo que a procesos

electorales se refiere y en los que la participación ciudadana desempeña un papel relevante para la vida democrática del país

Dado el impacto que genera la participación ciudadana en los resultados de los procesos electorales en las naciones de todo el mundo, múltiples investigaciones se han llevado a cabo para determinar los factores que inciden en el comportamiento de los ciudadanos.

En México, entre 1994 y 2009 hubo tres elecciones presidenciales, en donde la participación ciudadana pasó de 77% a 59%. En las elecciones intermedias dentro del mismo periodo el abstencionismo pasó del 24 a 55% (Centro de Estudios y Opinión Pública, 2012). En las elecciones federales del 2012, se registró un repunte en la tasa de participación electoral de 62.08, lo que representa un incremento en relación a lo observado en 2006, sin embargo, este porcentaje es aún inferior a los niveles de registrados en las votaciones de 1994 en las que se reportó una tasa de 77.16%. Esta situación nos puede llevar a una desestabilidad democrática así como la falta de legitimidad de las instituciones políticas (IFE, 2013).

En cada elección federal se tiene un gasto público que asciende a los miles de millones de pesos que a través de los impuestos, y lo que se esperaría es que se viese reflejado en una mayor participación ciudadana y por ende en la estabilidad y legitimidad de un estado democrático. Por lo tanto, se requiere de un análisis detallado de los resultados para comprender el origen los cambios de tendencia y el comportamiento de los votantes (LXI Legislatura de la Cámara de Diputados, 2012).

En este sentido, los procesos electorales que se han llevado a cabo en el periodo 1991-2012 en México, han generado un cúmulo de datos los cuales tienen un gran potencial para generar información relevante y útil, para el INE, los partidos políticos y los que estudian el tema.

Por otro lado, el uso eficiente de la inteligencia de negocios da la oportunidad de hacer uso de la tecnología y modelos matemáticos para generar conocimiento que permita soportar estrategias para el diseño de políticas y acciones en educación cívica y promoción del voto.

### 1.3. Objetivos

#### *Objetivo General*

El objetivo general del proyecto de investigación es hacer uso de las tecnologías de inteligencia de negocios para analizar los factores que inciden en la participación ciudadana, en el ejercicio del derecho al sufragio en las elecciones federales para representantes de la cámara de Diputados de México, con el fin de soportar la toma de decisiones que permitan elaborar estrategias de promoción al voto.

Los factores que se consideran son los proporcionados por las fuentes oficiales: INEGI e INE a nivel nacional.

## *Objetivos Específicos*

- Comprender el dominio de aplicación.
- Identificar los datos y seleccionar los métodos para análisis.
- Extraer, Transformar y Cargar los datos (ETL).
- Implementar modelo de inteligencia de negocios.
- Reportar resultados y conclusiones.

### 1.4. Alcances y limitaciones del proyecto de investigación

El trabajo de investigación considera el diseño de un modelo basado en Inteligencia de Negocios que brinde soporte la toma de decisiones de las instituciones políticas para la promoción de la participación ciudadana para emitir su voto en las elecciones federales de Diputados, Senadores y Presidente de México, por los procesos electorales de 1991 a 2012 considerando los datos del proceso electoral 2014- 2015 que aun el INE no ha publicado de manera oficial.

El análisis en el que se enfocará esta investigación será para las elecciones federales. Podría adaptarse la investigación en trabajos futuros a un estado en particular o a otro país, pero el alcance de este trabajo es hacer el diseño del modelo adaptado a México como país.

### 1.5. Metodología de la investigación

#### 1. Fase conceptual

- 1.1. Formulación y delimitación del problema
- 1.2. Revisión de literatura sobre la participación ciudadana en elecciones y trabajos relacionados con BI
- 1.3. Creación de un marco teórico
- 1.4. Formulación de hipótesis de investigación

#### 2. Fase de diseño y planificación

- 2.1. Selección de diseño que tendrá el modelo.
- 2.2. Identificación de los datos a estudiar
- 2.3. Especificación de los métodos para obtener los datos
- 2.4. Diseño del plan de análisis de datos
- 2.5. Terminación y revisión del plan de investigación
- 2.6. Realización y revisión del estudio sobre los datos recabados

#### 3. Fase empírica

- 3.1. Recolección de datos
- 3.2. Preparación de los datos para el análisis

#### 4. Fase analítica

- 4.1. Análisis de los datos
- 4.2. Interpretación de los resultados

## 5. Fase de difusión

### 5.1. Comunicación de los resultados

#### 1.6. Impacto social esperado

El impacto que tendrá nuestra investigación está en la población mexicana en edad de votar y en las instituciones políticas y de gobierno de México, ya que al aumentar la participación ciudadana en las elecciones se podrá garantizar una legitimización de las instituciones de gobierno.

# **CAPITULO 2**

## 2. MARCO TEORICO

### 2.1. Dominio de aplicación.

El Instituto Nacional Electoral, desde su creación en 1994, es el encargado de organizar las elecciones federales en México. En cada elección el Consejo General del IFE ha solicitado la realización de estudios de la documentación utilizada en los procesos electorales federales. Esto para generar información relevante para el propio instituto y partidos políticos considerando tres variables como lo son sexo, edad y tipo de sección, para generar políticas en materia electoral y la promoción del voto.

En el artículo de participación ciudadana y políticas públicas, Martínez define a la participación ciudadana como el involucramiento de los ciudadanos en el espacio físico del Estado, es decir de una comunidad política. Dicha participación está vinculada con la relación gobierno sociedad que se desea y a los modelos de democracia.

Se define democracia electoral como la existencia de elecciones libres y justas. La mayoría de los adultos deben de tener el derecho al voto y debe existir una competencia genuina entre los candidatos que buscan ocupar puestos públicos. La ausencia de elecciones significa que un país no es democrático. Para identificar el cambio político en cada país se clasifica de la siguiente manera (Smitt, 2004):

- Democrático, es cuando los líderes nacionales asumieron o mantuvieron el poder como resultado de elecciones libres y justas; esto es, cuando hubo una competencia abierta por el apoyo de una parte sustancial de la población adulta.
- Semidemocrático, cuando los líderes llegaron al poder mediante elecciones que fueron libres, pero no justas, ya que solo un candidato tenía posibilidades reales de ganar, o cuando los líderes electos fueron obligados a compartir el poder efectivo con grupos no electos o a cederles el poder.
- Oligárquico, cuando la competencia electoral fue libre y/o justa, pero limitada a las élites dominantes y el sufragio estaba restringido a una proporción muy pequeña de la población adulta.
- No democrático o autocrático, en cualquier otro caso durante los años de golpe militar.

Participación política tiene que ver con el involucramiento de los ciudadanos en las organizaciones formales y mecanismos del sistema político: partidos, parlamentos, ayuntamientos, elecciones. Es una participación mediada por los mecanismos de la representación política (Martínez, Participación ciudadana y políticas públicas).

En las sociedades modernas la participación no se limita sólo a procesos electorales – participación política-, sino representa también una forma de participar, controlar y moderar

el poder otorgado a los representantes políticos a través de formatos y mecanismos de participación ciudadana, que fortalezcan y nutran la vida democrática de la sociedad (Pérez, 1999).

La participación ciudadana busca que los gobiernos se democraticen, mediante mecanismos como: la información, consulta y participación activa, con los cuales se busca una transformación social mediante las siguientes estrategias: Políticas públicas de mayor calidad, Legitimidad de las instituciones, Ciudadanía activa. Para concluir, la participación ciudadana es un derecho, y como ciudadanos de un mundo moderno se deben ejercer esas libertades que se brindan al estar en un espacio público de una tierra en vías de la democracia (Guillen, Sáenz, Badii, & Castillo, 2009).

## 2.2. Inteligencia de Negocios

Rajan, (2009) define Inteligencia de Negocios como *“el proceso de tomar grandes cantidades de datos, analizarlos y realizar un reporte de alto nivel que condense la esencia de los datos en simples acciones de negocio, que permita la toma de decisiones para los negocios”*.

El término de Inteligencia de Negocios apareció a finales de los años 50, pero ese concepto se sigue usando hoy en día. El término de inteligencia de negocios está va de la mano con los sistemas de información para la toma de decisiones empresariales. Los sistemas de información de los proporciona ventajas como son: los reportes, prevención y predicción, tendencias de análisis, capacidades a profundidad y factores de éxito. Estas ventajas se integraron en los sistemas de BI. Pero fue hasta los años 2000 cuando el uso de las tecnologías de BI fueron evidentes ya que muchas herramientas de análisis de datos fueron implementadas. (Czernicki, 2010).

La Inteligencia de Negocios, tiene dos diferentes significados que se relacionan con el término de inteligencia. El primero es menos frecuente, que es la capacidad aplicar la inteligencia humana a las actividades de negocios. Business Es un campo de investigación que aplica las facultades del conocimiento humano y las tecnologías de inteligencia artificial a la toma de decisiones en diferentes problemas de negocios, la segunda está relacionada con el valor de la información (Rajan, 2009).

Teniendo en cuenta el panorama de las principales tendencias de negocio, lo que amplió del concepto de BI fue: la globalización, el rápido crecimiento de los negocios, los mercados más competitivos, las grandes cantidades de datos e información, las arquitecturas cliente servidor, los recursos de planeación empresarial, la tecnología de data warehouse, la inteligencia artificial y la web basado en tecnologías y aplicaciones (Czernicki, 2010).

BI tiene varios componentes como lo son:

**OLAP (On-line analytical processing)** proporciona multidimensionalidad, vistas resumidas de datos de negocios y es usado para reportes, análisis, modelado y planeación para optimizar negocios. Las técnicas de OLAP y sus herramientas pueden ser utilizadas como

almacenadores de datos o data marts diseñadas para sofisticadas empresas en sistemas de inteligencia. Estos procesos de sistemas de consulta son requeridos para descubrir tendencias y analizar factores críticos (Rajan, 2009).

OLAP permite obtener acceso a datos organizados y agregados de orígenes de datos empresariales, organiza subconjuntos de datos con una estructura multidimensional de manera que represente un significado especial o responda a una pregunta en particular y se pueden definir tres tipos de sistemas OLAP (Gomez & Butista, 2009).

- **ROLAP:** Implementación que almacena los datos en un motor relacional. Típicamente, los datos son detallados, evitando las agregaciones y las tablas se encuentran normalizadas.
- **MOLAP:** Esta implementación almacena los datos en una base de datos multidimensional. Para optimizar los tiempos de respuesta, el resumen de la información es calculado por adelantado.
- **HOLAP (OLAP Híbrido):** Almacena algunos datos en un motor relacional y otros en una base de datos multidimensional.

**Análisis avanzado**, se refiere a como la minería de datos, el análisis de prevención y predicción toma ventajas con las técnicas de análisis estadístico para predecir o prevenir posibles hechos (Rajan, 2009).

**La gestión del rendimiento corporativo**, es el que conecta las piezas para que puedan ser visibles los resultados, como lo son los tableros de control que muestran métricas financieras combinadas con aprendizaje organizacional y métricas de crecimiento.

**Inteligencia de Negocios de tiempo real**, permite en tiempo real la distribución de métricas mediante correo electrónico, sistemas de mensajes e interfaces interactivas.

El **Data Warehouse** es un componente muy importante para BI ya que apoya la propagación física de los datos por el manejo de los numerosos registros empresariales para tareas de integración, limpieza, agregación y tareas de consulta. También puede contener los datos operacionales con los cuales se puede definir como un conjunto actualizable de integración de datos para la toma de decisiones de toda empresa.

Como complemento del data warehouse la arquitectura de los sistemas de BI comprenden otros tres componentes: *businnes analitics* (análisis de negocios) como una colección de herramientas para consulta de datos y reportes, procesamiento analítico en línea, análisis estadístico, predicción y visualización de los datos, minería de datos. *Business Performance Management* usado para el monitoreo y análisis de rendimiento en donde se sacan los indicadores. Por último la *Interfaz* de usuario que es en la cual el usuario interactúe con los sistemas de BI (Czernicki, 2010).

**Data sources (datos fuente)**, pueden ser bases de datos operacionales, datos históricos, datos externos como por ejemplo de internet. Las bases de datos pueden ser relacionales o de cualquier otra estructura que soporte la línea de aplicaciones de negocios.

Para propósitos de análisis la Inteligencia de Negocios utiliza software de estadística y técnicas de Machine Learning. La combinación de estadísticas y análisis predictivo se considera como una de las tendencias cruciales en BI (Czernicki, 2010).

En conclusión, BI proporciona una manera rápida y efectiva de recopilar, abstraer, presentar, formatear y distribuir la información de sus fuentes de datos, permitiendo a los profesionales visualizar y analizar datos precisos sobre las actividades fundamentales del negocio y utilizarlas para mejorar la toma de decisiones y la planificación estratégica (Gómez & Bautista, 2009).

### 2.3. Redes Neuronales Artificiales

En el libro de Hilera & Martínez (2000), menciona que las redes neuronales artificiales, son redes interconectadas masivamente en paralelo de elementos simples y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico.

El cerebro humano contiene un complejo conjunto de procesos interconectados o neuronas que trabajan juntos para resolver problemas. La experiencia es vital para que la red trabaje bien. El cerebro trabaja más eficientemente cuando aprende a reconocer patrones. Es similar con una red neuronal, ya que puede ser entrenada a reconocer patrones en los datos que otras técnicas estadísticas o computacionales pueden fallar. Una vez entrenada la red puede ser usada para hacer pronósticos (Gill, 2008).

Las redes neuronales artificiales son consideradas como un moderno método matemático-computacional los cuales son usados para resolver problemas inesperados en sistemas conductuales. En cualquier problema que los métodos matemáticos y computacionales han fallado al reconocer patrones las redes neuronales son capaces de resolverlos; actualmente las redes neuronales son utilizados por la capacidad que tienen de predecir (Khaze, Masdari & Hojjatkah, 2013).

En el libro de Hilera y Martínez, (2000) se menciona las ventajas que ofrecen las redes neuronales y son:

*Aprendizaje adaptativo:* Capacidad de aprender a realizar tareas basadas en un entrenamiento o experiencia inicial.

*Auto-organización:* La red neuronal puede crear su propia organización o representación de la información que recibe mediante una etapa de aprendizaje.

*Tolerancia a fallos:* La destrucción parcial de una red conlleva a una degradación de estructura; sin embargo, algunas capacidades de la red se pueden retener, incluso sufriendo un gran daño.

*Operación en tiempo real:* Los computadores neuronales pueden ser realizados en paralelo, y se diseñan y fabrican máquinas con hardware especial para obtener esta capacidad.

*Fácil inserción dentro de la tecnología existente:* Se pueden obtener chips especializados para redes neuronales que mejoran su capacidad en ciertas tareas. Ello facilitará la integración modular en los sistemas existentes.

Las redes neuronales son una tecnología computacional emergente que puede utilizarse en un gran número y variedad de aplicaciones, tanto comerciales como militares. Se pueden desarrollar redes neuronales en un periodo de tiempo razonable y pueden realizar tareas concretas mejor que otras tecnologías convencionales, incluyendo los sistemas expertos.

Hay muchos tipos diferentes de redes neuronales, cada uno de los cuales tiene una aplicación particular más apropiada. Sus aplicaciones están en la Biología, en las Empresas, en Medio Ambiente, Finanzas, Manufacturación, Medicina, Militares.

### *Tipos de Redes Neuronales*

Nombre de la red	Año	Aplicaciones más importantes	Comentarios	Limitaciones	Inventada/ desarrollada por
<b>Avalancha</b>	1967	Reconocimiento de habla continua	Ninguna red sencilla puede hacer todo esto	No es fácil alterar la velocidad o interpolar el movimiento	Stephen Grossberg
<b>ART</b>	1986	Reconocimiento de patrones	Sofística y poco utilizada	Sensible a la translación y distorsión y escala	Gail Carpenter, Stephen Grossberg
<b>ADALINE/MADALINE</b>	1960	Filtrado de señales	Rápida y fácil de implementar	Solo es posible clasificar espacios linealmente separados	Bernard Widrow
<b>Back propagation</b>	1974-85	Síntesis de voz de texto. Reconocimiento de patrones	Red más popular	Necesita mucho tiempo para el aprendizaje y muchos ejemplos	Paul Werbos, David Parker, David Rumelhart
<b>Memoria asociativa bidireccional</b>	1985	Memoria heteroasociativa de acceso por contenido	Aprendizaje y arquitectura simples	Baja capacidad de almacenamiento. Los datos deben ser codificados	Bart Kosko
<b>Máquinas de Boltzmann y Cauchy</b>	1985-86	Reconocimiento de patrones	Redes simples. Capacidad de representación	La máquina de Boltzmann necesita un tiempo muy	Jeffrey Hinton, Terry Sejnowski, Harold Szu.

			óptima de patrones	largo de aprendizaje	
<b>Brain-Estate-in-a-Box</b>	1977	Extracción de conocimiento de bases de datos	Posiblemente mejor realización que las redes de Hopfield.	Realización y potenciales aplicaciones no estudiadas totalmente	James Anderson
<b>Cerebellatron</b>	1969	Control de movimiento de los brazos de un robot	Semejante a Avalanche	Requiere complicadas entradas de control	David Marr, James Albus, Andres Pellionez
<b>Counter-propagation</b>	1986	Compresión de imagenes	Combinacion de Perceptron y TPM	Numerosas Neuronas y Conexiones	Robert Hecht-Nielsen
<b>Hopfield</b>	1982	Reconstrucción de patrones y optimización	Puede implementarse en VLSI. Fácil de conceptualizar	Capacidad y estabilidad	John Hopfield
<b>Neocognitron</b>	1978-84	Reconocimiento de caracteres manuscritos	Insensible a la translación, rotación y escala.	Requiere muchos elementos de proceso, niveles de conexiones	K. Fukushima
<b>Perceptron</b>	1957	Reconocimiento de caracteres impresos	La red más Antigua construida en HW	No puede reconocer caracteres complejos	Frank Resenblatt
<b>Self-Organizing-Map (SOM). Topology Preserving-Map (TPM)</b>	1980-84	Reconocimiento de patrones, codificación de datos, optimización	Realiza mapas de características comunes de los datos aprendidos	Requiere mucho entrenamiento	Teuvo Kohonen

Cualquier modelo de red neuronal consta de dispositivos elementales de proceso: las neuronas. A partir de ellas, se pueden generar representaciones específicas, de tal forma que un estado conjunto de ellas puede significar una letra, un número o cualquier otro objeto.

Las conexiones que unen a las neuronas que forman una RNA tienen asociado un peso, que es el que hace que la red adquiera conocimiento. Cada conexión entre la neurona  $i$  y la neurona  $j$  está ponderada por peso  $w$ .

En las redes neuronales artificiales, se puede considerar que el conocimiento se encuentra representado en los pesos de las conexiones entre neuronas; se puede decir que se aprende modificando los valores de los pesos de la red.

La distribución de neuronas dentro de la red se realiza formando niveles o capas de un número determinado de neuronas cada una. A partir de su situación dentro de la red, se pueden distinguir tres tipos de capas:

*De entrada:* Es la capa que recibe directamente la información proveniente de las fuentes externas a la red.

*Ocultas:* Son internas a la red y no tienen contacto directo con entorno exterior. El número de niveles ocultos puede estar entre cero y un número elevado. Las neuronas de las capas ocultas pueden estar interconectadas de distinta manera, lo que determina, junto con su número, las distintas topologías de redes neuronales.

*De salida:* Transfieren información de la red hacia el exterior.

Las redes neuronales multicapa son aquellas que disponen de conjuntos de neuronas agrupadas en varios niveles o capas. En estos casos, una forma para distinguir la capa a la que pertenece una neurona, consistiría en fijarse en el origen de las señales que recibe a la entrada y el destino de la señal de salida.

Existen dos tipos de redes de múltiple capa, las redes feedforward o redes con conexiones hacia adelante y las redes que disponen de conexiones hacia adelante y hacia atrás o redes feedback.

El perceptron, diseñado por Frank Rosenblatt en 1957 despertó un enorme interés debido a su capacidad de aprender a reconocer patrones sencillos: Perceptron, formado por varias neuronas lineales para recibir las entradas a la red y una neurona de salida, es capaz de decidir cuándo una entrada presentada a la red pertenece a una de las clases que es capaz de reconocer.

La única neurona de salida del Perceptron realiza la suma ponderada de entradas, resta el umbral y pasa el resultado a una función de transferencia de tipo escalos.

Sin embargo, el Perceptron, al constar sólo de una capa de entrada y otra de salida con una única neurona, tiene una capacidad de representación bastante limitada. Este modelo sólo es capaz de discriminar patrones sencillos y literalmente separables.

Otro modelo del Perceptron es el conocido como multicapa que se compone una capa de entrada, capas ocultas y capas de salidas. Están compuestas por una serie de neuronas que se encargan de recibir, procesar y enviar datos hacia otras neuronas procesando la información mediante distintas funciones matemáticas (Bayona & Salcedo, 2012). Esta red permite establecer regiones de decisión mucho más complejas que las de dos semiplanos, como lo hacía el Perceptron de un solo nivel.

El Perceptron básico de dos capas sólo puede establecer dos regiones separadas por una frontera lineal en el espacio de patrones de entrada. Un Perceptron con tres niveles de neuronas puede formar cualquier región convexa. Las regiones convexas se forman mediante la intersección entre las regiones formadas por cada neurona de la segunda capa.

El Perceptron multicapa no requiere de más de cuatro capas, ya que puede generar regiones de decisión arbitrariamente complejas. Sólo en ciertos problemas se puede simplificar el aprendizaje mediante el aumento del número de neuronas ocultas. Sin embargo, la tendencia es el aumento de la complejidad de la red y esto lleva al problema del número de neuronas que se deben seleccionar para un Perceptron con cuatro capas.

## 2.4. Algoritmo de Red Neuronal de Microsoft

En el libro de Tang y MacLennan (2005), se expresa que los nodos de entrada forman las primeras capas de la red. En la mayoría de las redes, cada nodo de entrada está relacionado con un atributo de entrada.

Los nodos ocultos son los nodos en las capas ocultas. Un nodo oculto recibe entradas de los nodos en la capa de entrada. Combina todas las entradas basados en los pesos asociados, procesa algunos cálculos y emite un resultado a la siguiente capa de nodos.

Los nodos de salida por lo regular representan los atributos predecibles. Una red neuronal podría tener múltiples atributos de salida. Es posible separar los nodos de salida en varias redes. Pero en la mayoría de los casos, eso reduce el tiempo de procesamiento cuando son combinados ya que puede compartir el costo común de escanear la fuente de datos. El resultado de los nodos de salida frecuentemente es un número flotante entre 0 y 1.

La predicción para las redes neuronales es sencilla, los valores de atributo de un caso de entrada es normalizada y mapeada a las neuronas de la capa de entrada. Después cada nodo de la capa oculta procesa las entradas y asigna una salida a la capa de neuronas siguiente. Al final, las neuronas de salida comienzan a procesar y generar un valor de salida. Este valor es mapeado a la escala original o categoría original.

Cuando se configura la red neuronal especificando el número de nodos ocultos, el proceso de entrenamiento encuentra el mejor conjunto de pesos para cada nodo en la red. Inicialmente los pesos son asignados aleatoriamente. Durante cada iteración de entrenamiento, la red procesa los casos de entrenamiento para generar predicciones en la capa de salida basada en la red configurada. Después se calcula los errores de las salidas. Basados en estos errores, se ajustan los pesos de la red usando propagación backward.

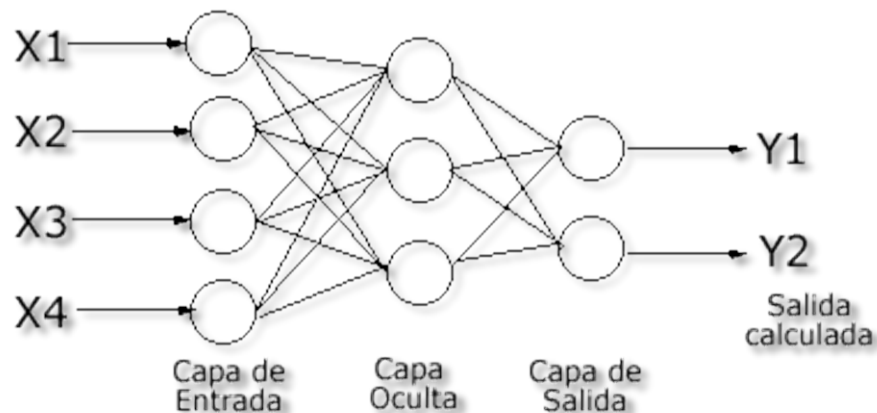


Fig. 1. Ejemplo de Red Neuronal.

### 2.4.1. Combinación y Activación

Cada neurona en la red neuronal es una unidad básica de procesamiento. Una neurona tiene un número de entradas y una salida. Combina todos los valores de entrada, hace los cálculos correspondientes y después asigna un valor de salida.

Una neurona cumple con dos funciones: que es la combinación de las entradas y el cálculo de las salidas. Se combinan los valores de las entradas en un solo valor. Hay diferentes formas de combinar las entradas y el método más popular es la suma ponderada. Esto significa que la suma de cada valor de entrada es multiplicada por su peso asociado. Otras funciones de combinación incluyen la media, el máximo lógico OR y el AND de los valores de entrada. La Red Neuronal de Microsoft utiliza el método de suma ponderada. Las combinaciones de salida son pasadas por una función de activación.

Cuando se usa la activación de funciones, pequeños cambios de los valores de entrada a veces ocasiona grandes cambios en la salida y en otras ocasiones grandes cambios en los valores de entrada tienen un impacto insignificante en las salidas. En particular, la salida es sensitiva a las entradas solo cuando la entrada está en un rango medio. Esta propiedad mejora la habilidad de la red neuronal de aprender. Varias funciones matemáticas satisfacen esta propiedad siendo las más conocidas la función sigmoid (logistic) y tanh. Hay funciones no lineales y resulta en un comportamiento no lineal. La definición de sigmoid y tanh es:

$$\text{Sigmoid: } o = 1/(1 + e^a)$$

$$\text{Tanh: } o = \frac{(e^a - e^{-a})}{(e^a + e^{-a})}$$

Donde  $a$  es el valor de entrada y  $O$  es el valor de salida.

La Red Neuronal de Microsoft usa la *Tanh* como función de activación en los nodos ocultos y la función *Sigmoid* para los nodos de salida.

### 2.4.2. Backpropagation (Retro propagación), Función de Error y Gradiente Conjugado

El entrenamiento de una red neuronal es un proceso iterativo. En cada iteración, el algoritmo compara los valores de salida con los valores actuales para obtener los errores de cada neurona de salida. Los pesos asociados a las neuronas de salidas son modificados basados en los cálculos de error. Estas modificaciones se propagan de la capa de salida hacia la capa oculta y terminando en la capa de entrada. En consecuencia, todos los pesos en la red neuronal son ajustados.

El proceso de entrenamiento de una red neuronal es descrito en los siguientes pasos:

1. El algoritmo asigna aleatoriamente valores para todos los pesos en la red en un escenario inicial (por lo regular es de -1.0 a 1.0).

2. Por cada ejemplo de entrenamiento (o cada conjunto de entrenamiento), son calculadas las salidas basados en los pesos actuales de la red.
3. Los errores de salida son calculados, y el proceso de retro propagación calcula los errores de cada salida y neurona oculta en la red. Los pesos en la red son actualizados.
4. Se repite el paso 2 hasta que la condición es satisfecha.

Algunas redes neuronales actualizan sus pesos después de examinar cada caso. Esto suele ser llamado *case[online] updating*. Otras redes neuronales actualizan sus pesos hasta que todos los casos muestra son analizados. Es llamado *epoch[batch] updating*. La interacción mediante el conjunto de datos de entrenamiento es llamado época. El algoritmo de redes neuronales de Microsoft usa la actualización de épocas porque es más robusta para los modelos de regresión.

La red neuronal necesita una medida para indicar la calidad del entrenamiento. Esta medida es el error de función. El propósito del entrenamiento de la red neuronal es la de minimizar el error de entrenamiento.

Hay diferentes opciones para la función de error, por ejemplo el, el cuadrado residual (el cuadrado de delta entre el valor predecible y el valor actual) o el umbral binario para la clasificación binaria (si la delta entre el valor de salida y el actual es menor a 0.5, entonces el error es 0; de otra manera es 1).

La siguiente formula da un método común para calcular el error para las neuronas de la capa de salida usando la derivación de la función logística (la Red Neuronal de Microsoft utiliza el error de la suma de cuadrados para atributos continuos y la entropía de cruce para los atributos discretos).

$$Err_i = O_i(1 - O_i) (T_i - O_i)$$

En este caso,  $O_i$  es la neurona de salida de la unidad  $i$ , y  $T_i$  es el valor actual para esta neurona de salida basado en la muestra de entrenamiento.

El error calculado de las neuronas ocultas está basado en los errores de las neuronas de la siguiente capa y los pesos asociados.

$$Err_i = O_i(1 - O_i) \sum_j Err_j w_{ij}$$

$O_i$  es la salida de la neurona oculta de la unidad  $i$ , el cual tiene salidas  $j$  en la siguiente capa.  $Err_j$  es el error de la neurona  $j$ ,  $w_{ij}$  es el peso entre estas dos neuronas.

Una vez que el error de cada neurona es calculado, el siguiente paso es ajustar los pesos en la red usando el siguiente método.

$$W_{ij} = W_{ij} + l * Err_j * O_i$$

Aquí  $l$  es un valor en el rango entre 0 y 1.

La variable  $l$  es llamada tasa de aprendizaje. Si el valor de  $l$  es más pequeño, los cambios de los pesos son más pequeños después de cada iteración, en consecuencia la tasa de aprendizaje es pausado. El valor de  $l$  usualmente decrementa durante el proceso de entrenamiento. En la etapa inicial del entrenamiento,  $l$  es grande, el cual permite a la red neuronal encontrar rápidamente hacia la solución óptima.

Muchas redes neuronales aplican un método llamado el gradiente conjugado en el proceso de ajustar los pesos después de cada iteración. El método de gradiente conjugado es un algoritmo para encontrar el mínimo local más cercano.

#### 2.4.3. Topología de la Red

La topología de la Red Neuronal tiene que ser fijada antes del procesamiento. El número de neuronas de entrada y de salida son fijados con el conjunto de datos de entrenamiento. Se puede configurar de muchas formas las capas ocultas, tal como el número de las capas ocultas y el número de neuronas por capa oculta.

La Red Neuronal puede tener cualquier número de capas ocultas. La capacidad de una función complicada por el número de nodos y el número capas ocultas. De tal manera que las múltiples capas ocultas podrían incrementar la capacidad de aprendizaje. Esto incrementará el tiempo de aprendizaje y también podría ocurrir un sobre entrenamiento. Con demasiadas capas ocultas y nodos ocultos, la red tiende a recordar los casos de entrenamiento en vez de generalizar los patrones. Ha sido probado que una capa oculta es más que suficiente, por lo tanto el Algoritmo de Redes Neuronales de Microsoft no permiten más de una capa oculta.

El número de neuronas que se ocupen en la capa oculta es muy importante. Cuando se tienen pocas, la red tendrá que hacer uso de sus recursos para resolver el problema, y cuando se tienen demasiadas el tiempo de procesamiento es más prolongado. Los investigadores proponen la pauta para escoger el número de neuronas ocultas:  $c * \sqrt{m * n}$  donde  $n$  es el número de neuronas de entrada,  $m$  es el número de neuronas de salida y  $c$  es una constante. En la red Neuronal de Microsoft el valor por default de  $c$  es 4.

Similar a otros algoritmos de Microsoft, un modelo de minería de datos basado en la red Neuronal de Microsoft puede tener múltiples atributos predecibles.

Cada atributo de entrada será mapeado a múltiples neuronas de entrada. En algunas ocasiones esto puede resultar en una gran cantidad de neuronas de entrada si hay muchos atributos discretos con valores distintos.

Cuando hay demasiados atributos de entrada, el algoritmo de Red Neuronal de Microsoft llama al proceso de selección de características el cual escoge los 255 atributos de entrada más importantes.

#### 2.4.4. Entrenamiento de la Red Neuronal

El proceso de entrenamiento de las redes neuronales es iterativo dependiendo en la complejidad de patrones en el conjunto de datos muestra, se puede tener cientos o miles de iteraciones en los datos. En la lista siguiente se muestra las condiciones de paro de una red neuronal:

- **Llega a la máxima precisión posible:** La tasa de errores de clasificación está por debajo de un determinado umbral.
- **Un máximo de iteración:** el proceso de entrenamiento ha encontrado el número máximo de iteraciones.
- **Convergencia de los pesos:** el cambio de los pesos después de cada iteración cae por debajo del umbral.
- **Tiempo excedido:** el número de iteraciones excede el límite.

El algoritmo de Red Neuronal de Microsoft usa las primeras tres condiciones como condición de paro. El entrenamiento se detiene cuando alguna de las tres primeras condiciones es satisfecha.

#### 2.4.5. Interpretación del Algoritmo de Red Neuronal

Después de que el modelo de Red Neuronal de Microsoft es procesado, se puede visualizar su contenido usando el visor de red neuronal. El propósito principal del visor es la de mostrar el impacto de los atributos sobre pares de valores (AV) relacionadas con el atributo de predicción.

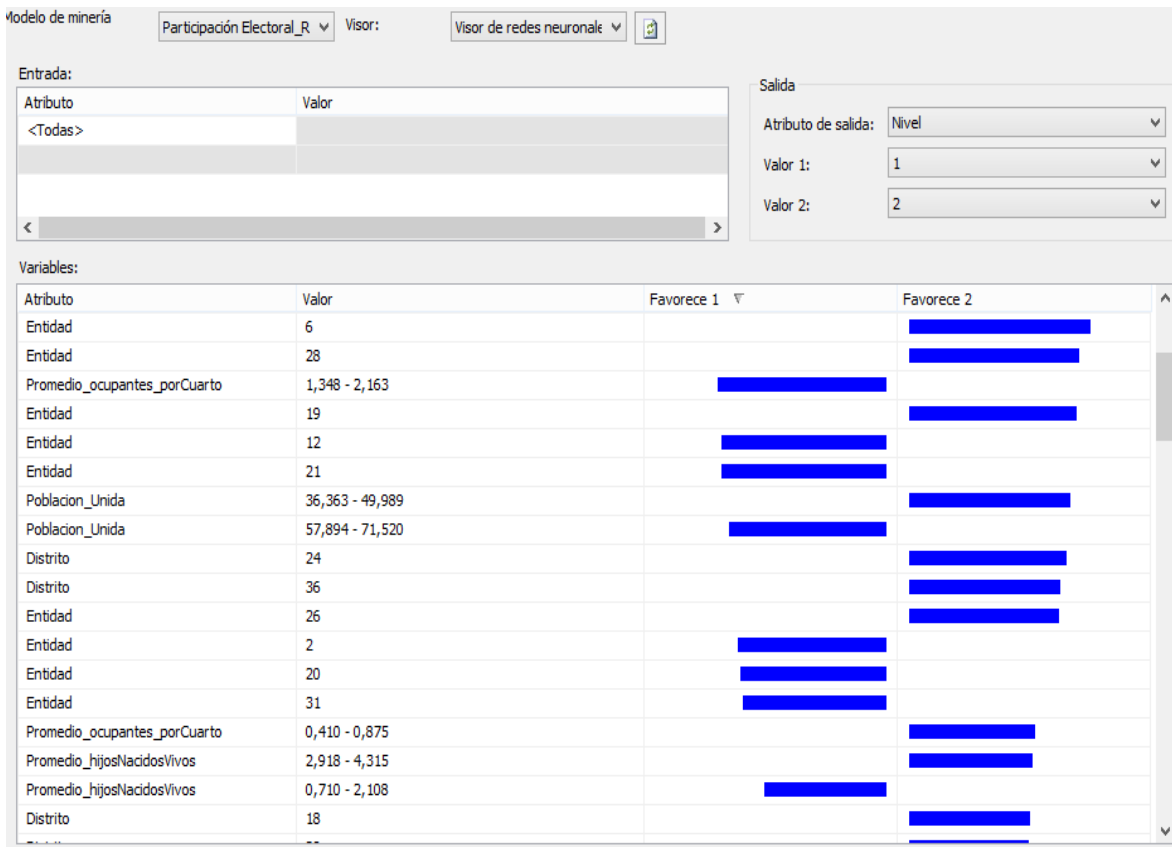


Fig. 2. Visor de contenido de Red Neuronal

En la imagen 2.2 se muestra una vista del visor de Red Neuronal. Es un visor de una sola pestaña con tres partes. En la parte superior izquierda es en donde están las opciones con las entradas, donde se puede especificar los valores de los atributos de entrada. Cuando las entradas no son especificadas el visor muestra la información de todas las entradas AVs relacionado a los estados de predicción. En la parte superior derecha es para la selección de salidas. Se puede seleccionar dos estados de un atributo de predicción. En la parte principal del visor muestra el impacto del atributo sobre los valores pares relacionados con los estados a predecir. La red neuronal puede calcular la probabilidad el estado de un atributo a predecir.

## 2.5. Modelo de análisis multidimensional.

El libro de Tang y MacLennan (2005) en el capítulo 11, que habla de los cubos de información, menciona que OLAP es el término que se le da a los sistemas que anteriormente fueron llamados sistemas de soporte de decisiones (DSS) o base de datos multidimensionales.

OLAP juega un papel muy importante hoy en día para la inteligencia de negocios (BI). Una base de datos OLAP contiene un número de cubos similar a las bases de datos relacionales que contienen un número de tablas. Un cubo contiene un conjunto de dimensiones y medidas bien definidas. Cada dimensión tiene una o más jerarquías.

OLAP y la minería de datos tienen dos tecnologías complementarias para la inteligencia de negocios. Los cubos grandes frecuentemente tienen millones de miembros en algunas dimensiones y las tablas de hechos pueden contener billones de transacciones. Existen requerimientos para aplicar las técnicas de minería de datos para encontrar patrones en estos cubos.

OLAP es usado para sistemas de toma de decisiones para analizar información en ventas, finanzas, presupuesto y muchas otras aplicaciones. El esquema de base de datos no está organizada de una manera que pueda ser fácilmente provista la información resumida requerida por los tomadores de información de las organizaciones. En resumen, OLAP se trata de medidas en jerarquías de dimensión y almacenar esta información pre calculada en una estructura de datos especial.

Existen dos esquemas para formar un cubo de información. El primer esquema de base de datos es llamado estrella. Este consiste en todas las tablas de dimensiones están conectadas a la tabla de hechos o tabla principal. El segundo esquema es el llamado de cascada, en donde la tabla de hechos es la tabla raíz y sus tablas de dimensiones reflejan su organización jerárquica.

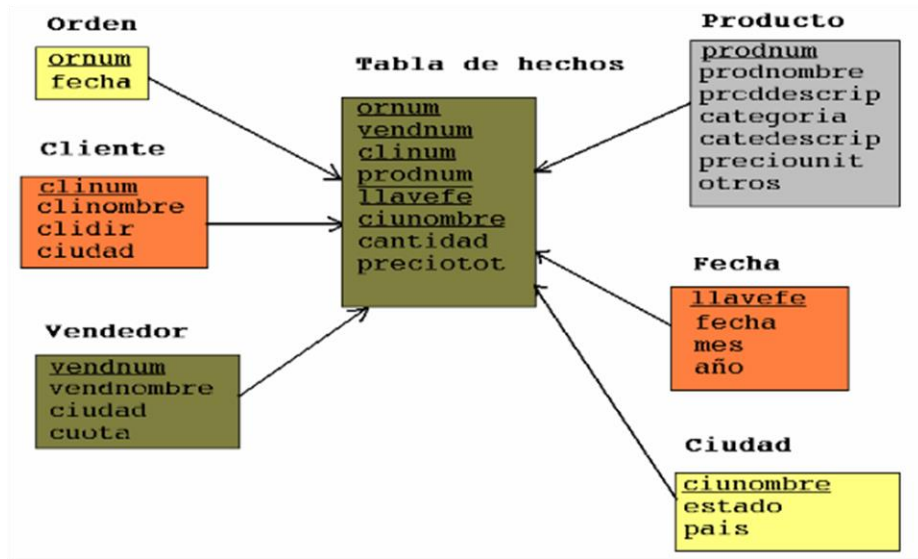


Fig. 3. Ejemplo de Modelo de Estrella.

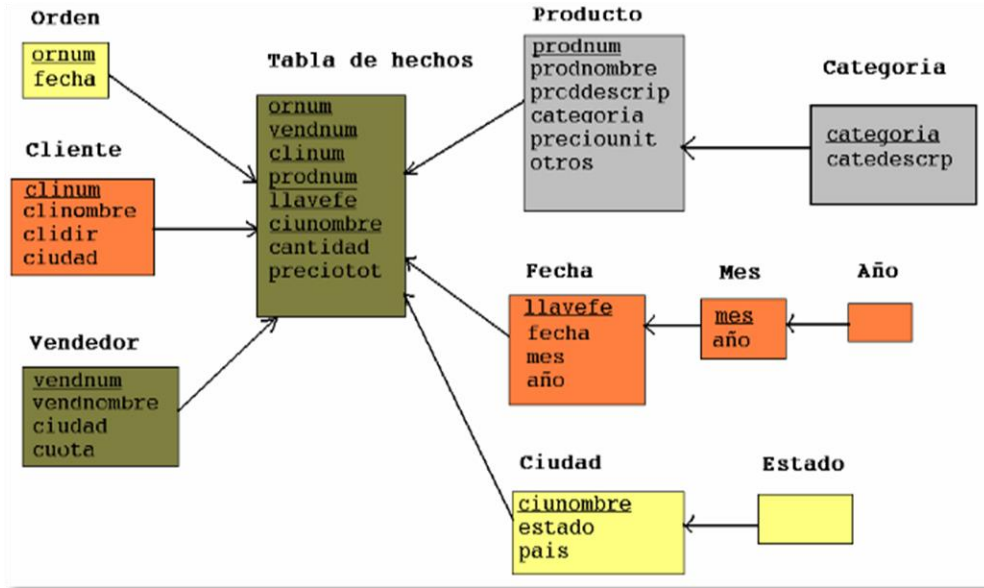


Fig. 4. Ejemplo de Modelo de Cascada.

Existen browsers para visualizar los datos del cubo, ya sea en formato numérico o en forma de gráficas, que ayudan para analizar los datos que están almacenados en el cubo.

Dimensión	Jerarquía	Operador	Expresión de filtro	Parám...
<Seleccionar dimensión>				
Indice Economicamente...	Indice Economicamente...	Indice Unida	Res Elecciones	
1	1	3	375	
1	1	4	529	
1	5	1	985	
1	5	2	3132	
1	5	3	3814	
1	5	4	9722	
1	5	5	1974	
2	2	1	609	
2	2	3	1423	
2	2	4	1128	
2	2	5	499	
2	3	1	214	
2	3	5	71	
2	4	1	2159	
2	4	2	6076	
2	4	3	48205	
2	4	4	69279	
2	4	5	32667	
2	5	1	4736	
2	5	2	42472	

Fig. 5. Visor para analizar datos del cubo

# **CAPITULO 3**

### 3. TRATAMIENTO DE LOS DATOS.

Czernicki (2011), define al proceso de selección, extracción, transformación y carga de los datos conocido como ETL por sus siglas en inglés (Extract, Transform and Load) como a continuación se presenta:

*Extracción:* En esta parte se realiza la extracción de datos brutos de una fuente de datos en específico.

*Transformación:* Esta es la parte que más tiempo lleva del proceso de ETL ya que es la parte en donde los datos se transforman en base al diseño que requiere el almacén de datos (data warehouse). El origen de los datos se transforma a un formato con el cual se nos haga fácil las operaciones de BI.

*Carga (Load):* En esta parte del proceso en la que se toman los datos ya transformados y se suben al data warehouse apropiado.

#### 3.1. Selección de datos.

##### 3.1.1. Fuente de datos INE

El propósito de este trabajo de investigación es la diseñar un modelo de inteligencia de negocios que soporte la toma de decisiones para generar estrategias de promoción de la participación ciudadana para el ejercicio del derecho al sufragio en las elecciones federales para representantes en la cámara de Diputados, Senadores y Presidente de México. De este modo la primera instancia en la que se piensa para obtener datos confiables y que tiene registro de todas las elecciones desde 1991 es el Instituto Nacional Electoral (INE).

El INE en su portal ([http://www.ine.mx/archivos3/portal/historico/contenido/Historico\\_de\\_Resultados\\_Electorales/](http://www.ine.mx/archivos3/portal/historico/contenido/Historico_de_Resultados_Electorales/)) cuenta con un histórico de los resultados electorales por año desde 1991 al 2012.

Los datos se pueden mostrar de diferentes formas ya que tiene diferentes filtros, ya que se puede elegir las tablas de datos por año electoral, tipo de elección y la agregación geográfica electoral.



Fig. 6. Sistema de Consulta de la Estadística de las Elecciones Federales.

Los datos que se puede observar que llegarán a servir para los propósitos de esta investigación están divididos en varias columnas diferentes cómo lo son el número de fila, distrito, municipio sección, el número de casillas, columnas con los diferentes partidos contendientes en el año de elección seleccionado con sus respectivos resultados, y por último cantidad de datos obtenidos en dicha zona geográfica del país.

NUM FILA	DISTRITO	MUNICIPIO	SECCION	CASILLAS	PAN	PRI	PRD	PVEM	PT	CONVERGENCIA	NUEVA ALIANZA	PSD	PRIMERO MEXICO
1	1	ASIENTOS	338	2	226	81	15	44	4	12	114	6	
2	1	ASIENTOS	339	2	260	82	10	47	12	9	119	1	
3	1	ASIENTOS	340	2	266	221	7	69	3	13	32	2	
4	1	ASIENTOS	341	3	260	195	14	53	10	21	101	4	
5	1	ASIENTOS	342	2	174	208	18	32	32	7	92	3	
6	1	ASIENTOS	343	4	557	238	25	54	22	21	173	3	
7	1	ASIENTOS	344	3	309	269	39	52	46	61	180	8	
8	1	ASIENTOS	345	3	188	266	55	50	17	11	145	5	
9	1	ASIENTOS	346	3	206	243	11	26	20	21	117	3	
10	1	ASIENTOS	347	3	421	252	21	36	18	10	166	3	
11	1	ASIENTOS	348	2	119	138	22	20	9	16	179	7	
12	1	ASIENTOS	349	3	185	240	17	22	5	8	155	1	
13	1	ASIENTOS	350	3	231	319	40	39	10	15	160	1	
14	1	ASIENTOS	351	2	189	171	66	28	24	6	113	3	
15	1	ASIENTOS	352	3	352	164	17	64	46	12	134	1	
TOTALES			589	1,320	95,746	92,583	14,476	42,273	6,428	10,799	36,148	3,394	0

Fig. 7. Datos que arroja el INE después de aplicar los filtros seleccionados.

### 3.1.2. Fuente de datos INEGI

Planteando el problema de investigación y haciendo una búsqueda exhaustiva en diferentes artículos científicos, se llegó a la conclusión de que se necesitarían variables a considerar para poder proponer el un modelo, como promedio de hijos por familia, el acceso a servicios básicos y de salud, etc.

En México la institución oficial que nos puede proporcionar esa información es el INEGI (Instituto Nacional de Estadística y Geografía).

El INEGI en colaboración con el IFE (ahora INE) realizaron la tarea de pasar toda la información del Censo de Población y Vivienda 2010 a cada las sección electoral del país

para poder realizar consultas estadísticas a nivel secciones y distritos electorales (Estadísticas Censales a Escalas Goelectorales).

El resultado del trabajo que tuvieron en conjunto las dos instituciones es una base de datos con 190 indicadores con una desagregación máxima que es por sección y corresponden a: estructura por sexo y edad, fecundidad, discapacidad, características educativas, derechohabencia a los servicios de salud, situación conyugal, religión, tipo de viviendas y acceso a los servicios básicos como también los bienes en los hogares.

Al observar la información que nos proporciona el INEGI, se puede visualizar los potenciales beneficios para el proyecto ya que se podrían cruzar con los datos que proporciona el IFE ya que también tienen una desagregación máxima que es por sección y hacer los análisis requeridos.

Los datos son proporcionados en el portal del INEGI para la consulta pública (<http://gaia.inegi.org.mx/geoelectoral/viewer.html#>).

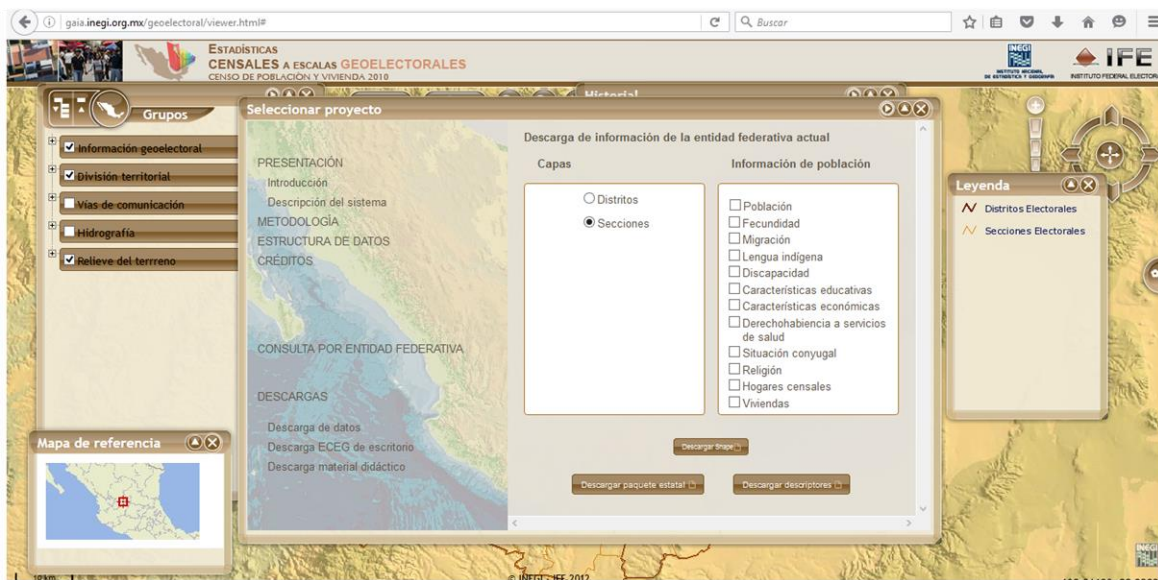


Fig. 8. Portal del INEGI donde están los datos a escalas Goelectorales.

### 3.2. Extracción de datos

Como ya se había comentado los datos que proporciona el IFE y el INEGI son públicos para su consulta.

En el caso del IFE, se cuenta con los resultados de las elecciones desde el año de 1991 al 2012 en las que se muestran ocho contiendas electorales intermedias o referente únicamente del tipo de elección de Diputados.

El primer problema que se encuentra con la información que proporciona el IFE es que se tiene que extraer una por una las tablas de los resultados electorales, seleccionando el año de

elección y a la entidad. Por lo tanto en una labor extenuante se extrajeron un total 256 tablas de datos, ocho por cada estado de la república. Las tablas se extrajeron en un formato que puede reconocer Excel.

En el portal del INEGI se encontró la información basada en el censo de población 2010, que se tiene que extraer por entidad federativa, siendo un total de 32 tablas en formato .dbf, siendo posible visualizarlas con Excel.

### 3.3. Transformación y depuración de los datos

Al tener una gran cantidad de tablas de ambas bases de datos, para un mejor control de la información lo primero que se hizo fue unir en una sola tabla mediante un query con SQL las tablas con los datos respectivos.

Las tablas que contienen los resultados de las elecciones, no se pudieron unir con facilidad, ya que cada una contiene un número diferente de columnas, siendo porque la cantidad de partidos políticos ha variado en los diferentes años electorales, haciéndolas no homogéneas.

De lo anterior, el objetivo del trabajo de tesis es la de crear un modelo de la participación ciudadana en las elecciones federales y no la de predecir los resultados de cada uno de los partidos políticos, entonces sólo se seleccionaron aplicando un query las columnas que nos servirían de cada una de las tablas como son año, entidad, el distrito, municipio, la sección y la más importante la cantidad de votos que hubo.

El resultado fue una tabla que contiene 7 columnas y 510103 registros para poder ser explotados en el modelo de BI.

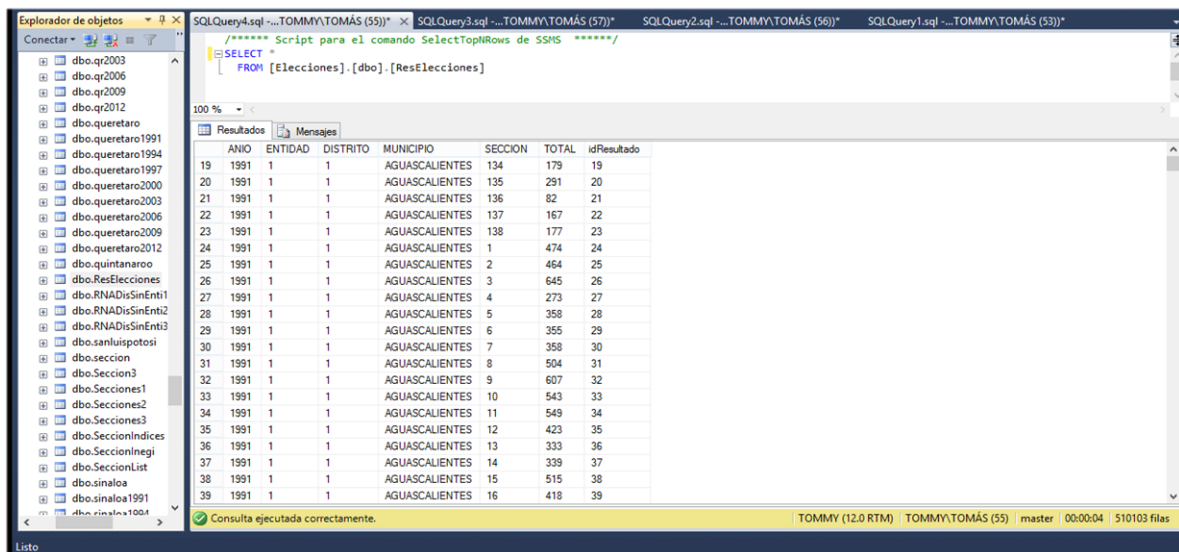


Fig. 9. Resultado de unir las 256 tablas con los resultados electorales.

Al momento de trabajar con los datos que se sacaron del INEGI, a pesar de que sólo eran 32 tablas el problema que se encontró fue que cada tabla contiene 194 columnas, pero de igual manera se aplicó un query para unir las tablas.

El resultado de lo anterior fue una tabla con 194 columnas y 66740 registros.

	CLAVEGEO	ENTIDAD	DISTRITO	POBTOT	POBMAS	POBFEM	P_0A2	P_0A2_M	P_0A2_F	P_3YMAS	P_3YMAS_M	P_3YMAS_F	P_5YMAS	P_5YMAS_M	P_5YMAS_F
1	0100100200338	1	1	2367	1156	1211	185	86	99	2180	1069	1111	2089	1038	1051
2	0100100200339	1	1	2166	1071	1095	157	78	79	2009	993	1016	1917	948	969
3	0100100200340	1	1	2307	1169	1138	179	86	93	2127	1083	1044	2013	1019	994
4	0100100200341	1	1	2621	1287	1334	216	100	116	2401	1185	1216	2298	1135	1163
5	0100100200342	1	1	2138	1060	1078	162	94	68	1972	964	1008	1879	912	967
6	0100100200343	1	1	4042	1987	2055	295	155	140	3745	1830	1915	3543	1726	1817
7	0100100200344	1	1	3482	1731	1751	256	119	137	3223	1611	1612	3057	1520	1537
8	0100100200345	1	1	2857	1383	1474	196	89	107	2661	1294	1367	2548	1239	1309
9	0100100200346	1	1	2373	1178	1195	172	86	86	2201	1092	1109	2098	1045	1053
10	0100100200347	1	1	3421	1698	1723	254	126	128	3167	1572	1595	2988	1483	1505
11	0100100200348	1	1	2110	1014	1096	161	78	83	1948	936	1012	1850	874	976
12	0100100200349	1	1	2928	1465	1463	173	96	77	2745	1363	1382	2626	1297	1329
13	0100100200350	1	1	3555	1748	1807	231	99	132	3321	1648	1673	3141	1562	1579
14	0100100200351	1	1	1966	986	980	117	49	68	1848	937	911	1751	885	866
15	0100100200352	1	1	3155	1561	1594	190	107	83	2964	1454	1510	2821	1377	1444
16	0100100200353	1	1	2101	1068	1033	144	83	61	1957	985	972	1848	930	918
17	0100100200354	1	1	1945	968	977	119	54	65	1825	913	912	1746	868	878
18	0100100300355	1	1	1568	750	818	86	36	50	1482	714	768	1412	682	730
19	0100100300356	1	1	3336	1596	1740	226	112	114	3109	1484	1625	2939	1402	1537
20	0100100300357	1	1	1617	776	841	80	41	39	1536	735	801	1475	702	773
21	0100100300358	1	1	4364	2103	2261	323	167	156	4041	1936	2105	3837	1829	2008
22	0100100300359	1	1	1196	545	651	69	29	40	1121	512	609	1070	491	579
23	0100100300360	1	1	977	458	514	54	24	30	915	433	487	870	406	464

Fig. 10. Resultado de unir las tablas del Censo de Población 2010.

Lo primero que se buscó para depurar la base de datos fue que hubiera registros que tuvieran datos NULL, pero en ninguna de las dos bases de datos se encontró. De hecho en la que contiene los resultados electorales se vio que en la forma que se le dio se podría utilizar.

Con la base de datos de INEGI no fue igual, ya que se encontraron los primeros registros que no servirían para la investigación, porque estas filas indicaban que la población total de ese sector de la población nacional es cero, algo imposible si en la base de datos que contiene los resultados electorales indica que hubo participación ciudadana. Por lo tanto se eliminaron esos casos que fueron un total de 69.

```

USE ELECCIONES
GO
SELECT *
FROM InegiTotales2010
where P_15YMAS=0
OR (P15YH_SE+P15PRI_IN+P15PRI_CO+P15SEC_IN+P15SEC_CO+P18YM_PB)=0
OR TVIVPARHAB=0
OR (VPH_S_ELEC+VPH_C_ELEC)=0
OR (VPH_AGUAFV+ VPH_AGUADV)=0
OR (VPH_PISOTI + VPH_PISODT)=0
OR (POB15_64)=0
OR (P_12YMAS)=0
OR (POBTOT)=0
OR (FROM HNVI)=0

```

	CLAVEGEO	ENTIDAD	DISTRITO	POBTOT	POBMAS	POBFEM	P_0A2	P_0A2_M	P_0A2_F	P_3YMAS	P_3YMAS_M	P_3YMAS_F	P_5YMAS	P_5YMAS_M	P_5YMAS_F
1	0200800401744	2	8	0	0	0	0	0	0	0	0	0	0	0	0
2	0500302800718	5	3	2	2	0	0	0	0	2	2	0	2	2	0
3	0500503501575	5	5	0	0	0	0	0	0	0	0	0	0	0	0
4	0700709701529	7	7	0	0	0	0	0	0	0	0	0	0	0	0
5	0800103701942	8	1	3450	1725	1725	0	0	0	0	0	0	0	0	0
6	0900200501040	9	2	0	0	0	0	0	0	0	0	0	0	0	0
7	0901001604981	9	10	280	172	108	0	0	0	280	172	108	280	172	108
8	0901001604988	9	10	0	0	0	0	0	0	0	0	0	0	0	0
9	0901001604991	9	10	0	0	0	0	0	0	0	0	0	0	0	0
10	0901001605011	9	10	62	57	5	0	0	0	62	57	5	62	57	5
11	0901001605095	9	10	0	0	0	0	0	0	0	0	0	0	0	0
12	0901001605097	9	10	0	0	0	0	0	0	0	0	0	0	0	0
13	0901101705501	9	11	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 11. Ejemplo de Filas que se eliminaron

### 3.4. Carga de datos

La carga de datos requirió un trabajo exhaustivo, ya que se habla de más de 280 tablas que se tuvieron que cargar una por una al data warehouse.

Por conveniencia de la investigación se pensó en que se necesitaría un data warehouse que tuviera las ventajas de poder modificar las tablas de la base de datos y principalmente que se pudiera aplicar inteligencia de negocios. Por ello el almacenador de datos que más se adapta a las necesidades es SQL Server.

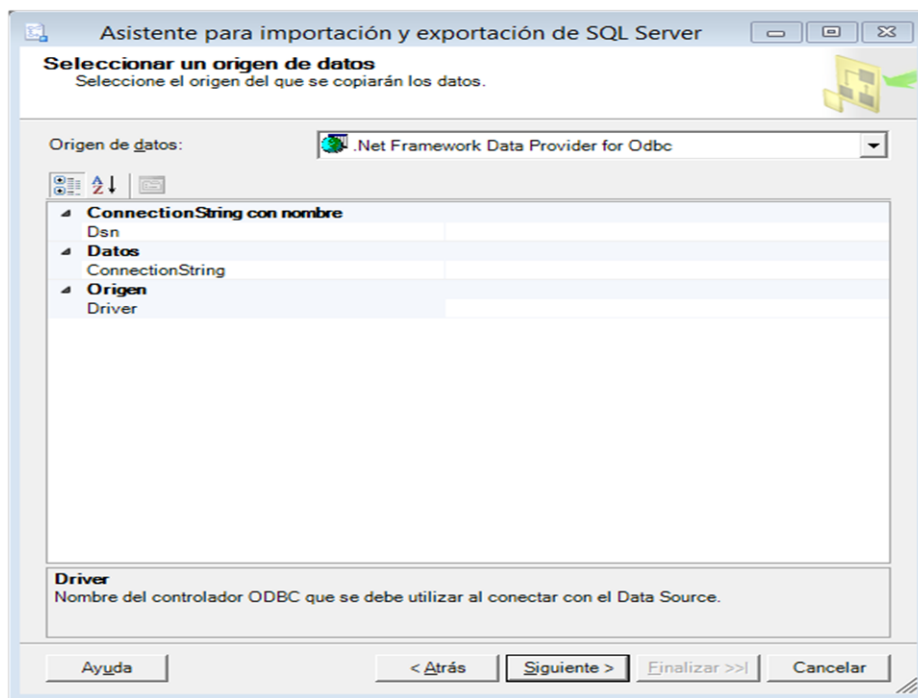


Fig. 12. Interfaz del asistente para la importación de tablas en SQL Server.

### 3.5. Indicadores clave

La base de datos del INEGI se considera que están en crudo, ya que sólo nos proporcionan totales de cada una de las características que ofrece, por ello para que se pueda encontrar mediante BI algún patrón primero se deben de tratar los datos.

Derivado de lo anterior se decidió sacar los principales índices de población que se pueden obtener con la información que proporciona la base de datos. Para obtenerlos se tomaron como base las fórmulas que se muestran en el documento llamado Panorama Sociodemográfico de México basado en el censo de población 2010 realizado por el INEGI.

Los indicadores son una guía en donde se puede visualizar los aspectos sociodemográficos más relevantes del país (Panorama Sociodemográfico de México). Estos son variables que en la literatura nos recomendó usar para el modelo de BI que se pretende proponer.

Las fórmulas para calcular los índices son las siguientes.

No	Índice	Formula	Descripción
1	Población de 15 años o más analfabeta	$(\text{Población de 15 años o más analfabeta} / \text{Total de población de 15 años o más}) * 100$	Porcentaje de población de 15 años o más analfabeta.
2	Población de 15 años o más sin primaria completa	$(\text{Población de 15 años o más sin escolaridad} + \text{Población de 15 años o más con primaria}) / (\text{población de 15 años y más con primaria completa} + \text{población de 15 años y más con secundaria incompleta} + \text{población de 15 años y más con secundaria completa} + \text{población de 18 años y más con educación post basica}) * 100$	Porcentaje de población con primaria incompleta.
3	Viviendas sin excusado	$((\text{total de viviendas habitadas} - \text{total de viviendas con excusado}) / \text{total de viviendas habitadas}) * 100$	Porcentaje de viviendas habitadas que no disponen de excusado.
4	Viviendas sin electricidad	$\text{Total de viviendas habitadas sin electricidad} / (\text{Total de viviendas habitadas con Electricidad} + \text{Total de viviendas})$	Porcentaje de viviendas habitada que no disponen de electricidad

		habitadas sin electricidad)*100	
5	Viviendas habitadas sin agua entubada	(Total de viviendas habitadas sin agua entubada / total de viviendas habitadas sin agua entubada +total de viviendas con agua entubada)*100	Porcentaje de viviendas habitadas que no disponen de agua entubada
6	Ocupantes por cuarto en viviendas habitadas	Promedio de ocupantes por cuarto en viviendas habitadas	Es porcentaje la lo proporciona por default el Censo.
7	Viviendas habitadas con piso de tierra	Viviendas habitadas con piso de tierra/(viviendas habitadas con piso de tierra + viviendas habitadas sin piso de tierra)	Porcentaje de Viviendas habitadas que tienen piso de tierra.
8	Viviendas particulares sin refrigerador	(Total de viviendas habitadas – total de viviendas habitadas con refrigerador)/ total de viviendas habitadas *100	Porcentaje de viviendas particulares habitas que no tienen refrigerador
9	Razón de dependencia por edad	(Poblacion menor a 15 años + la población mayor a 64 años)/población de 15 a 64 años	Es el cociente de la suma de la población menor de 15 años y la población mayor a 64 años entre la población de 15 a 64 años.
10	Viviendas con drenaje	(Viviendas que disponen de drenaje / total de viviendas particulares habitadas)*100	Es el número de viviendas particulares habitadas que cuentan con drenaje respecto al total de viviendas particulares habitadas expresado por cada 100
11	Viviendas con disponibilidad de teléfono	(Viviendas que disponen de teléfono/ total de viviendas habitadas)*100	Es el número de viviendas habitadas particulares que cuentan con teléfono respecto al total de viviendas

			particulares habitadas expresado por cada 100
12	Viviendas con disponibilidad de Celular	(Viviendas que disponen de celular/total de viviendas habitadas particulares)*100	Es el porcentaje de viviendas que disponen con teléfono celular.
13	Viviendas con computadora	Viviendas que disponen de computadora/ total de viviendas habitadas	Porcentaje de viviendas que cuentan con computadora.
14	Viviendas con disponibilidad de Internet	Viviendas que disponen de internet/ total de viviendas particulares habitadas	Porcentaje de viviendas particulares habitadas que cuenta con Internet
15	Promedio de hijos nacidos vivos		Se proporciona directo del censo de población y vivienda 2010
16	Población Económicamente Activa	(Población de 12 años y más económicamente activa/ población de 12 años y más)*100	Es la población que tiene trabajo o están buscando trabajo
17	Población Económicamente no Activa	(Población de 12 años y más económicamente no activa/ población de 12 años y más)*100	Es la población que no tiene trabajo
18	Población Económicamente ocupada	(Población de 12 años y más económicamente ocupada/ población de 12 años y más)*100	Población que trabaja
19	Población Económicamente no ocupada	(Población de 12 años y más económicamente no ocupada/ población de 12 años y más)*100	Es la población que no tiene trabajo pero están buscando trabajo
20	Población derechohabiente	(Población derechohabiente/población total)*100	Porcentaje de población que tiene derecho a recibir atención médica de instituciones públicas o privadas.
21	Población no derechohabiente	(Población no derechohabiente/población total)*100	Porcentaje de población que no tiene derecho a

			recibir atención médica de instituciones públicas o privadas.
22	Población soltera	(Población soltera de 12 años y más/ población total de 12 años y más)*100	Porcentaje de población que es soltera
23	Población unida	(Población unida de 12 años y más/ población total de 12 años y más)*100	Porcentaje de población que está casada o vive en unión libre
24	Población separada	(Población separada de 12 años y más/ población total de 12 años y más)*100	Porcentaje de población que es divorciada o viuda
25	Población con limitaciones	(Población que declaró tener una limitación física o mental/ Población Total)*100	Porcentaje de población que tiene alguna limitación física o mental

Después de aplicar todas estas fórmulas se consiguió crear una nueva tabla que contuviera todos los índices de cada una de las secciones electorales. De esta manera hemos transformado los datos de tal manera que se puedan usar y obtener mejores resultados en el modelo de BI.

idIndicesTotal	Entidad	Distrito	Seccion	Pob_Analfabeta_15ymas	Pob_SinPrimaria_15ymas	Viviendas_sin_excusado	Viviendas_sin_Electricidad	Viviendas_sin_agi
1	1	1	338	3,41880341880342	20,248790601244	2,4390243902439	0,37593984962406	2,819548872180
2	2	1	339	6,10361958836054	24,9637155297533	4,33884297520661	2,68595041322314	0,621118012422
3	3	1	340	6,43598615916955	27,9630943931867	5,13347022587269	2,6694045174538	1,437371663244
4	4	1	341	6,07164541590771	22,328242748092	1,49253731343284	0,747663551401869	1,492537313432
5	5	1	342	4,29403202328966	28,3232628398792	8,1858407079646	2,66666666666667	4,656319290465
6	6	1	343	4,81262327416174	21,9333874898457	5,47445255474453	2,07317073170732	1,958384332925
7	7	1	344	6,74259681093394	29,8556124825338	16,010854816825	3,15068493150685	6,675749318801
8	8	1	345	3,80604796663191	25,6989247311828	9,10518053375196	2,1978021978022	5,660377358490
9	9	1	346	5,95939751146038	28,8708586883029	12,3339658444023	1,70777988614801	12,16730038022
10	10	1	347	5,69591180523656	29,277566539924	10,02604166666667	2,21354166666667	4,836601307189
11	11	1	348	4,66812545587163	20,3957382039574	5,22088353413655	1,40845070422535	1,405622489959
12	12	1	349	5,04587155963303	21,943231441048	3,02114803625378	1,06221547799697	0,151975683890
13	13	1	350	5,8721183123097	30,964015992892	14,9606299212598	4,60526315789474	2,894736842105
14	14	1	351	6,24518118735544	30,3272146847566	7,60626398210291	2,68456375838926	2,013422818791
15	15	1	352	7,74577954319762	36,1928934010152	17,7197802197802	4,6767537826685	2,341597796143
16	16	1	353	6,0560181680545	29,2152292152292	10,062893801761	4,0169133192389	1,898734177215
17	17	1	354	3,71248025276461	23,6497545008183	8,18965517241379	3,23275862068966	3,663793103448
18	18	1	355	2,88640595903166	15,5255544840887	0,5249343832021	0,263157894736842	0
19	19	1	356	7,24106324472961	27,3111215391835	0,550206327372765	0,137741046831956	0
20	20	1	357	2,04778156996587	18,6170212765957	0,232558139534884	0	1,864801864801
21	21	1	358	6,04735883424408	26,4870931537598	1,04166666666667	0,104602510460251	0,524109014675

Consulta ejecutada correctamente. TOMMY (12.0 RTM) | TOMMY,TOMÁS (54) | master | 00:00:01 | 62881 filas

Fig. 13. Tabla con los índices socioeconómico.

Con la base de datos que tiene los resultados electorales también se tiene que calcular de participación por sección electoral y se hace de la siguiente manera.

**Índice de participación ciudadana**

$$= (\text{Total\_de\_votos} / \text{Poblacion\_Total\_18y más}) * 100$$

Para poder sacar el índice se tuvo que considerar información que nos proporciona el censo de población 2010 ya que el listado nominal sólo es proporcionado hasta el filtro por distrito.

El INE mantiene definió niveles de participación como baja, media y alta en los rangos [<60%], [60%-68%] y [>68%] respectivamente, por lo tanto se creó una columna llamada nivel de valores discretos referente a cada uno de los rangos aceptados.

	idIndicesTotal	Entidad	Distrito	Seccion	Res_Elecciones	Participacion_Electoral	Nivel
1	1	1	1	338	530	38,4057971014493	1
2	2	1	1	339	559	44,0851735015773	1
3	3	1	1	340	630	49,4893951296151	1
4	4	1	1	341	693	47,3360655737705	1
5	5	1	1	342	580	47,7759472817133	1
6	6	1	1	343	1138	49,8685363716039	1
7	7	1	1	344	993	50,7668711656442	1
8	8	1	1	345	772	44,988344988345	1
9	9	1	1	346	669	49,0469208211144	1
10	10	1	1	347	962	49,8187467633351	1
11	11	1	1	348	527	42,8107229894395	1
12	12	1	1	349	665	38,4615384615385	1
13	13	1	1	350	846	40,8695652173913	1
14	14	1	1	351	617	53,1438415159345	2
15	15	1	1	352	825	46,1409395973154	1
16	16	1	1	353	547	46,2383770076078	1
17	17	1	1	354	626	55,1056338028169	2
18	18	1	1	355	419	42,8863868986694	1
19	19	1	1	356	789	40,4407995899539	1
20	20	1	1	357	482	45,0467289719626	1
21	21	1	1	358	970	39,8520953163517	1
22	22	1	1	359	382	51,6216216216216	2
23	23	1	1	360	289	47,2995090016367	1
24	24	1	1	361	631	48,650732459522	1
25	25	1	1	362	762	47,333562224730	1

Fig. 14. Índice de Participación Ciudadana en Elecciones.

De igual manera la base de datos con los índices de población se discretizaron en cinco niveles y en algunos casos son los que contempla el INEGI para sacar el índice de marginación por localidad.

Grado de marginación	Número de localidades	Población Total	Límites del IM		Límites del IM 0-100	
			Inferior	Superior	Inferior	Superior
<b>Total</b>	<b>107 458</b>	<b>111 855 519</b>				
Muy bajo	3 480	50 760 382	[-1.83197	-1.32309]	[0.10849	4.14836]
Bajo	7 164	23 109 308	(-1.32309	-1.06870]	(4.14836	6.16786]
Medio	12 045	12 596 754	(-1.06870	-0.81425]	(6.16786	8.18785]
Alto	62 326	23 071 141	(-0.81425	0.71231]	(8.18785	20.30683]
Muy alto	22 443	2 317 934	(0.71231	8.34515]	(20.30683	80.90188]

Fig. 15. Niveles que se consideran para la discretización de los índices.

En otros casos no sirvió discretizar los datos de la forma anterior, por lo tanto mediante la media de los datos se sacaron sus cinco niveles correspondientes a cada variable. Las siguientes tablas se muestran los rangos que abarcan los niveles de cada índice.

<b>Indice_Analfabetismo</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.10849] - [4.14836]
2	Bajo	[4.14837] - [6.16786]
3	Medio	[6.16787] - [8.18785]
4	Alto	[8.18786] - [20.30683]
5	Muy Alto	[20.30684] - [100]

<b>Indice_Viviendas_Celular</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [20]
2	Bajo	[20] - [40]
3	Medio	[40] - [60]
4	Alto	[8.18786] - [20.30683]
5	Muy Alto	[20.30684] - [100]

<b>Indice_Viviendas_Drenaje</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [34]
2	Bajo	[34] - [51]
3	Medio	[51] - [68]
4	Alto	[68] - [85]
5	Muy Alto	[85] - [100]

<b>Indice_Viviendas_Internet</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.10849] - [4.14836]
2	Bajo	[4.14837] - [6.16786]
3	Medio	[6.16787] - [8.18785]
4	Alto	[8.18786] - [20.30683]
5	Muy Alto	[20.30684] - [100]

<b>Indice_Poblacion_Limitaciones</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.10849] - [4.14836]
2	Bajo	[4.14837] - [6.16786]

3	Medio	[6.16787] – [8.18785]
4	Alto	[8.18786] – [20.30683]
5	Muy Alto	[20.30684] – [100]

<b>Indice_Viviendas_PisoTierra</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.10849] - [4.14836]
2	Bajo	[4.14837] – [6.16786]
3	Medio	[6.16787] – [8.18785]
4	Alto	[8.18786] – [20.30683]
5	Muy Alto	[20.30684] – [100]

<b>Indice_Viviendas_Telefono</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [10.5]
2	Bajo	[10.5] – [21]
3	Medio	[21] – [31.5]
4	Alto	[31.5] – [42]
5	Muy Alto	[42] – [100]

<b>Indice_Poblacion_Derechohabiente</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [40]
2	Bajo	[40] – [55]
3	Medio	[55] – [70]
4	Alto	[70] – [85]
5	Muy Alto	[85] – [100]

<b>Indice_Poblacion_Economicamente_Activa</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [20]
2	Bajo	[20] – [40]
3	Medio	[40] – [50]
4	Alto	[50] – [60]
5	Muy Alto	[60] – [100]

<b>Indice_Poblacion_Economicamente_NoActiva</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [20]
2	Bajo	[20] – [40]
3	Medio	[40] – [50]
4	Alto	[50] – [60]
5	Muy Alto	[60] – [100]

<b>Indice_Poblacion_Economicamente_NoOcupada</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [1]
2	Bajo	[1] – [2]
3	Medio	[2] – [3]
4	Alto	[3] – [4]
5	Muy Alto	[4] – [100]

<b>Indice_Poblacion_Economicamente_Ocupada</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [20]
2	Bajo	[20] - [40]
3	Medio	[40] - [50]
4	Alto	[50] - [60]
5	Muy Alto	[60] - [100]

<b>Indice_Poblacion_NoDerechohabiente</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [15]
2	Bajo	[15] - [30]
3	Medio	[30] - [40]
4	Alto	[40] - [50]
5	Muy Alto	[50] - [100]
<b>Indice_Poblacion_Separa</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [4]
2	Bajo	[4] - [8]
3	Medio	[8] - [11]
4	Alto	[11] - [15]
5	Muy Alto	[15] - [100]

<b>Indice_Viviendas_SinAgua</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.10849] - [4.14836]
2	Bajo	[4.14837] - [6.16786]
3	Medio	[6.16787] - [8.18785]
4	Alto	[8.18786] - [20.30683]
5	Muy Alto	[20.30684] - [100]

<b>Indice_Viviendas_SinElectricidad</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.10849] - [4.14836]
2	Bajo	[4.14837] - [6.16786]
3	Medio	[6.16787] - [8.18785]
4	Alto	[8.18786] - [20.30683]
5	Muy Alto	[20.30684] - [100]

<b>Indice_Viviendas_SinExcusado</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.10849] - [4.14836]
2	Bajo	[4.14837] - [6.16786]
3	Medio	[6.16787] - [8.18785]
4	Alto	[8.18786] - [20.30683]
5	Muy Alto	[20.30684] - [100]

<b>Indice_Poblacion_SinPrimaria</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.10849] - [4.14836]
2	Bajo	[4.14837] - [6.16786]

3	Medio	[6.16787] – [8.18785]
4	Alto	[8.18786] – [20.30683]
5	Muy Alto	[20.30684] – [100]

<b>Indice_Viviendas_SinRefri</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.10849] - [4.14836]
2	Bajo	[4.14837] – [6.16786]
3	Medio	[6.16787] – [8.18785]
4	Alto	[8.18786] – [20.30683]
5	Muy Alto	[20.30684] – [100]

<b>Indice_Poblacion_Soltera</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [30]
2	Bajo	[30] – [35]
3	Medio	[35] – [40]
4	Alto	[40] – [45]
5	Muy Alto	[45] – [100]

<b>Indice_Viviendas_ConComputadora</b>	<b>Nivel</b>	<b>Rango</b>
1	Muy Bajo	[0.0] - [20]
2	Bajo	[20] – [40]
3	Medio	[40] – [60]
4	Alto	[60] – [80]
5	Muy Alto	[80] – [100]

<b>Indice_Promedio de hijos</b>	<b>Nivel</b>	<b>Rango</b>
1	Un 1	[0] - [1]
2	al menos 1	[1] – [2]
3	Al menos 2	[2] – [3]
4	Al menos 3	[3] – [4]
5	Al menos 4	[< 4]

### 3.6. Cruce de datos y última depuración

Desde un principio se buscó trabajar con datos que puedan identificar el mismo registro, es decir que se identifiquen las mismas zonas geográficas. En este caso las dos tablas tienen los identificadores que son las columnas Entidad, Distrito y Sección.

Lo último que se hizo para terminar con el proceso de ETL, fue cruzar los datos de ambas tablas para que tanto el índice de participación ciudadana y los índices socioeconómicos estuvieran en la misma tabla. Esto se logró haciendo uso de técnicas de análisis, explotación y actualización de datos comparando las columnas mencionadas en el párrafo anterior.

Se detectaron dos problemas que se generaron al momento de cruzar los datos por lo que se tuvieron que eliminar esos registros.

El primer problema fue que en algunos registros del campo *Participación\_Electora*, el número de votos resultantes era mayor al número total de población haciendo que el índice superara el rango de 0-100. Esa situación fue la misma en 1776 registros por lo que se descartaron.

También se encontró que a la hora de cruzar los datos la tabla con los índices de población consideraba unas secciones más que la tabla que contiene los resultados electorales, en consecuencia hubo 2014 registros que en el campo de *Participación\_Electoral* eran NULL, por lo que se eliminaron. Al final se contó con 62881 registros para realizar el análisis de datos.

El algoritmo de redes neuronales empleado en este trabajo de investigación combina cada posible estado del atributo de entrada con cada posible estado del atributo de predicción y usa los datos de entrenamiento para calcular las probabilidades. Posteriormente, puede usar estas probabilidades para la clasificación o la regresión, así como para predecir un resultado del atributo de predicción basándose en los atributos de entrada.

Los modelos construidos con este algoritmo de redes neuronales pueden contener varias redes, en función del número de columnas que se utilizan para la entrada y predicción, o sólo para la predicción. El número de redes que contiene un único modelo de minería de datos depende del número de estados que contienen las columnas de entrada y las columnas de predicción que utiliza el modelo.

Este algoritmo de redes neuronales es útil para analizar datos de entrada complejos o problemas empresariales para los que hay una cantidad importante de datos de entrenamiento disponibles pero en los que no es fácil derivar reglas mediante otros algoritmos.

El algoritmo de redes neuronales crea una red formada por has tres niveles de neuronas. Estas capas son una capa de entrada, una capa oculta opcional y una capa de salida.

**Nivel de entrada:** las neuronas de entrada definen los valores de los atributos de entrada para el modelo de minería de datos, así como sus probabilidades.

**Nivel oculto:** las neuronas ocultas reciben entradas de las neuronas de entrada y proporcionan salidas a las neuronas de salida. El nivel oculto es donde se asignan pesos a las distintas probabilidades de las entradas. Un peso describe la relevancia o importancia de una entrada determinada para la neurona oculta. Cuando mayor sea el peso asignado a una entrada, más importante será el valor de dicha entrada. Los pesos pueden ser negativos, lo que significa que la entrada puede desactivar, en lugar de activar, un resultado correcto.

**Nivel de salida:** las neuronas de salida representan valores de atributo de predicción para el modelo de minería de datos.

El modelo de redes neuronales de contener una columna clave, una o más columnas de entrada y una o más columnas de predicción.

Los modelos de minería de datos que usan el algoritmo de redes neuronales están muy influenciados por los valores que se especifican en los parámetros disponibles para el algoritmo. Los parámetros definen como se muestrean los datos, como se distribuyen o como se espera que estén distribuidos en cada columna, y cuando se invoca la selección de características para limitar los valores usados en el modelo final.

# **CAPITULO 4**

## 4. MULTIDIMENSIONALIDAD: INTELIGENCIA DE NEGOCIOS.

### 4.1. Modelo multidimensional en el dominio de aplicación.

Los datos hasta el momento está en una forma plana, pero para visualizar la información de una forma más dinámica y aplicar Inteligencia de negocios en un modelo multidimensional pasamos los datos a un cubo de datos, siendo cada índice una dimensión del cubo.

Con los datos sobre un modelo relacional, y de acuerdo al objetivo de la investigación se realizó el análisis y diseño de las estructuras de datos para poder pasarlos a un modelo multidimensional el cual ofrece mayores ventajas para la operación en tiempo real.

Se analizó la viabilidad de trabajar el modelo de copo de nieve así como el modelo de estrella. Se determinó que la mejor estructura era estrella como se muestra en la figura 4.1 en la que se consideraron los índices totales como la tabla de hechos y los índices como las dimensiones.

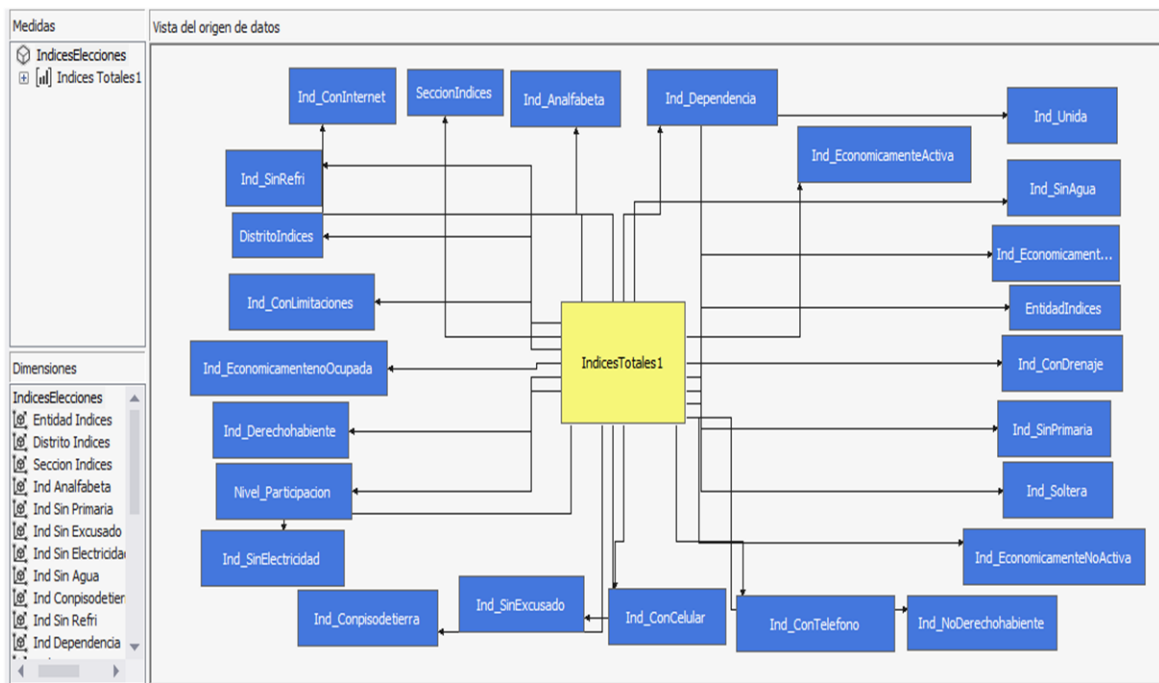


Fig. 16. Estructura del Cubo de información.

Ya que se tiene el cubo, se puede hacer un análisis de los datos que contiene. El visor de datos que maneja SQL Server 2014 facilita dicha tarea.

El objetivo del análisis de los datos mediante el cubo es la de encontrar la combinación de índices en los cuales se pueden encontrar la mayor cantidad de votos en la base de datos.

El visor funciona seleccionando una columna que contenga datos con totales, de ahí se selecciona una dimensión y automáticamente se muestra la suma total de los datos que están distribuidos en los niveles o rangos de dicha dimensión como se muestra en la figura 17.

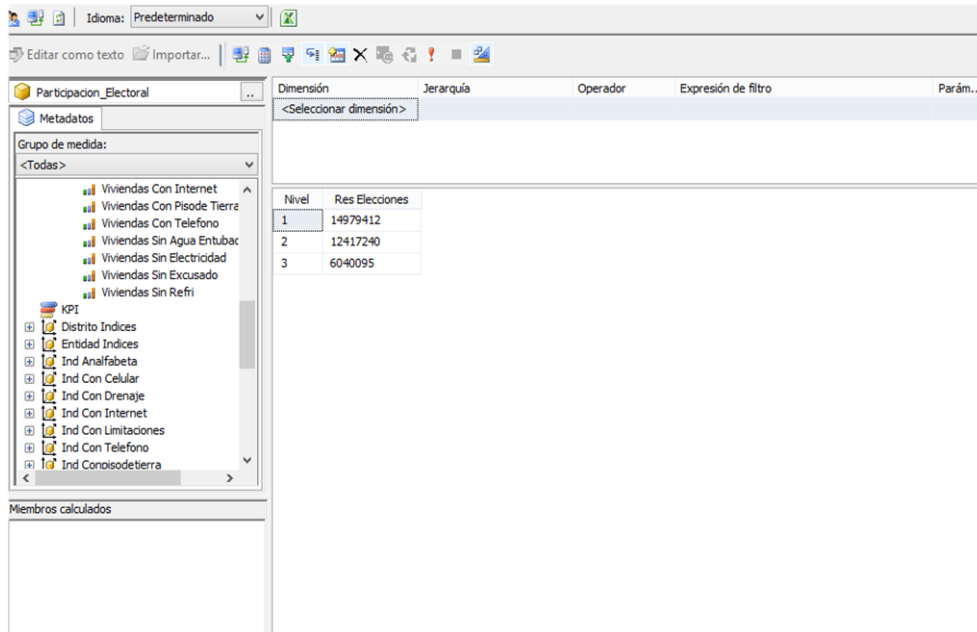


Fig. 17. Ejemplo de Distribución de Totales

En la tabla de hechos del cubo la columna que tiene totales que podemos medir es [Res\_Elecciones], ya que es la que contiene el total de votos que hubo en las diferentes secciones que hay en la República mexicana y está registrada en la base de datos.

Primero se selecciona la medida que como se mencionó en el párrafo sería la columna [Res\_Elecciones], la cual muestra la suma de todos los votos que hubo en las secciones con un total de 33436747 en las elecciones federales del 2009.

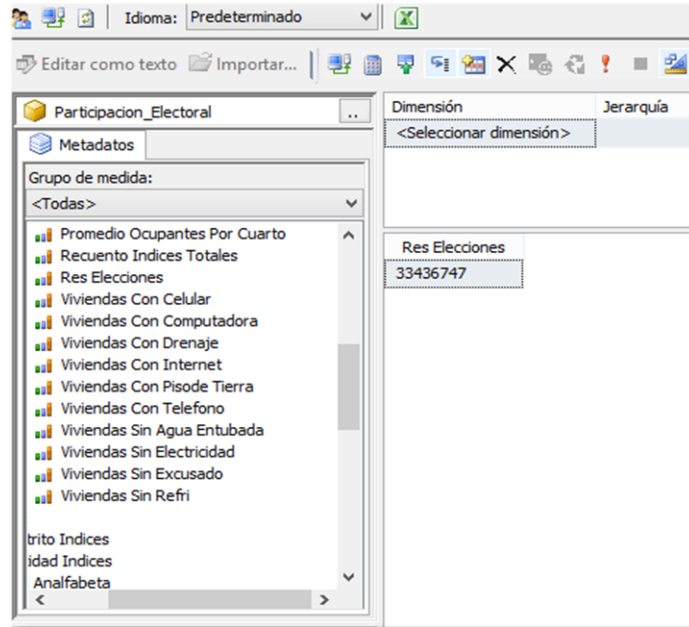


Fig. 18. Medida [Res\_Elecciones] .

Ya teniendo claro que es lo que se va a medir, lo que procede es agregar las dimensiones para ver cómo está distribuido el total de votos. De cierta manera esta forma de hacer el análisis es manual ya que se tiene que buscar que combinación de índices es la que es más frecuente.

Se cuenta con un total de 25 índices de los cuales en este análisis se tiene que encontrar las combinaciones de niveles en los cuales se encuentra el mayor número de votos. El problema es que al ser tantos índices y cada uno con sus propios niveles el análisis se hace muy complicado por la gran cantidad de combinaciones que se pueden generar. Por lo tanto sólo se utilizarán los índices del modelo final que se generó el cual se explica en el siguiente capítulo.

Indice Economicamente A...	Indice Economicamente ...	Indice Viv Con Telefono	Indice Viv Con Celular	Indice Promediodehijos	Indice No Derecho Habiente	Indice Unida	Res Elecciones
1	1	1	2	3	1	5	82
1	1	3	4	3	1	3	232
1	1	5	4	2	1	3	375
1	1	5	4	2	1	4	647
1	1	5	4	3	1	4	529
1	1	5	5	2	1	3	595
1	2	5	4	2	1	4	700
1	5	1	1	3	1	3	579
1	5	1	1	3	1	4	90
1	5	1	1	3	1	5	469
1	5	1	1	3	2	3	304
1	5	1	1	3	2	4	1066
1	5	1	1	3	3	1	108
1	5	1	1	3	3	4	927
1	5	1	1	3	4	2	494
1	5	1	1	3	4	3	777
1	5	1	1	3	5	2	154
1	5	1	1	3	5	3	515
1	5	1	1	3	5	4	330
1	5	1	1	4	1	1	371

Fig. 19. Combinación de los 7 índices.

Aun con sólo 7 índices existen muchas combinaciones posibles, pero aun así es posible hacer el análisis.

Individualmente cada índice presenta el nivel en donde existe la mayor cantidad de votos, como se muestra en la siguiente tabla.

Índice	Nivel	Votos
Indice_Economicamente_Activa	4	20,362,938
Indice_Economicamente_Ocupada	3	20,748,397
Indice_No_Derechohabiente	2	13,769,919
Indice_Unida	4	11,454,121
Indice_ViviendasconTelefono	5	18,609,235
Indice_ViviendasconCelular	4	15,951,671
Indice_PromediodeHijos	3	21,380,643

La combinación que muestra una mayor cantidad de votos a comparación de otras combinaciones es la que se muestra a continuación.

Índice	Nivel
Indice_Economicamente_Activa	4
Indice_Economicamente_Ocupada	3
Indice_ViviendasconTelefono	5
Indice_ViviendasconCelular	4
Indice_PromediodeHijos	3
Indice_No_Derechohabiente	3
Indice_Unida	3
Res_Elecciones	940538

Como se puede ver en la tabla anterior la mayor cantidad de votos que se obtuvo de la combinación de los niveles de los índices es de 940538. El análisis anterior es para dar información a los tomadores de decisiones, entonces hay que explicar que es lo que quiere decir la tabla.

La conclusión que se le puede dar a los tomadores, es que hay más votos en las elecciones cuando la población cae en los rangos de las variables que se considera en el modelo.

<b>Índice</b>	<b>Nivel</b>	<b>Rango</b>
Indice_Economicamente_Activa	Alto	50-60 %
Indice_Economicamente_Ocupada	Medio	40-50%
Indice_ViviendasconTelefono	Muy Alto	42-100%
Indice_ViviendasconCelular	Alto	60-80%
Indice_PromediodeHijos	De 2 a 3 hijos	2-3
Indice_No_Derechohabiente	Medio	30-40 %
Indice_Unida	Medio	50-55 %

La tabla anterior quiere decir que cuando entre el 50 y 60 por ciento de la población es económicamente activa, entre 40 y 50 por ciento es económicamente ocupada, existen familias que en promedio tienen de 2 y 3 hijos, entre el 50 y 55 por ciento de la población vive en pareja, de 30 a 40 por ciento de población no tiene acceso a los servicios de salud públicos, el porcentaje de viviendas que cuentan con teléfono fijo es mayor al 42 por ciento y haya entre 60 y 80 por ciento de viviendas que cuenten con celular se puede esperar que la votación sea alta.

# **CAPITULO 5**

## 5. INTELIGENCIA DE NEGOCIOS

### 5.1. Análisis de la Estructura de minería de datos

Al haber completado la exhaustiva tarea que es el ETL teniendo la seguridad de que ya no existen campos vacíos o null se comenzó con la aplicación de técnicas para la inteligencia de negocios. La herramienta que se utilizó para hacer el análisis fue SQL Server y el lenguaje DMX.

Hay que realizar un proceso de cuatro pasos para poder analizar los datos que son: crear la estructura de minería de datos, agregar el algoritmo de minería de datos, procesar la estructura de minería de datos. Esto se realiza mediante código DMX.

La estructura de minería de datos requiere de un origen de datos, que es la base de datos que contiene las tablas en el Sql Server. También solicita las vistas de origen de datos, que son las tablas que utilizaremos para crear el cubo y de las que hará uso la estructura de minería de datos.

Al asignar la base de datos y las tablas con las que se trabajará la estructura de minería de datos, lo que sigue es realizar el proceso para el análisis de los datos.

#### 5.1.1. Crear Estructura de minería de datos

Para crear la estructura de base de datos se usa `CREATE MINING STRUCTURE` en lenguaje DMX, con lo que también se define el nombre de la estructura, la columna llave, las columnas que serán para la minería de datos y por ultimo un conjunto de datos de prueba.

```

CREATE MINING STRUCTURE [Participacion_Electoral_24_continuo_001] //Nombre de la
                                                                    // Estructura de
                                                                    // Minería de Datos

(
  [idIndicesTotal] long key // Columna llave para la Estructura de minería

  ,[Entidad] long discrete
  ,[Distrito] long discrete
  ,[Seccion] long discrete
  ,[Pob_Analfabeta_15ymas] double continuous //Columnas para la minería
  ,[Pob_SinPrimaria_15ymas]double continuous
  ,[Viviendas_sin_excusado]double continuous
  ,[Viviendas_sin_Electricidad]double continuous
  ,[Viviendas_sin_agua_entubada]double continuous
  ,[Promedio_ocupantes_porCuarto]double continuous
  ,[Viviendas_conPisodeTierra]double continuous
  ,[Viviendas_sin_Refri]double continuous
  ,[Viviendas_conDrenaje]double continuous
  ,[Viviendas_conTelefono]double continuous
  ,[Viviendas_conCelular]double continuous
  ,[Viviendas_con_Computadora]double continuous
  ,[Viviendas_con_internet]double continuous
  ,[Promedio_hijosNacidosVivos]double continuous
  ,[Poblacion_Economicamente_Activa]double continuous
  ,[Poblacion_Economicamente_NoActiva]double continuous
  ,[Poblacion_Econimicamente_Ocupada]double continuous
  ,[Poblacion_Econimicamente_No_Ocupada]double continuous
  ,[Poblacion_Derechohabiente]double continuous
  ,[Poblacion_NoDerechohabiente]double continuous
  ,[Poblacion_Soltera]double continuous
  ,[Poblacion_unida]double continuous
  ,[Poblacion_Separada]double continuous
  ,[Poblacion_con_Limitaciones]double continuous
  ,[Nivel] long discrete
)

```

Fig. 20. Crear Estructura de minería de datos.

### 5.1.2. Algoritmo de minería de datos

En esta parte se agrega el modelo de minería que permitirá analizar los datos y poder crear predicciones.

Para agregar el algoritmo de minería de datos se usa ALTER MINING STRUCTURE (DMX). El código puede ser dividido en las siguientes partes:

- Identificar la estructura de minería de datos
- Nombrando del modelo de minería
- Definición de la columna llave
- Definición de las columnas de entrada y la de predicción
- Identificación del algoritmo y selección de sus parámetros.

```

ALTER MINING STRUCTURE [Participacion_Electoral_24_continuo_001]//Identifica la
Estructura de minería
ADD MINING MODEL [Participación Electoral_RNA_024]//Nombre del modelo de minería
(
  [idIndicesTotal] //Definición de la columna llave
  ///Definición de las columnas que se usaran para la minería
    ,[Entidad]
    ,[Distrito]
    ,[Seccion]
    ,[Pob_Analfabeta_15ymas]
    ,[Pob_SinPrimaria_15ymas]
    ,[Viviendas_sin_excusado]
    ,[Viviendas_sin_Electricidad]
    ,[Viviendas_sin_agua_entubada]
    ,[Promedio_ocupantes_porCuarto]
    ,[Viviendas_conPisodeTierra]
    ,[Viviendas_sin_Refri]
    ,[Viviendas_conDrenaje]
    ,[Viviendas_conTelefono]
    ,[Viviendas_conCelular]
    ,[Viviendas_con_Computadora]
    ,[Viviendas_con_internet]
    ,[Promedio_hijosNacidosVivos]
    ,[Poblacion_Economicamente_Activa]
    ,[Poblacion_Economicamente_NoActiva]
    ,[Poblacion_Econimicamente_Ocupada]
    ,[Poblacion_Econimicamente_No_Ocupada]
    ,[Poblacion_Derechohabiente]
    ,[Poblacion_NoDerechohabiente]
    ,[Poblacion_Soltera]
    ,[Poblacion_unida]
    ,[Poblacion_Separada]
    ,[Poblacion_con_Limitaciones]
    ,[Nivel] Predict //Definición de la columna que será usada para la

                prediccion
)
///Definición del algortimo y los atributos que contendrá
using Microsoft_Neural_Network ( HOLDOUT_PERCENTAGE=50,
HOLDOUT_SEED=0,MAXIMUM_INPUT_ATTRIBUTES=255,MAXIMUM_OUTPUT_ATTRIBUTES=255,
MAXIMUM_STATES=200,SAMPLE_SIZE=6000,HIDDEN_NODE_RATIO=10)

```

Fig. 21. Modelo de minería de datos y el algoritmo a emplear.

### 5.1.3. Procesar Estructura de minería de datos

Para procesar la estructura de minería y el modelo de minería de datos se utiliza INSERT INTO MINING STRUCTURE en DMX.

Al momento de procesar la estructura de minería, se lee la fuente de los datos y construye las estructuras que soporta los modelos de minería. Los datos definidos por la estructura de minería son pasados por el algoritmo de minería de datos que se eligió. El algoritmo busca

patrones y tendencias los cuales son almacenados en el modelo de minería. En este momento es cuando el modelo de minería contiene la información descubierta por el algoritmo.

El código puede ser dividido en tres partes:

- Identificar la estructura de minería
- Listar las columnas de la estructura de minería
- Definir los datos de entrenamiento.

```
INSERT INTO MINING STRUCTURE [Participacion_Electoral_24_continuo_001] //Nombre de
la estructura de Minería
(
  [idIndicesTotal]
  ,[Entidad]
  ,[Distrito]
  ,[Seccion]
  ,[Pob_Analfabeta_15ymas]
  ,[Pob_SinPrimaria_15ymas]
  ,[Viviendas_sin_excusado]
  ,[Viviendas_sin_Electricidad]
  ,[Viviendas_sin_agua_entubada]
  ,[Promedio_ocupantes_porCuarto]
  ,[Viviendas_conPisodeTierra]
  ,[Viviendas_sin_Refri] //Columnas de la estructura de minería
  ,[Viviendas_conDrenaje]
  ,[Viviendas_conTelefono]
  ,[Viviendas_conCelular]
  ,[Viviendas_con_Computadora]
  ,[Viviendas_con_internet]
  ,[Promedio_hijosNacidosVivos]
  ,[Poblacion_Economicamente_Activa]
  ,[Poblacion_Economicamente_NoActiva]
  ,[Poblacion_Econimicamente_Ocupada]
  ,[Poblacion_Econimicamente_No_Ocupada]
  ,[Poblacion_Derechohabiente]
  ,[Poblacion_NoDerechohabiente]
  ,[Poblacion_Soltera]
  ,[Poblacion_unida]
  ,[Poblacion_Separada]
  ,[Poblacion_con_Limitaciones]
  ,[Nivel] )

  //Definicion de los datos de entrenamiento
openquery ([ParticipacionElectoral],
'SELECT [idIndicesTotal]
,[Entidad]
,[Distrito]
,[Seccion]
,[Pob_Analfabeta_15ymas]
,[Pob_SinPrimaria_15ymas]
,[Viviendas_sin_excusado]
,[Viviendas_sin_Electricidad]
,[Viviendas_sin_agua_entubada]
,[Promedio_ocupantes_porCuarto]
,[Viviendas_conPisodeTierra]
,[Viviendas_sin_Refri]
```

```

,[Viviendas_conDrenaje]
,[Viviendas_conTelefono]
,[Viviendas_conCelular]
,[Viviendas_con_Computadora]
,[Viviendas_con_internet]
,[Promedio_hijosNacidosVivos]
,[Poblacion_Economicamente_Activa]
,[Poblacion_Economicamente_NoActiva]
,[Poblacion_Econimicamente_Ocupada]
,[Poblacion_Econimicamente_No_Ocupada]
,[Poblacion_Derechohabiente]
,[Poblacion_NoDerechohabiente]
,[Poblacion_Soltera]
,[Poblacion_unida]
,[Poblacion_Separada]
,[Poblacion_con_Limitaciones]
,[Nivel]
from dbo.IndicesTotales')

```

Fig. 22. Procesamiento de la estructura de minería de datos

#### 5.1.4. Predicción de la Estructura de minería de datos

Para realizar las predicciones con DMX se utiliza `SELECT FROM <MODEL> PREDICTION`, esta forma contiene tres formas.

- Una lista de las columnas de minería de datos y las funciones de predicción que son regresados en los resultados. Los resultados también pueden contener las columnas de entrada de los datos fuente.
- El query fuente que definen los datos que están siendo usadas para crear predicciones.
- Un mapeo entre las columnas del modelo de minería y los datos fuente.

```

SELECT
  [Participación Electoral_RNA_024].[Nivel],
  t.[Nivel],
  t.[idIndicesTotal],    //Columnas que regresaran los resultados de la predicción
  t.[Entidad],
  t.[Distrito],
  t.[Seccion]
From
  [Participación Electoral_RNA_024]

//////////Datos que están siendo usados para realizar la predicción
PREDICTION JOIN
  OPENQUERY([ParticipacionElectoral],
    'SELECT
      [Nivel],
      [idIndicesTotal],
      [Entidad],
      [Distrito],
      [Seccion],
      [Pob_Analfabeta_15ymas],
      [Pob_SinPrimaria_15ymas],
      [Viviendas_sin_excusado],

```

```

[Viviendas_sin_Electricidad],
[Viviendas_sin_agua_entubada],
[Promedio_ocupantes_porCuarto],
[Viviendas_conPisodeTierra],
[Viviendas_sin_Refri],
[Viviendas_conDrenaje],
[Viviendas_conTelefono],
[Viviendas_conCelular],
[Viviendas_con_Computadora],
[Viviendas_con_internet],
[Promedio_hijosNacidosVivos],
[Poblacion_Economicamente_Activa],
[Poblacion_Economicamente_NoActiva],
[Poblacion_Economicamente_Ocupada],
[Poblacion_Economicamente_No_Ocupada],
[Poblacion_Derechohabiente],
[Poblacion_NoDerechohabiente],
[Poblacion_Soltera],
[Poblacion_unida],
[Poblacion_Separada],
[Poblacion_con_Limitaciones]
FROM
[dbo].[IndicesTotales1]
') AS t
//Mapeo entre las columnas del modelo de minería y los datos fuente
ON
[Participación Electoral_RNA_024].[Entidad] = t.[Entidad] AND
[Participación Electoral_RNA_024].[Distrito] = t.[Distrito] AND
[Participación Electoral_RNA_024].[Seccion] = t.[Seccion] AND
[Participación Electoral_RNA_024].[Pob_Analfabeta_15ymas] = t.[Pob_Analfabeta_15ymas]
AND
[Participación Electoral_RNA_024].[Pob_SinPrimaria_15ymas] =
t.[Pob_SinPrimaria_15ymas] AND
[Participación Electoral_RNA_024].[Viviendas_sin_excusado] =
t.[Viviendas_sin_excusado] AND
[Participación Electoral_RNA_024].[Viviendas_sin_Electricidad] =
t.[Viviendas_sin_Electricidad] AND
[Participación Electoral_RNA_024].[Viviendas_sin_agua_entubada] =
t.[Viviendas_sin_agua_entubada] AND
[Participación Electoral_RNA_024].[Promedio_ocupantes_porCuarto] =
t.[Promedio_ocupantes_porCuarto] AND
[Participación Electoral_RNA_024].[Viviendas_conPisodeTierra] =
t.[Viviendas_conPisodeTierra] AND
[Participación Electoral_RNA_024].[Viviendas_sin_Refri] = t.[Viviendas_sin_Refri] AND
[Participación Electoral_RNA_024].[Viviendas_conDrenaje] = t.[Viviendas_conDrenaje]
AND
[Participación Electoral_RNA_024].[Viviendas_conTelefono] = t.[Viviendas_conTelefono]
AND
[Participación Electoral_RNA_024].[Viviendas_conCelular] = t.[Viviendas_conCelular]
AND
[Participación Electoral_RNA_024].[Viviendas_con_Computadora] =
t.[Viviendas_con_Computadora] AND
[Participación Electoral_RNA_024].[Viviendas_con_internet] =
t.[Viviendas_con_internet] AND
[Participación Electoral_RNA_024].[Promedio_hijosNacidosVivos] =
t.[Promedio_hijosNacidosVivos] AND
[Participación Electoral_RNA_024].[Poblacion_Economicamente_Activa] =
t.[Poblacion_Economicamente_Activa] AND
[Participación Electoral_RNA_024].[Poblacion_Economicamente_NoActiva] =
t.[Poblacion_Economicamente_NoActiva] AND

```

```

[Participación Electoral_RNA_024].[Poblacion_Econimicamente_Ocupada] =
t.[Poblacion_Econimicamente_Ocupada] AND
[Participación Electoral_RNA_024].[Poblacion_Econimicamente_No_Ocupada] =
t.[Poblacion_Econimicamente_No_Ocupada] AND
[Participación Electoral_RNA_024].[Poblacion_Derechohabiente] =
t.[Poblacion_Derechohabiente] AND
[Participación Electoral_RNA_024].[Poblacion_NoDerechohabiente] =
t.[Poblacion_NoDerechohabiente] AND
[Participación Electoral_RNA_024].[Poblacion_Soltera] = t.[Poblacion_Soltera] AND
[Participación Electoral_RNA_024].[Poblacion_unida] = t.[Poblacion_unida] AND
[Participación Electoral_RNA_024].[Poblacion_Separada] = t.[Poblacion_Separada] AND
[Participación Electoral_RNA_024].[Poblacion_con_Limitaciones] =
t.[Poblacion_con_Limitaciones] AND
[Participación Electoral_RNA_024].[Nivel] = t.[Nivel]

```

Fig. 23. Predicción de la Participación Ciudadana en Elecciones.

## 5.2. Análisis de datos

Para probar la estructura de minería de datos, la tabla en la que tenemos todos los índices se dividen en dos. Una que será la tabla con datos de prueba y entrenamiento, y una segunda tabla que será de predicción.

El objetivo del proyecto de investigación es la de proponer un modelo de inteligencia de negocios para la participación ciudadana en las elecciones federales, por lo tanto se consideraron 25 índices o variables de las cuales tendremos que determinar cuáles son las que influyen para que una sección de determinada zona geográfica sea más participativa en las elecciones. Por lo tanto el primer análisis se tiene que hacer introduciendo todas las variables para predecir el nivel de participación.

Se cuentan con dos tipos de datos que acepta el algoritmo de red neuronal, que son los valores discretos y los continuos. De esta manera se tienen las dos primeras pruebas, considerando las 25 variables con los distintos tipos de datos.

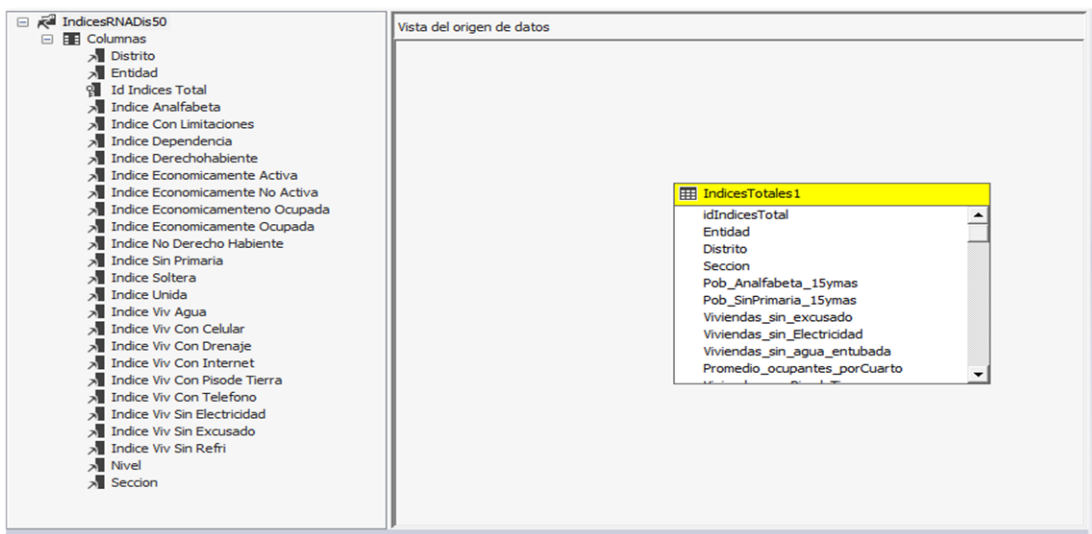


Fig. 24. Estructura de minería de datos con datos Discretos.

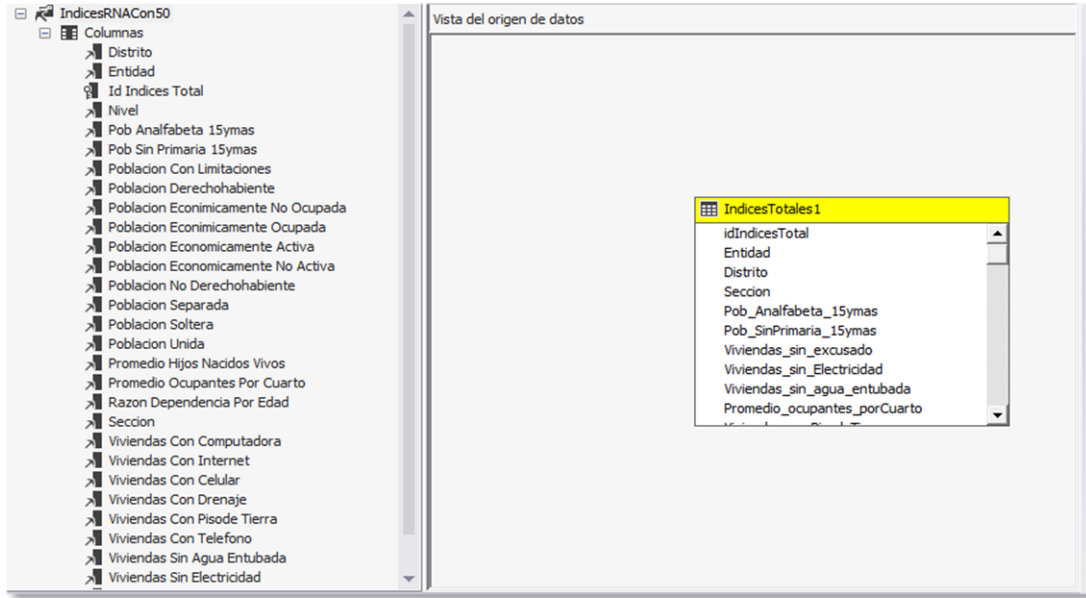


Fig. 25. Estructura de minería de datos con datos Continuos.

A parte de las variables que contiene los índices se incluyeron en este modelo las variables de entidad, distrito y sección.

Cuando se generó la estructura de minería de datos se eligió el algoritmo de red neuronal con un porcentaje de población para aprendizaje de 30%.

Se comparó el algoritmo de red neuronal con el de árboles de decisión para ver cual ofrece una mejor probabilidad de predicción.

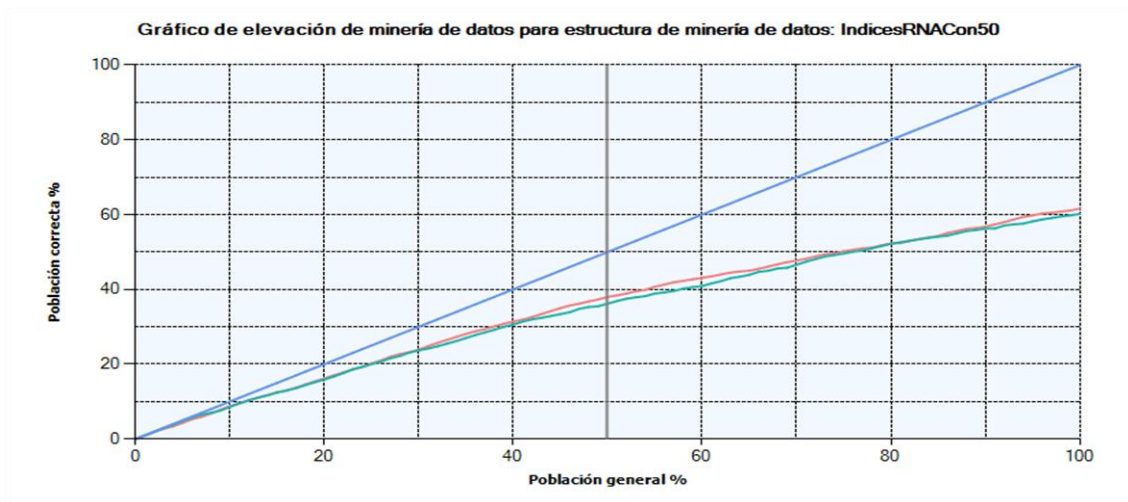


Fig. 26. Gráfico de elevación con modelo de datos continuos.

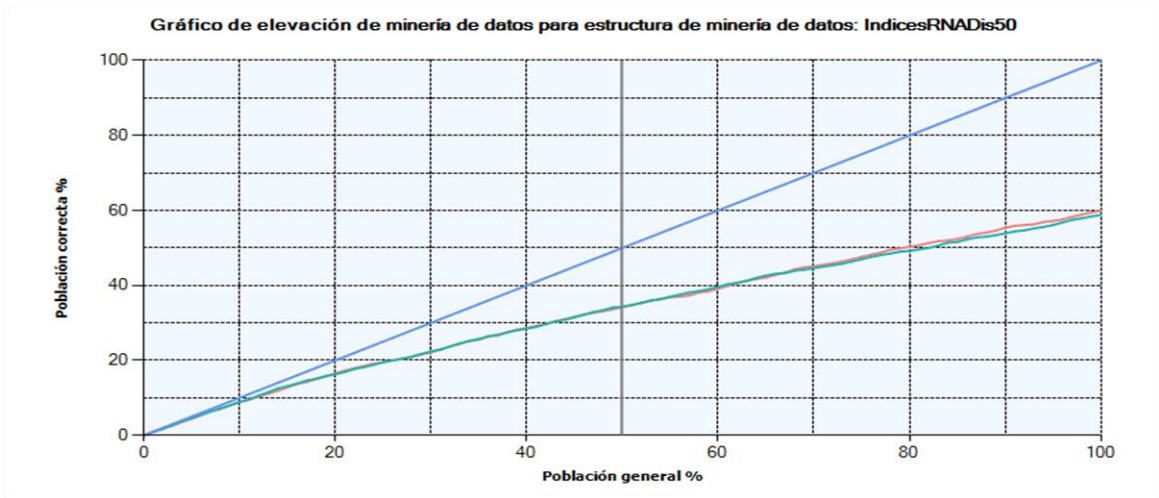


Fig. 27. Gráfico de elevación con modelo de datos continuos.

En las gráficas anteriores la línea azul es la proyección ideal de modelo, la curva rosa es el modelo de red neuronal y la verde de árboles de decisión. Lo que nos dice cada curva es que el primer algoritmo se acerca más a la proyección ideal por lo que tendrá una probabilidad mayor de predicción.

### 5.3. Resultados de la predicción

Después de haber hecho lo expuesto en los capítulos anteriores y haber creado las estructuras de minería de datos con DMX se realiza la predicción teniendo como resultado una tabla con 6 columnas: Predicción, Nivel, Id, Entidad, Distrito y Sección. El primer campo representa el nivel de participación electoral que el algoritmo predice se obtendrá según las variables de ese registro. El segundo campo representa el nivel participación real de dicho registro, los demás campos representa a que zona geográfica se hizo la predicción.

Predicción	Nivel	idIndicesTotal	Entidad	Distrito	Seccion
3	3	49800	24	6	1024
2	2	49804	24	6	1028
3	2	49808	24	6	1032
3	3	49812	24	6	1036
2	1	49816	24	6	1045
2	2	49820	24	6	1049
2	2	49824	24	6	1053
2	1	49828	24	6	1057
2	1	49832	24	6	1061
2	1	49836	24	6	1065
2	1	49840	24	6	1069
3	1	49844	24	6	1073
3	1	49848	24	6	1108
1	2	49852	24	6	1112
2	1	49856	24	6	1125
3	3	49860	24	7	35
2	3	49864	24	7	39
3	3	49868	24	7	43
3	3	49872	24	7	47
3	2	49876	24	7	51
3	2	49880	24	7	55
3	3	49884	24	7	59

Ejecución de consulta finalizada con 31443 filas recuperadas

Fig. 28. Tabla con la predicción del algoritmo de red neuronal.

Con la tabla que resultó se verifica la cantidad de registros en los cuales el algoritmo predijo correctamente el nivel de participación ciudadana comparando el campo Predicción con el campo Nivel. Los resultados se muestran en la siguiente tabla.

Algoritmo	Características	Porcentaje de clasificación
Red Neuronal	Variables de tipo continuo con variable de predicción discreto	$(18788/31443)*100=59.75\%$
Red Neuronal	Variables de tipo discreto con variable de predicción discreto	$(17619/31443)*100=56.04\%$

Se consideraron dos modelos más, considerando las mismas variables con el detalle que las columnas de Entidad, Distrito y Sección no son consideradas. Los resultados en la predicción fueran las siguientes.

Algoritmo	Características	Porcentaje de clasificación
Red Neuronal	Variables de tipo continuo con variable de predicción discreto sin las columnas Entidad, Distrito y Sección	$(15341/31443)*100=48.7\%$
Red Neuronal	Variables de tipo discreto con variable de predicción discreto sin las columnas Entidad, Distrito y Sección	$(16807/31443)*100=53.4\%$

De los cuatro modelos anteriores propuestos se observa que las variables de Entidad, Distrito y Sección ayudan a aumentar el porcentaje de clasificación.

También uno puede darse cuenta que los modelos que consideran variables continuas tienen mejores resultados que los que se conforman con valores discretos.

En la siguiente figura se muestra la red de dependencias, que representa a las variables en las que el algoritmo encontró patrones de predicción. Se observa que en sólo 15 de las primeras 25 variables iniciales se encontró información para predecir la participación ciudadana en las elecciones.

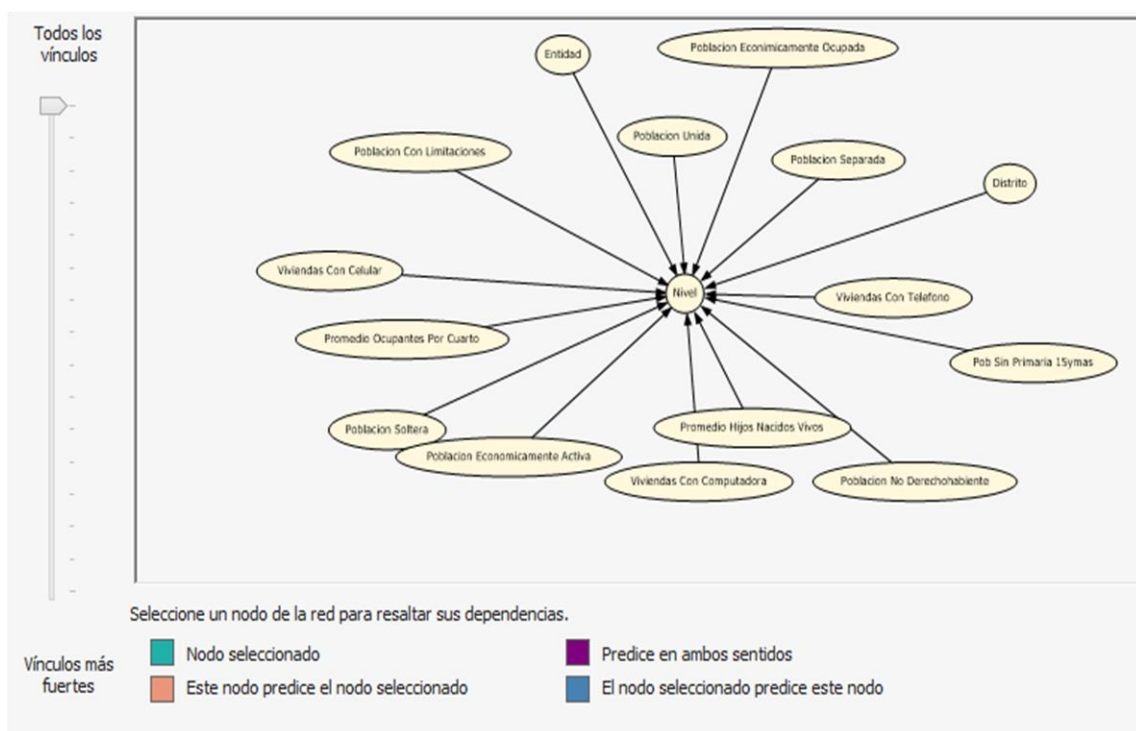


Fig. 29. Red de dependencia.

Se generó más modelos de prueba en los que sólo se consideraron las que se muestran en la imagen anterior. Teniendo como base las predicciones anteriores en donde los modelos con columnas con datos continuos presentaron mejores resultados, sólo se consideró este tipo de datos para la predicción lo cual mostró los siguientes resultados.

Algoritmo	Características	Porcentaje de clasificación
Red Neuronal	15 variables de tipo continuo con variable de predicción discreto	$(19084/31443) * 100 = 58.70\%$

Se observa que el porcentaje de clasificación aumentó aunque con menos variables a considerar pero con las mismas características descritas al principio, y como se muestra en

la Fig.30 el algoritmo redujo aún más la cantidad de variables que predicen el nivel de participación ciudadana en las elecciones.

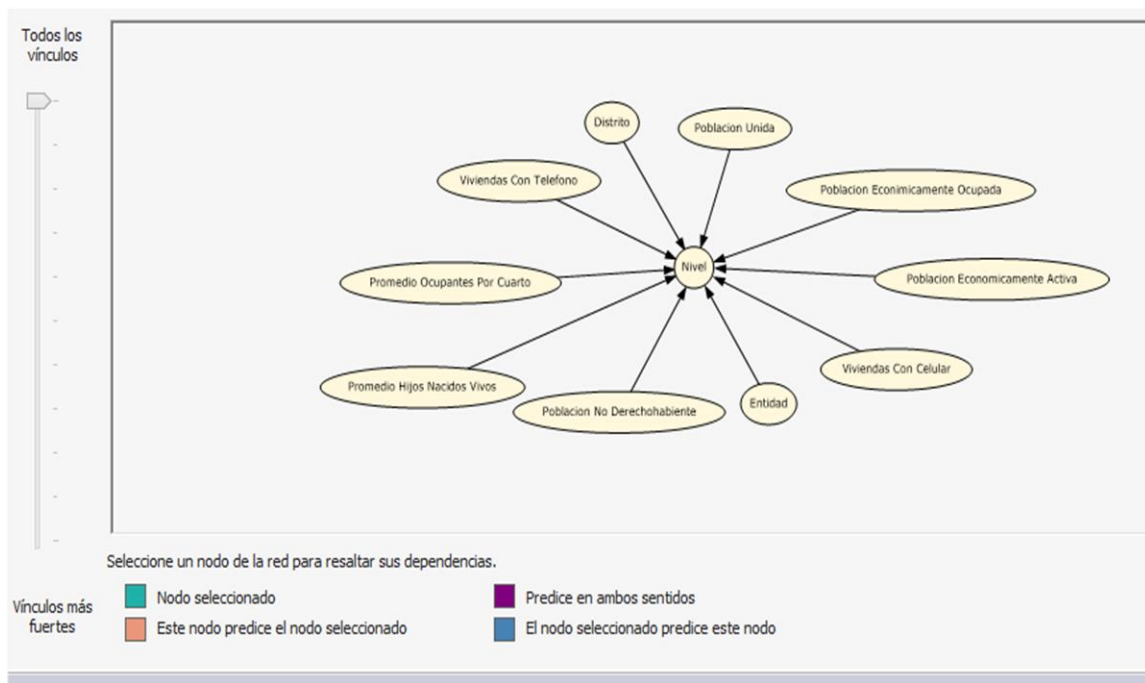


Fig. 30. Nueva Red de Dependencia.

Se creó un nuevo modelo con el mismo algoritmo de redes neuronales, seleccionando el 30% de la población y con las variables arriba descritas obteniendo el resultado siguiente.

Algoritmo	Características	Porcentaje de clasificación
Red Neuronal	10 variables de tipo continuo con variable de predicción discreto	$(18999/31443)*100=60.43\%$

Como se muestra en la tabla anterior, no cambió mucho el mejor resultado en comparación del resultado del modelo que se creó anteriormente. Se podría decir que con menos variables se logró prácticamente el mismo porcentaje de clasificación, mostrando también que el modelo no redujo el número de variables en las que encontró patrones de predicción. Las pruebas siguientes que se hicieron aumentado la población de aprendizaje a 50% o más no mejoraron el mejor porcentaje de clasificación.

De lo anterior y de todos los modelos que se propusieron, se puede concluir que el modelo que se busca para predecir la participación ciudadana en las elecciones federales está conformado por diez variables que son: **Entidad, Distrito, índice de población unida, índice de población económicamente ocupada, índice de población económicamente activa, índice de viviendas con celular, índice de población no derechohabiente, promedio de hijos nacidos vivos, promedio de ocupantes por cuarto y el índice de viviendas que tienen teléfono fijo (Fig. 5.11).**

Ya teniendo el modelo definido y la red neuronal entrenada, se hicieron más pruebas de predicción. En la primera sólo se consideró datos en el cual su nivel de participación electoral era de 1, donde mayor porcentaje de predicción fue el siguiente.

<b>Algoritmo</b>	<b>Características</b>	<b>Salidas</b>	<b>Porcentaje de clasificación</b>
Red Neuronal	10 variables de tipo continuo con variable de predicción discreto	1	77.43%

En la siguiente prueba sólo se consideraron datos en donde el nivel de participación electoral es 2.

<b>Algoritmo</b>	<b>Características</b>	<b>Salidas</b>	<b>Porcentaje de clasificación</b>
Red Neuronal	10 variables de tipo continuo con variable de predicción discreto	2	51.68%

Por ultimo sólo se usaron datos en donde el nivel de participación electoral era de 3, considerado el más alto. El porcentaje de clasificación fue el siguiente.

<b>Algoritmo</b>	<b>Características</b>	<b>Salidas</b>	<b>Porcentaje de clasificación</b>
Red Neuronal	10 variables de tipo continuo con variable de predicción discreto	3	35.55 %

Lo que se puede observar en las tres últimas tablas, es que cuando las variables son únicamente con salida de nivel 1, el porcentaje de clasificación llega casi al 80 por ciento y va disminuyendo ese porcentaje cuando las muestras son de salida de los dos niveles siguiente. Esto es porque la red neuronal en sus datos de entrenamiento existen en su mayoría datos que corresponden a salidas de nivel 1 y en menor cantidad datos con salidas de nivel 2 y 3. En pocas palabras tiene más datos de aprendizaje de nivel 1 que los demás.

Las siguientes pruebas que se hicieron fueron en Matlab, se probó con el algoritmo de Redes Neuronales que se encuentra disponible en su toolbox. Se creó un archivo en .txt en donde se encuentran los datos con las variables que conforman el modelo final de inteligencia de negocios propuesto con sus respectivas salidas de predicción.

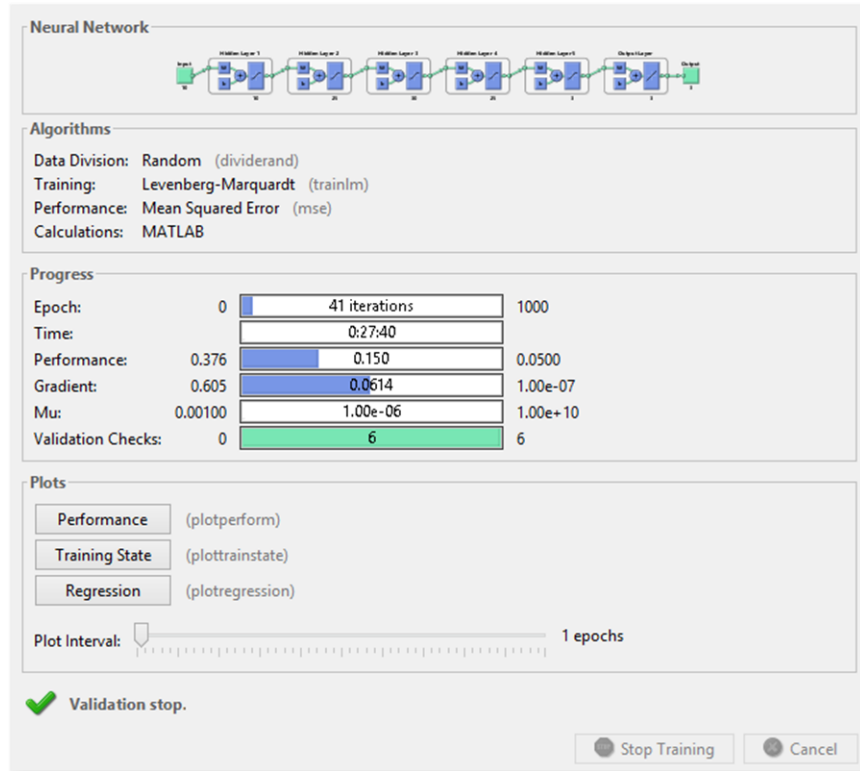


Fig. 31. Entrenamiento de Red Neuronal de la toolbox de Matlab.

Se probaron diferentes estructuras para la red neuronal. A continuación se presentan las estructuras que se utilizaron y las que mejor resultados presentaron.

Entradas	Salidas	Nodos Capa de salida	Capas ocultas	Nodos	Porcentaje de Predicción
10	3	3	3	[6 4 7]	50.15%
10	3	3	5	[10 6 4 7 3]	58.14%
10	3	3	5	[10 10 15 10 3]	57.2271 %
10	3	3	5	[10 25 30 25 3]	60.8960 %
10	3	3	3	[30 35 30]	55.4571%
10	3	3	5	[10 30 30 30 3]	57.286%

Estas tan sólo es una muestra de las estructuras que se probaron. Como se puede observar, la estructura que tiene 5 capas ocultas con y con la distribución de nodos [10 25 30 25 3] fue la que mejor porcentaje de predicción se obtuvo con 60.8960%.

Cuando se hacía el análisis y se probaban las estructuras, se vio que las estructuras con menos de 20 nodos y menos de 2 capas ocultas el porcentaje de predicción era menos del 50 por ciento. Entonces se procedió a aumentar el número de nodos en cada capa oculta y la cantidad de capas ocultas hasta que se llegó a la estructura [10 25 30 25 3] que fue la que mejor porcentaje de predicción consiguió. Pensando en la misma dinámica, se pensó en aumentar aún más el número de nodos de la red pero el porcentaje de predicción no mejoraba, sino que al contrario la predicción disminuía.

## 5.4. Rangos para Predicción

Hasta ahora se han propuesto las variables que predicen la participación ciudadana en las elecciones federales, pero con eso, la información que necesitan los tomadores de decisiones está incompleta, ya que se requiere de los rangos en los que los tres niveles de participación electoral se encuentran.

Por todo lo anterior se tienen que analizar los patrones que el algoritmo de Redes Neuronales encontró en cada variable que considera para determinar a qué nivel de participación pertenece cada salida.

Como se mencionó en capítulos anteriores SQL Server 2014 tiene un visor que muestra el impacto de los atributos sobre un par de valores de salidas relacionadas con la predicción. En las siguientes imágenes se muestra el visor de salidas del modelo con mejor porcentaje de predicción.

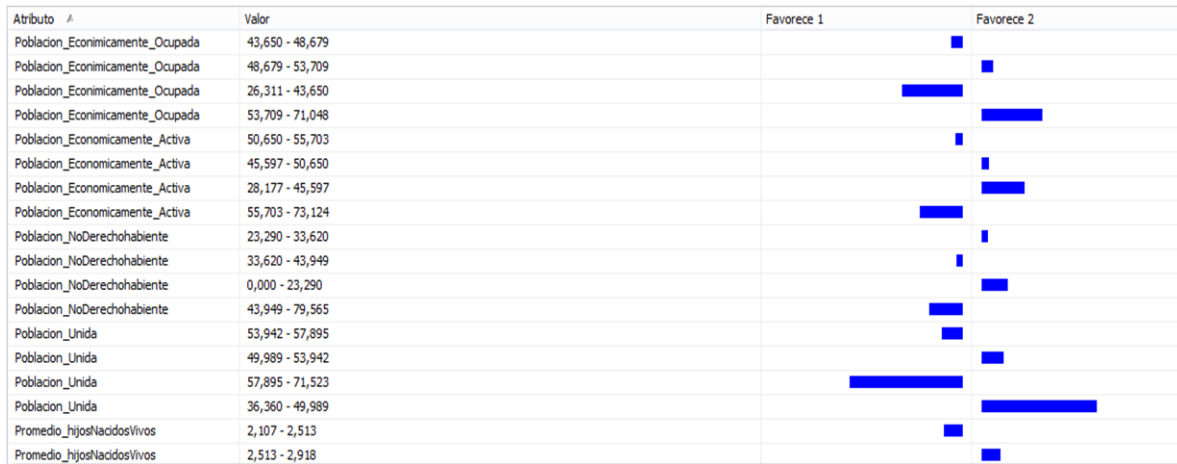


Fig. 32. Rangos de predicción 1.



Fig. 33. Rangos de predicción 2.

En las imágenes anteriores se observan 4 columnas, la primera es el nombre del atributo, la segunda representa el rango en el que el algoritmo matemático encontró como patrón y las últimas dos columnas son las salidas que se están comparando donde las barras azules significan que salida es más probable que sea beneficiada.

Cada rango de cada atributo asigna una probabilidad a cada una de las salidas, siendo que la suma de estas tres probabilidades es 100 por ciento. Estas probabilidades ayudan a predecir el nivel de participación que habrá en la combinación de los atributos del modelo. En la siguiente tabla se muestran los rangos de los atributos y las probabilidades de ocurrir en cada salida.

Atributo	Rango	Nivel 1	Nivel 2	Nivel 3
Poblacion_Unida	36.360 – 49.989	17.07 %	56.53 %	26.38 %
Poblacion_Unida	57.895 - 71.523	77.74 %	17.77 %	4.47 %
Poblacion_Unida	49.989 – 53.942	39.80 %	44.21 %	15.97 %
Poblacion_Unida	53.942 – 57.895	53.16 %	35.42 %	11.40 %
Promedio_hijosNacidosVivos	2.918 – 4.314	18.29 %	53.03 %	28.66 %
Promedio_hijosNacidosVivos	0.711 – 2.107	75.57 %	19.88 %	4.54 %
Promedio_hijosNacidosVivos	2.513 – 2.918	40.27 %	43.56 %	16.16 %
Promedio_hijosNacidosVivos	2.107 – 2.513	52.64 %	36.03 %	11.32 %
Promedio_Ocupantes_Cuarto	1.347 – 2.161	72.74 %	21.79 %	21.79%
Promedio_Ocupantes_Cuarto	0.430 – 0.875	26.94 %	50.93 %	22.12 %
Promedio_Ocupantes_Cuarto	0.815 - 1.111	40.90 %	43.37 %	15.71 %
Promedio_Ocupantes_Cuarto	1.111 – 1.347	52.02 %	36.30 %	11.67 %
Poblacion_Economicamente_Ocupada	53.709 – 71.048	27.57 %	48.32 %	24.09 %
Poblacion_Economicamente_Ocupada	26.311 – 43.650	65.54 %	27.68 %	6.76 %
Poblacion_Economicamente_Ocupada	48.679 – 53.709	42.63 %	42.01 %	15.34 %
Poblacion_Economicamente_Ocupada	43.650 – 48.679	50.25 %	37.71 %	12.03 %
Viviendas_conCelular	0.000 – 41.015	31.68 %	48.90 %	19.40 %
Viviendas_conCelular	75.014 – 100.00	58.43 %	31.74 %	9.80 %
Viviendas_conCelular	41.015 – 58.015	42.90 %	42.22 %	14.87 %
Viviendas_conCelular	58.015 -75.014	49.99 %	37.57 %	12.42 %
Viviendas_conTelefono	58.350 – 99.195	34.18 %	48.89 %	16.92 %
Viviendas_conTelefono	0.0 – 23.163	56.48 %	32.46 %	11.04 %
Viviendas_conTelefono	23.167 – 40.759	49.50 %	37.66 %	12.82 %
Viviendas_conTelefono	40.759 - 58.350	43.40 %	42.16 %	14.41 %
Poblacion_Economicamente_Activa	55.703 – 73.124	60.02 %	31.26 %	8.70 %
Poblacion_Economicamente_Activa	28.177 – 45.597	42.90 %	46.70 %	20.38 %
Poblacion_Economicamente_Activa	45.597 – 50.650	43.77 %	41.45 %	14.76 %
Poblacion_Economicamente_Activa	50.650 – 55.703	49.10 %	38.34 %	12.54 %
Poblacion_NoDerechohabiente	43.949 – 79.565	59.10 %	34.01 %	6.88 %
Poblacion_NoDerechohabiente	0.0 – 23.290	35.66 %	42.01 %	22.31 %
Poblacion_NoDerechohabiente	33.620 – 43.949	48.96 %	39.01 %	12.01 %
Poblacion_NoDerechohabiente	23.290 – 33.620	43.88 %	40.71 %	15.40 %

Fig. 34. Salidas de la Red Neuronal.

Cada variable perteneciente al modelo propuesto, al aplicar el algoritmo de Redes Neuronales obtiene distintos rangos en donde cada salida tiene una probabilidad de ocurrir, como se muestra en la tabla anterior. Con esta información que la Inteligencia de Negocios proporciona, el tomador de decisiones puede pensar en estrategias para que la participación ciudadana en las elecciones federales se mantenga siempre arriba del 51%.

# **CAPITULO 6**

## 6. RESULTADOS Y CONCLUSIONES

En los capítulos anteriores se describió todo el procedimiento que se realizó, desde dar formato y limpieza de los datos que se encontraban en una forma plana hasta encontrar los patrones y un porcentaje de predicción correcta.

El cubo dinámico de información OLAP es una herramienta muy eficiente de la Inteligencia de Negocios en donde se puede hacer análisis de datos, pero dicho análisis nunca estará completo ya que requiere de un algoritmo matemático que indique los patrones de predicción con un porcentaje de aceptación.

El algoritmo de Redes Neuronales ayudó a modelar el comportamiento humano en la participación en las elecciones de diputados federales, ayudando a proponer diez variables que predicen el nivel de participación que habrá en las elecciones. Dichas variables son: la Entidad, el Distrito, la Población Económicamente Activa, la Población Económicamente Ocupada, Población No Derechohabiente, Población Unida, Promedio de Hijos Nacidos Vivos y el Promedio de Ocupantes por cuarto.

De las variables mencionadas en el párrafo anterior, el algoritmo de Redes Neuronales encontró patrones que se necesitan para predecir el nivel de participación y los cuales tienen que ser analizados ya que es la información que necesita el tomador de decisiones para implementar políticas o estrategias que ayuden a aumentar la participación en las elecciones.

Con los patrones que el algoritmo de Redes Neuronales se encontró, se puede concluir lo siguiente:

En primer lugar, se encontraron las situaciones que cumplen las variables del modelo propuesto de Inteligencia de Negocios, en las que se fomenta que la participación ciudadana sea baja o menos 51 % en las elecciones o su probabilidad sea mayor.

Se encontró que cuando el índice de población unida se encuentre en el rango de 53.942 a 57.895 por ciento, la probabilidad de que el nivel de participación electoral sea menor al 51% es de 53.16%. Cuando el rango del mismo índice está entre 57.895 y 71.523 la probabilidad de que la participación electoral sea de nivel 1 aumenta al 77.74%. Esto nos podría indicar que en las poblaciones donde su mayoría es casada o vive en pareja, la participación es baja.

En poblaciones donde el promedio de hijos nacidos vivos está entre 0.711 y 2.107, la probabilidad de que la participación ciudadana en las elecciones federales sea de nivel 1 es de 75.57%, y cuando se encuentra entre 2.107 y 2.513 la probabilidad es de 52.64%. Esto nos dice que la población que no tiene hijos o tiene menos de 2, casi no participa en las épocas electorales.

Con la variable que respecta al promedio de ocupantes por cuarto en las viviendas, el algoritmo de Redes Neuronales encontró que la probabilidad de que la participación ciudadana sea baja debe estar entre 1.111 y 1.347 con 52.02%. Para que la probabilidad sea de 72.74% el rango debe de ser de 1.347 a 2.161. Con esta información se puede concluir que

las familias que tienen menos hijos o a lo mucho 2, son las que menos participan en las elecciones federales.

Un dato relevante es que en poblaciones en donde índice de población económicamente ocupada va de 26.311 a 43.650 por ciento, la probabilidad de que la participación ciudadana en las elecciones sea de nivel bajo es de 65.54%. Esto nos podría indicar que en poblaciones donde es bajo el porcentaje de personas que tienen trabajo, estas no están interesadas en votar.

En una situación similar a la del párrafo anterior, una variable del modelo de inteligencia de negocios que se propone en este proyecto de investigación es la del índice de población económicamente activa, en donde se encontró que cuando es alto la cantidad de personas que están buscando trabajo o para ser más exactos entre 55.703 y 73.124, la probabilidad de que no se participe es de 60.02%. Esto nos indicaría que cuando la población no siente estabilidad económica, lo que menos le preocupa es participar en una elección de diputados federales.

Cuando en una sección electoral existen viviendas donde en su mayoría tienen celular o para ser exactos en un rango de 75.014 a 100 por ciento la probabilidad de que su índice de participación ciudadana es de 58.43%. Este podría ser un dato curioso y que no se esperaba, ya que con el resultado que se consiguió se puede concluir que el celular es un importante inhibidor de participación ciudadana en las elecciones a diputados.

En un situación inversa cuando en la sección electoral existen viviendas donde del 0.0 al 23.163 por ciento cuentan con teléfono fijo, la probabilidad de que la participación electoral sea de menos del 51% es de 56.48%. Se puede inferir que en las poblaciones que menos acceso a un teléfono tiene, es la que menos acudirá a votar.

Un dato que debería de ser muy importante para los tomadores de decisiones, ya que es algo de lo que se pueden generar estrategias a mediano plazo, es en el índice de población que no es derechohabiente o no tiene acceso a los servicios de salud públicos. Se encontró que cuando en la población de un sector el rango de personas que no tienen acceso va del 43.949 al 79.565 por ciento, la probabilidad de que la participación sea baja es del 59.10%. Lo que indicaría que cuando los ciudadanos no tienen garantizado el derecho a la salud, no les interesa participar en los comicios electorales federales.

Por otro lado, el algoritmo de Redes Neuronales encontró los rangos de cada una de las variables del modelo en donde la probabilidad de que la participación ciudadana es mayor al 51 por ciento. Esta información ayuda al tomador de decisiones a ver que se está haciendo bien y seguir fomentando las condiciones para que el ciudadano salga a votar. Dichos rangos son descritos a continuación.

En las secciones electorales donde el índice de población unida este en el rango de 36.360 a 49.989, la probabilidad de que la participación en elecciones sea de nivel 2 es de 56.53% y también se tiene la probabilidad de que sea de nivel 3 que es de 26.38%. Es nos quiere decir que en las poblaciones donde del 50 al 75 por ciento aproximadamente es soltera, la gente sale más a votar.

En poblaciones donde el promedio de hijos que tienen por familia va de 2.918 a 4.314 la probabilidad que la participación ciudadana sea de nivel 2 y nivel 3 es de 53.03% y 28.66% respectivamente. Es un dato interesante que las familias que en su núcleo este conformado de tres a más hijos sean las que más votan en las elecciones.

Se encontró que en las secciones electorales donde la probabilidad de que la participación sea de nivel 2 sea de 50.93% y de nivel 3 con 22.12%, es en donde el promedio de ocupantes por cuarto esta entre 0.43 y 0.875.

En un párrafo se mencionó que cuando en las secciones electorales en donde poca gente tiene trabajo, no participa en los comicios electorales. En una situación contraria cuando el índice de población económicamente ocupada esta entre 53.709 y 71.048 la probabilidad de que las elecciones alcancen una participación de nivel 2 y nivel 3 respectivamente es de 48.38% y de 24.06%. Por lo tanto cuando la gente siente una seguridad económica se motiva a participar en las elecciones.

En el caso del índice de población económicamente activa, el sector electoral alcanza un nivel de participación 2 con una probabilidad de 46.70% y el nivel 3 con una probabilidad de 20.28%, todo esto en un rango del 28.177 a 45.597.

Una de las conclusiones a la que se llegó, es que cuando las viviendas de un sector electoral tiene acceso a celular en un rango que va 0.00 al 41.015 por ciento, es muy probable que la participación electoral de nivel 2 es del 48.9 por ciento y 19.40 por ciento de nivel 3. Es un dato que los tomadores de decisiones tienen que analizar el por qué, pero desde una perspectiva que se tiene, es que podría estar relacionado con lo que se menciona en artículo del IFE del 2013, donde indica que las comunidades rurales son las que más participan en los comicios electorales.

En secciones electorales donde el promedio de población que tiene teléfono está entre 58.350% y 99.195%, la probabilidad de que la participación sea de nivel 2 es 48.89% y de que sea de nivel 3 es de 16.92%.

Un dato importante que debe ser relevante, es a cerca de del índice de población que no es derechohabiente, porque se descubrió que las personas que más votan son las que menos acceso a los servicios de salud tienen. Eso se ve en los datos que el algoritmo de Redes Neuronales ayudó a descubrir. Se encontró que cuando el índice se encuentra entre 0.0 a 23.290 por ciento en las secciones electorales, la probabilidad de que la votación sea de nivel 2 es de 42.01% y la probabilidad de que sea de nivel 3 alcanza 22.31%. El llegar a la conclusión a la conclusión que las poblaciones en donde el índice de derechohabientes es bajo, la participación ciudadana en las elecciones tiende a ser mayor al 51%, es algo que los tomadores de decisiones deben estudiar a profundidad para descubrir cuáles son los fenómenos que fomentan este hecho.

En este proyecto se describieron dos tipos de análisis, el primero se usó mediante el cálculo directo en el cubo de información OLAP, seleccionando las dimensiones que pertenecen al modelo de Inteligencia de Negocios y buscar que combinación de los distintos niveles que tiene cada dimensión, cual es la que consiguió la mayor cantidad de votos.

El otro análisis que se hizo, fue determinar las variables pertenecientes al modelo de Inteligencia de Negocios con datos continuos, teniendo cada registro una variable de salida o de predicción para definir a que nivel de participación ciudadana en elecciones federales pertenece. Teniendo lo anterior estos datos se analizaron con el algoritmo de Redes Neuronales, el cual encuentra patrones que ayudan a determinar a qué nivel de participación pertenecen el conjunto de características del modelo. Esto ayuda a que si metemos un nuevo conjunto de datos que contengan las mismas variables del modelo, se pueda predecir el nivel de participación electoral con un porcentaje de confianza.

Lo que es prudente hacer a estas alturas, es la de comparar que resultados de los dos análisis que se hicieron en esta investigación. A continuación se presenta la comparación de los análisis, la cual muestra las características en las que la ciudadanía tiene una mayor participación en las elecciones federales a Diputados.

<b>Atributo del Modelo de BI</b>	<b>Análisis con Algoritmo de Redes Neuronales</b>	<b>Análisis en cubo OLAP</b>
Poblacion_Unida	36.360 – 49.989 %	50 - 55%
Promedio_hijosNacidosVivos	2.918 – 4.314 %	2 - 3 hijos
Promedio_Ocupantes_Cuarto	0.430 – 0.875 %	-----
Poblacion_Economicamente_Ocupada	53.709 – 71.048 %	40 - 50 %
Viviendas_conCelular	0.000 – 41.015 %	60 - 80%
Viviendas_conTelefono	58.350 – 99.195 %	42 - 100%
Poblacion_Economicamente_Activa	28.177 – 45.597 %	50 - 60%
Poblacion_NoDerechohabiente	0.0 – 23.290 %	30 - 40%

*Fig. 35. Comparación de los dos metodos de análisis.*

Se puede observar de la tabla anterior que existen inconsistencias entre los resultados de cada uno de los análisis, pero también hay algunas similitudes.

En uno de los atributos donde se ve claramente las inconsistencias, es en el de Viviendas\_conCelular. Mientras que el cubo nos indica que donde hay más votos es en el rango de 60 a 80 por ciento, el análisis con el algoritmo de redes neuronales nos dice que en las poblaciones donde más participan son en las que se encuentran entre 0.0 y 41.015 por ciento de acceso a telefonía celular.

Otra de las inconsistencias está en la variable de Poblacion\_NoDerechohabiente, donde el cubo nos muestra que la población más participativa está en el rango de 30 a 40%, y el algoritmo de Redes Neuronales nos dice que está entre 0.0 y 23.290%.

En las similitudes que se encontraron fue en la variable de Viviendas\_conTelefono. El algoritmo de Redes Neuronales encontró que las personas que más participan son las que se encuentran en el rango de 58.350 a 99.195 por ciento con acceso a telefonía fija, mientras que el análisis con el cubo encontró el rango entre 42 y 100 por ciento.

Al analizar los resultados obtenidos, se puede concluir que el mejor método de análisis para este trabajo de investigación es aplicando el algoritmo de Redes Neuronales. El análisis con los cubos de información OLAP son de gran ayuda con pocas variables, pero cuando existen muchas variables a considerar se complica el observar los resultados. En la búsqueda de los rangos las Redes Neuronales pueden ser más exactas que el simple análisis con el cubo ya que el primero encuentra patrones y el segundo hace su análisis en base a los niveles ya establecidos cuando se discretizaron los datos, por ende, ahí puede dejar de ser exacto ya que el nivel encontrando puede no estar completo.

Durante la búsqueda de las variables que conformaría el modelo de Inteligencia de Negocios, se procesaron y probaron diferentes modelos hasta llegar al modelo final conformado por las ocho variables descritas en capítulos anteriores y qué alcanzó el porcentaje de predicción más alto. Al procesar uno de esos modelos se tuvo alcanzó el 60.437% de predicción, un porcentaje muy bueno y similar al modelo final.

El modelo que presentó resultados similares se presenta en la siguiente imagen.

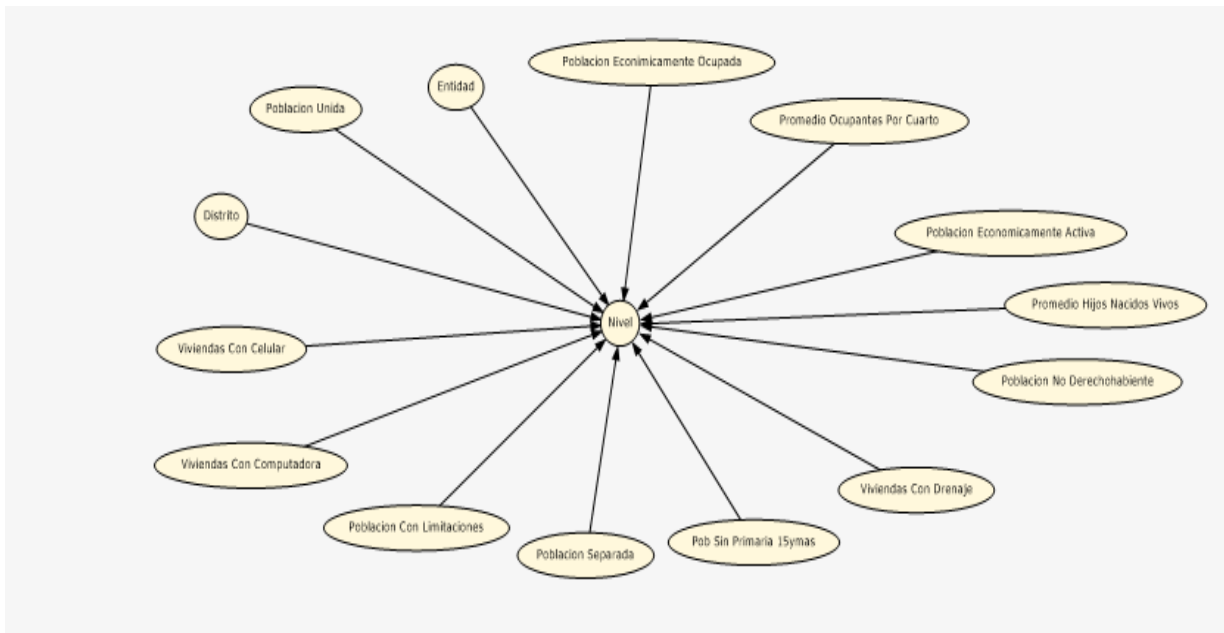


Fig. 36. Segundo Modelo Encontrado.

Se siguió la misma metodología que la descrita en el capítulo 5, tratando de descartar variables en las que el algoritmo de redes neuronales no encontraba ningún patrón, pero el modelo siempre se mantuvo y no descartó ninguna otra variable manteniendo los porcentajes de predicción entre 58 y 60.437 por ciento, aunque en otras ocasiones fluctuó entre 53 y 56.

Al comparar los dos modelos propuestos, se encontró que ambos consideran nueve variables iguales que son: Entidad, Distrito, Promedio de Ocupantes por Cuarto, Población Económicamente Activa, Población Económicamente Ocupada, Población No Derechohabiente, Población Unida, Promedio de Hijos Nacidos Vivos y Viviendas con Celular.

La variable que el modelo final considera y el nuevo modelo encontrado descarta es la de Viviendas con Teléfono. En su lugar el nuevo modelo considera cinco variables nuevas que son las Viviendas con Computadora, Población con Limitaciones, Población Separada, Población sin Primaria, y Viviendas con Drenaje; estando conformado el nuevo modelo por 14 variables.

Se llegó a la conclusión de que el modelo principal propuesto en esta tesis es el descrito en el capítulo 5, ya que el porcentaje de predicción correcta que arrojaba promediaba entre 59.2% y 60.4%, mientras que el modelo que está conformado por 14 variables fluctuó entre 53% y 60.4% de predicción correcta. Siendo que con menos variables, en todos los casos la predicción se mantenía cerca del mejor resultado obtenido.

Se puede concluir que los dos modelos son válidos, pero si el tomador de decisiones al querer hacer predicciones de alguna elección futura y no cuenta con la variable del índice de viviendas que cuentan con teléfono, podría sustituir esa variable por las cinco que fueron descritas en el párrafo anterior; sabiendo que el porcentaje de predicción correcta puede llegar a ser el mismo que el modelo de 10 variables propuesto.

Después de realizar la investigación y analizar los resultados obtenidos, se puede decir que el objetivo que se planteó se cumplió. Ya que se ha diseñado un modelo basado en Inteligencia de Negocios, el cual ha generado información con la cual el tomador de decisiones puede generar estrategias para fomentar y mejorar la participación ciudadana en las elecciones federales para Diputados de México.

La aportación más importante de esta tesis, fue la de poder modelar el comportamiento humano, proponiendo las características que influyen en la participación ciudadana en las elecciones federales. Encontrando que el algoritmo matemático de Redes Neuronales son una poderosa herramienta para analizar los datos del modelo y encontrar patrones para llegar a predecir la decisión de la población para votar o no en la elección de representantes a Diputados federales de México.

# Bibliografía

Akinwonmi, A. E., Kuboye, B. M., & Thompson, A. F. (2013). A Neural Network Approach to Selection of Candidates for Electoral Offices by Political Parties. *International Journal of Information Science*, 3(3), 63-69.

Bayona, G. A. T., & Salcedo, I. A. L. (2012). Evaluación de las Redes Neuronales Artificiales Perceptron Multicapa y Fuzzy-Artmap en la Clasificación de Imágenes Satelitales. *Ingeniería*, 17(1), 61-72.

Bevelander, P. (2015). Voting Participation of Immigrants in Sweden—a Cohort Analysis of the 2002, 2006 and 2010 Elections. *Journal of International Migration and Integration*, 16(1), 61-80.

Czernicki, B. (2010). *Silverlight 4 Business Intelligence Software*. Apress.

Czernicki, B. (2011). *Silverlight 4 Business Intelligence Software*. Apress.

Gibbons, P., Y Prescott, J. (1996). Parallel competitive intelligence processes in organizations. *International Journal of Technology, Special Issue on Information Flow Management*, Vol. 11, n° 1-2.

Gill, G. S. (2008). Election Result Forecasting Using Two Layer Perceptron Network. *Journal of Theoretical and Applied Information Technology*, 47(11), 1019-1024.

Gómez, A. A. R., & Bautista, D. W. R. (2010). Inteligencia de negocios: Estado del arte. *Scientia et Technica*, 1(44), 321-326.

Guillen, A., Sáenz, K., Badii, M. H., & Castillo, J. (2009). Origen, espacio y niveles de participación ciudadana (Origin, space and levels of participation). *Daena: International Journal of Good Conscience*, 4(1), 128-148.

Hartley, K., & Seymour, L. F. (2011, October). Towards a framework for the adoption of business intelligence in public sector organisations: the case of South Africa. In *Proceedings of the South African Institute of Computer Scientists and Information Technologists Conference on Knowledge, Innovation and Leadership in a Diverse, Multidisciplinary Environment* (pp. 116-122). ACM.

Hayashi, Y., Hsieh, M. H., & Setiono, R. (2010). Understanding consumer heterogeneity: A business intelligence application of neural networks. *Knowledge-Based Systems*, 23(8), 856-863.

Hilera, J.R., Martínez, V.J. (2000). Redes Neuronales Artificiales. Fundamentos, modelos y aplicaciones. *Alfaomega*.

Khaze, S. R., Masdari, M., & Hojjatkah, S. (2013). Application of Artificial Neural Networks in Estimating Participation in Elections. *arXiv preprint arXiv:1309.2183*.

Kolovos, I., & Harris, P. (2005). Voter apathy in British elections: Causes and Remedies.

López, I., Fernandez, A., Machado, E., Agustín, L., Zabala, E. (2006). *La inteligencia Competitiva como herramienta de innovación*. Recuperado el 20 de junio 2011, de <http://www.ingefraf.es/XVIII/PDF/Comunicacion17012.pdf>.

MacLennan, J., & Tang, Z. (2005). *Data Mining with SQL Server 2005*. Wiley.

Maila, H. M. (2009). Performance management and service delivery in the Department of Water Affairs and Forestry (DWAF).

Martínez, M. T. V. PARTICIPACIÓN CIUDADANA Y POLÍTICAS PÚBLICAS.

Mendieta, A. (2012). Participación Política en América Latina y sus repercusiones en México. Una Propuesta de Análisis, *Revista de Comunicación Vivat Academia*. ISSN: 1575-2844.

México. Centro de Estudios Sociales y de Opinión Pública, LXI Legislatura de la Cámara de Diputados. (2012). *Abstencionismo y cultura política en México*. Abril 2012.

México. Instituto Federal Electoral. (2013). *Estudio Censal de la participación ciudadana en las elecciones federales del 2012*, Octubre 2013.

Pérez, H., Y Valdez, A. (2014). Inteligencia competitiva en campañas electorales, El caso del sistema de partidos políticos en Jalisco, México. *Revista Iberoamericana de Ciencias*, ISSN 2334-2501. Vol.1 N°.5.

Pérez, M., Y., & Amador, S. (2009). Características de los sistemas de información que permiten la gestión oportuna de la información y le conocimiento institucional, *ACIMED*. 20(5), 66-71.

Pérez, S. (1999). Gobierno y Participación Ciudadana En Quórum Año VIII, No. 68, septiembre- octubre, México, p. 178.

Radcliff, B. (2001). Organized labor and electoral participation in American national elections. *Journal of Labor Research*, 22(2), 405-414.

Rajan, J. (2009). Business Intelligence: Concepts, Components, Techniques and Benefits. *Journal of Theoretical and Applied Information Technology*, JATIT 2005-2009.

Smith, P. H. (2004). Los ciclos de democracia electoral en América Latina, 1900-200. *Política y gobierno*, 11(2), 189-228.

Valdez, A. (2012). Inteligencia competitiva en campañas electorales. *RMC 132*.  
Noviembre 2012.