**CARRERA INGENIERÍA DE SISTEMAS**

UNIVERSIDAD PÚBLICA DE EL ALTO



**TESIS DE GRADO**

**“MODELO DE PROYECCIÓN DE CONSUMO DE ENERGÍA ELÉCTRICA CON MINERÍA DE DATOS**

**CASO: EL ALTO”**

Para obtener el título de Licenciatura en Ingeniería De Sistemas

**Mención: Gestión y Producción**

**Postulante: Univ. Daniel Choque Canaviri**

**Tutor Metodológico: Ing. Wilfredo Pérez Aquino**

**Tutor Especialista: Ing. Enrique Flores Baltazar**

**EL ALTO - BOLIVIA**

**2016**

1. MARCO PRELIMINAR
   1. INTRODUCCIÓN

La energía eléctrica es un recurso valioso e imprescindible, para cualquier sociedad desarrollada debido a sus ventajas tanto en su producción como en la diversidad de sistemas y estrategias para obtenerla, puesto que su fácil distribución y transporte hacen factible la disponibilidad del mismo para ser transformada en otros tipos de energía como la luz, calor y energía mecánica.

En Bolivia el 85,4%1 de las viviendas cuenta con energía eléctrica lo que genera una alta demanda del mismo por lo que se requiere una estrategia para una eficiente distribución mediante un modelo que proporcione una proyección para el consumo de energía eléctrica con minería de datos.

Se describe a la Minería de Datos como el proceso de extraer información valida, autentica que se puede procesar de las base de datos de gran tamaño para encontrar patrones o tendencias en los datos para generar modelos.

Una proyección, es la estimación a futuro de un evento mediante un modelo, el cual generara resultados para su posterior interpretación ayudando a la prevención o simulación de eventos.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1 Instituto Nacional de Estadística (INE), 2012

* 1. ANTECEDENTES

**INTERNACIONAL**

**[Ignacio Perversi; 2007]; APLICACIÓN DE MINERÍA DE DATOS PARA LA EXPLORACIÓN Y DETECCIÓN DE PATRONES DELICTIVOS EN ARGENTINA** Diseñar un modelo para detectar los patrones delictivos en Argentina

**Métodos**

Para obtener resultados se utilizó los Algoritmo K-means, Algoritmo ID3 y Algoritmo C4.5 los cuales generan un árbol de decisiones

**Herramientas**

Las herramientas que se utilizaron para el análisis fueron

**Tabla de Centroides**, permite conocer cuál es el centroide de cada cluster. En un sentido geométrico, el centroide es el lugar del hiper-espacio de posibles estados que equidista de todos los casos que corresponden a un determinado cluster.

**Gráficos de dispersión**, Los gráficos de dispersión se representan en ejes cartesianos: cada eje representa un atributo y cada punto un hecho. Estos gráficos tienen la particularidad que permiten incorporar virtualmente una tercera dimensión mediante la asignación de distintos colores a los puntos.

**Gráficos de distribución**, Estos gráficos de dispersión permiten visualizar 3 atributos al mismo tiempo e identificar cual es la interrelación que subyace entre ellos. Por lo general el atributo que se encuentra en la dimensión de color es el cluster.

**Matriz de confusión**, permiten entender cuál es el error que comete un árbol de clasificación al intentar clasificar todos los registros.

**Resumen**

Se logró demostrado no sólo que es factible aplicar minería de datos a la información criminal en Argentina, sino también su alto valor agregado para el análisis y la generación de nuevo conocimiento.

INSTITUTO TECNOLÓGICO DE BUENOS AIRES – ÁREA DE INGENIERÍA, BUENOS AIRES - ARGENTINA

**[MAGDALENA SERVENTE; 2002]; ALGORITMOS TDIDT APLICADOS A LA MINERIA DE DATOS INTELIGENTE**

Se logró desarrollar un sistema multiagente para simulación social con el objetivo del estudio del consumo eléctrico de una sociedad.

**Métodos**

Utiliza métodos de estimación por demanda introduciendo como datos de entrada las características de las poblaciones.

**Herramientas**

**Swarm**, Probablemente la herramienta de simulación más conocida, debido a que fue una de las pioneras, consiste en una colección de librerías portables y orientadas a objetos de componentes reutilizables que proporcionan un conjunto de herramientas flexible.

**Mason**, es una plataforma para la simulación de eventos discretos en máquinas monoprocesador, dotada con herramientas de visualización muy potentes, gratuita y de código abierto

**SeSAm,** (Shell for Simulated Agent Systems) es otro intento de facilitar el modelado y la experimentación de sistemas ABM (agent-based model). Está escrito en Java y es código abierto.

**Resumen**

Se desea enmarcar en el área de la simulación la aparición de los sistemas multiagente como herramienta de modelado de sistemas complejos, así como hacer una breve introducción general a los conceptos básicos relacionados con dichos sistemas.

UNIVERSIDAD COMPLUTENSE DE MADRID - FACULTAD DE INFORMÁTICA

 MADRID- ESPAÑA

**[Zoraida Emperatriz Mamani Rodríguez;2015] APLICACIÓN DE LA MINERIA DE DATOS DISTRIBUIDA USANDO ALGORITMO DE CLUSTERING K-MEANS PARA MEJORAR LA CALIDAD DE SERVICIOS DE LAS ORGANIZACIONES MODERNAS CASO: PODER JUDICIAL**

Desarrolla un prototipo que aplica minería de datos distribuida sobre datos nominales para determinar patrones de comportamiento basado en el petitorio que presenta la carga procesal de los periodos 2008 al 2010 correspondiente a órganos jurisdiccionales casatorios.

**Métodos**

**Minería de datos distribuida**, es el proceso de descubrimiento de conocimiento en arquitecturas de datos que son totalmente diferentes al enfoque centralizado. Esto comprende las fuentes de datos distribuidas, el cómputo distribuido y las comunicaciones.

**Herramientas**

**DMS:** Suite de minería de datos; este tipo de herramienta está ampliamente enfocado en minería de datos e incluye gran número de métodos

**MAT**: Este tipo de herramienta no está enfocada precisamente en minería de datos, pero ofrece un conjunto de métodos y rutinas de visualización de resultados

**BI:** No está enfocada estrictamente en minería de datos pero incluye métodos básicos de minería de datos orientadas a métodos estadísticos para aplicaciones de negocios.

**Resumen**

Se orientó a aportar en la factibilidad de aplicar minería de datos distribuida en los organismos públicos permitiendo cubrir los objetivos centrados en el gobierno electrónico contemplado en el PNMSP

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS -FACULTAD DE CIENCIAS MATEMÁTICAS

LIMA – PERÚ

* 1. PLANTEAMIENTO DEL PROBLEMA
     1. PROBLEMA PRINCIPAL

La atención a las consultas realizadas por los contribuyentes en el Servicio de Impuestos Nacionales del GSCCT se realiza de manera manual lo que provoca en algunos casos provoca mucho en los operadores.

¿De qué manera coadyuvaría el modelo de Redes Neuronales para la atención automática a con ChatBot al Servicio de Impuestos Nacionales?

La planificación en la ciudad de El Alto para la generación del consumo eléctrico, se realiza según la demanda que presenta la población sin seguir un modelo o patrón específico lo que genera una deficiente producción y distribución provocando importantes pérdidas en la población y en su economía.

¿De qué manera coadyuvaría el modelo de proyección de consumo de energía eléctrica con minería de datos la ciudad de El Alto?

* + 1. PROBLEMA SECUNDARIOS
* Difícil almacenamiento de la energía eléctrica.
* Los cortes de energía con tención cero provocan una pérdida parcial o total de datos en los equipos de la población.
* La caída de tensión inferior al 80% o 85% provocan efectos similares a los de una sobre tensión dañado equipos en las industrias afectadas.
* la alta demanda de
  1. OBJETIVOS
     1. OBJETIVO GENERAL

Diseñar un modelo capaz de realizar una proyección de consumo de energía eléctrica con minería de datos para la ciudad de EL ALTO.

* + 1. OBJETIVOS ESPECÍFICOS
* Diagnosticar el sistema actual.
* Analizar y construir modelo para la proyección de consumo eléctrico.
* Desarrollar un sistema que permita visualizar la proyección.
* Diseñar el sistema se control.
* Realizar un análisis de costos.
  1. HIPÓTESIS

El modelo para realizar las proyecciones de consumo de energía eléctrica con minería de datos optimizará la generación de energía eléctrica para la ciudad de EL ALTO.

* + 1. IDENTIFICACIÓN DE VARIABLES

**Variable Dependiente**

Consumo de Energía Eléctrica

**Variable Independiente**

Modelo de Proyección con Minería de Datos

* + 1. OPERACIONALIZACION DE VARIABLES

|  |  |  |  |
| --- | --- | --- | --- |
| **Variables** | **Dimensiones** | **Indicadores** | **Actividades** |
| **Dependiente** | Variable cuantitativa | Consumo de energía eléctrica | Estudio |
| Consumo de Energía Eléctrica | Documental |
| **Independiente** | Proyección | estimación a futuro | Estudio Experimental |
| Modelo con Minería de Datos | Modelos | aproximación a la realidad | Estudio Experimental |

* + 1. CONCEPTUALIZACION DE VARIABLES

**Variable Dependiente**

**Consumo de Energía Eléctrica.-** Es la cantidad de energía que se gasta en los diferentes aparatos utilizados dentro de una vivienda, establecimiento o industria.

La cantidad de energía utilizada por hogar, institución o industria varía ampliamente dependiendo del nivel de vida del país, el clima, y la edad y el tipo de residencia.

**Variable Independiente**

**Modelo con Minería de Datos.-** Es un campo de la estadística y las ciencias de la computación referida al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior. Además de la etapa de análisis en bruto, que involucra aspectos de bases de datos y de gestión de datos, de procesamiento de datos, del modelo y de las consideraciones de inferencia, de métricas de Intereses, de consideraciones de la Teoría de la complejidad computacional, de post-procesamiento de las estructuras descubiertas, de la visualización y de la actualización en línea.2

* 1. JUSTIFICACIÓN

La ciudad de El Alto ha tenido un crecimiento alto en su población por lo que la demanda del consumo de energía eléctrica se incrementó considerablemente por lo que el siguiente trabajo se justifica

* + 1. TÉCNICA

El instituto cuenta con los datos y equipos necesarios para realizar proyecciones pero no tiene un modelo para generar proyecciones.

* + 1. ECONÓMICA

Al ser un modelo generado por computadora se logra ahorrar tiempo y costos en el análisis de los datos para generar el modelo evitando una costosa inversión.

* + 1. JUSTIFICACIÓN SOCIAL

La ciudad de El Alto es una de las localidades que experimenta un elevado impacto negativo con los cortes de energía eléctrica debido a su elevado índice de empresas e instituciones con las que cuenta, el estudio que se realizara servirá para tomar proyecciones para evitar los mismos.

* 1. METODOLOGÍA
     1. MÉTODO CIENTÍFICO

El método científico es un método de investigación usado principalmente en la producción de conocimiento en las ciencias.

**Observación**: Es aplicar atentamente los sentidos a un objeto o a un fenómeno, para estudiarlos tal como se presentan en realidad, puede ser ocasional o causalmente.

**Inducción**: La acción y efecto de extraer, a partir de determinadas observaciones o experiencias particulares, el principio particular de cada una de ellas.

**Hipótesis**: Consiste en elaborar una explicación provisional de los hechos observados y de sus posibles causas.

Probar la hipótesis por experimentación.

Demostración o refutación (antítesis) de la hipótesis.

Tesis o teoría científica.

* + 1. INGENIERIA DE SOFTWARE

La ingeniería de software requiere llevar a cabo numerosas tareas agrupadas en etapas, al conjunto de estas etapas se le denomina ciclo de vida. Las etapas comunes a casi todos los modelos de ciclo de vida son las siguientes:

#### Obtención de los requisitos.- Se debe identificar sobre qué se está trabajando, es decir, el tema principal que motiva el inicio del estudio y creación del nuevo software o modificación de uno ya existente.

**Análisis de requerimientos**.- Extraer los requisitos de un producto software es la primera etapa para crearlo.

**Especificación.-**La especificación de requisitos describe el comportamiento esperado en el software una vez desarrollado

**Diseño y arquitectura**.- La integración de infraestructura, desarrollo de aplicaciones, bases de datos y herramientas gerenciales, requieren de capacidad y liderazgo para poder ser conceptualizados y proyectados a futuro, solucionando los problemas de hoy.

**Programación**.- Implementar un diseño en código puede ser la parte más obvia del trabajo de ingeniería de software, pero no necesariamente es la que demanda mayor trabajo y ni la más complicada.

**Prueba de software**.- Consiste en comprobar que el software responda/realice correctamente las tareas indicadas en la especificación. Es una buena praxis realizar pruebas a distintos niveles (por ejemplo primero a nivel unitario y después de forma integrada de cada componente) y por equipos diferenciados del de desarrollo (pruebas cruzadas entre los programadores o realizadas por un área de test independiente).

**Implementación**.- Una Implementación es la realización de una especificación técnica o algoritmos con un programa, componente software, u otro sistema de cómputo

**Documentación**.- Realización del manual de usuario, y posiblemente un manual técnico con el propósito de mantenimiento futuro y ampliaciones al sistema. Las tareas de esta etapa se inician ya en la primera fase pero sólo finalizan una vez terminadas las pruebas.

**Mantenimiento**.- En esta etapa se realizan un mantenimiento correctivo (resolver errores) y un mantenimiento evolutivo (mejorar la funcionalidades y/o dar respuesta a nuevos requisitos).

* 1. HERRAMIENTAS

**Java.-**Un lenguaje de programación de propósito general, concurrente, orientado a objetos que fue diseñado específicamente para tener tan pocas dependencias de implementación como fuera posible3.

**Matlab**.- Es una herramienta de software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio (lenguaje M). Está disponible para las plataformas Unix, Windows, Mac OS X y GNU/Linux. Entre sus prestaciones básicas se hallan: la manipulación de matrices, la representación de datos y funciones, la implementación de algoritmos, la creación de interfaces de usuario (GUI) y la comunicación con programas en otros lenguajes y con otros dispositivos hardware.

* 1. LÍMITES Y ALCANCES
     1. LIMITES

El siguiente proyecto está estimado para realizar una proyección para el consumo eléctrico en la ciudad de El Alto para toma de decisiones e investigaciones futuras.

* + 1. ALCANCES

La siguiente investigación se efectuara para realizar la proyección del consumo de energía eléctrica que se demanden en un tiempo estimado de 5 años.

* 1. APORTES

La tesis aportara a la sociedad con un modelo de proyección, para la toma de daciones en la ciudad de El Alto en su producción y distribución.

ESQUEMA TENTATIVO

CAPITULO I

1. MARCO PRELIMINAR
   1. INTRODUCCIÓN
   2. ANTECEDENTES
   3. PLANTEAMIENTO DEL PROBLEMA
      1. PROBLEMA PRINCIPAL
      2. PROBLEMA SECUNDARIOS
   4. OBJETIVOS
      1. GENERAL
      2. OBJETIVOS ESPECÍFICOS
   5. HIPÓTESIS
      1. IDENTIFICACIÓN DE VARIABLES
      2. OPERACIONALIZACION DE VARIABLES
      3. CONCEPTUALIZACION DE VARIABLES
   6. JUSTIFICACIÓN
      1. TÉCNICA
      2. ECONÓMICA
      3. JUSTIFICACIÓN SOCIAL
   7. METODOLOGÍA
      1. MÉTODO CIENTÍFICO
      2. INGENIERIA DE SOFTWARE
   8. HERRAMIENTAS
   9. LÍMITES Y ALCANCES
      1. LIMITES
      2. ALCANCES
   10. APORTES
   11. ESQUEMA
   12. CRONOGRAMA DE EJECUCIÓN

2. MARCO TEÓRICO

2.1. MINERÍA DE DATOS

2.1.1. EL PROCESO DE MINERÍA DE DATOS

2.2. MINERÍA DE DATOS DISTRIBUIDA

2.2.1. ARQUITECTURA DE MDD

2.3. TÉCNICAS DE MODELADO EN MINERÍA DE DATOS

2.3.1. CLASIFICACIÓN

2.3.2. ANÁLISIS DE DEPENDENCIAS

3. MARCO APLICATIVO

3.1. ANÁLISIS DEL CASO DE ESTUDIO

3.2. MODELO DIMENSIONAL

3.3. RESULTADOS

4. ANÁLISIS DE COSTOS

5.1. CONCLUSIONES

5.2. RECOMENDACIONES

BIBLIOGRAFÍA

* Oded Maimon and Lior Rokach (2010). Data Mining and Knowledge Discovery Handbook. Springer, New York.
* https://es.wikipedia.org/wiki/Java\_ (lenguaje\_de\_programación)