



BENÉMERITA UNIVERSIDAD AUTÓNOMA DE
PUEBLA

FACULTAD DE CIENCIAS DE LA COMPUTACIÓN

Herramientas Estadísticas Para El Cálculo Del Pronóstico
De La Demanda Máxima De Potencia Eléctrica.

TESIS PROFESIONAL

Para Obtener El Título De:

Lic. En Ciencias De La Computación

Presenta: MARÍA MERCEDES MIRÓN FLORES

Asesores:

Dr. Manuel Martín Ortiz

MC. Jorge Ocampo Jiménez

Puebla, Pue. Marzo de 2013

Agradezco el apoyo de los profesores que me apoyaron en la elaboración de esta tesis, además de reconocer el apoyo de CONACYT por la beca durante la elaboración de la misma.

Resumen

La Comisión Federal de Electricidad (CFE) es una empresa que se encarga de la distribución de energía eléctrica, por ello es de suma importancia conocer la cantidad de energía eléctrica que se deberá suministrar en un lugar y tiempo determinado, actualmente tiene un sistema que le permite obtener el pronóstico de la Demanda Máxima de Potencia Eléctrica (DMPE) sin embargo este resulta poco certero. En consecuencia CFE requiere de un sistema en plataforma web que permita obtener el pronóstico de la DMPE que sea confiable.

Para determinar cuál es la relación de las variables de entrada con la respecto a la variable pronosticada DMPE se obtiene la correlación entre todas ellas.

La solución propuesta es la utilización de algunos modelos de aprendizaje computacional para la predicción de la DMPE a un determinado lapso de tiempo, para ello se requiere conocer los valores históricos de las variables de entrada (variables de energía, económicas, y de usuario) que son aquellas que influyen en el pronóstico. Posteriormente los métodos estadísticos obtienen el pronóstico de la DMPE utilizando combinaciones matemáticas de los valores predichos por modelos de aprendizaje.

Este trabajo de tesis está enfocado en implementar los métodos estadísticos necesarios para la obtención del pronóstico de la demanda máxima de potencia eléctrica, además de los métodos necesarios para determinar la relación entre variables, para elaborarlo se seguirá la metodología ágil de desarrollo Método de Desarrollo de Sistemas Dinámicos (DSDM).

Contenido

CAPÍTULO 1. SISTEMAS DE GENERACIÓN Y DISTRIBUCIÓN ELÉCTRICA	7
Centrales termoeléctricas	7
Centrales hidroeléctricas	7
Centrales eólicas	7
Distribución de la energía eléctrica.....	8
Parámetros De Un Sistema De Distribución.....	11
Planeación Del Sistema De Distribución	13
Distribución De La Energía Eléctrica En México	15
Mercado eléctrico de distribución	16
Importancia Del Pronóstico De La Demanda Eléctrica	17
CAPITULO 2. PRONÓSTICOS Y SU APLICACIÓN EN LA ELECTRICIDAD.....	18
Métodos cualitativos	20
Método Delphi.....	20
Investigación de Mercados	20
Analogía Histórica.....	20
Métodos cuantitativos.....	20
Modelos de series de tiempo.....	20
Modelos causales o explicativos	25
Análisis de regresión	25
Modelos econométricos	26
Modelos insumo-producto	26
Medidas De Precisión Del Pronóstico	27
Estado Del Arte	28
CAPÍTULO 3. METODOLOGÍAS ÁGILES	30
Programación Extrema (XP)	31
SCRUM	31
Desarrollo de Software Adaptable (ASD)	32
Método de Desarrollo de Sistemas Dinámicos (DSDM)	33
CAPÍTULO 4. MARCO TEÓRICO DE HERRAMIENTAS ESTADÍSTICAS.....	36

Métodos no paramétricos	36
Coeficientes De Correlación	37
Prueba Kruskal Wallis.....	37p
Prueba De Fisher	38
Distribución Weibull	39
Intervalos De Confianza	41
CAPÍTULO 5. PLANTEAMIENTO DEL PROBLEMA Y ANÁLISIS	42
Planteamiento del problema	42
CAPÍTULO 6. DISEÑO Y ARQUITECTURA.....	57
Diseño.....	57
Arquitectura.....	64
CAPÍTULO 7. PRUEBAS Y RESULTADOS.....	66
Pruebas	66
Resultados	66
CONCLUSIONES.....	74
Trabajo a futuro	74
REFERENCIAS	75
Apéndice A. Reglas de Usabilidad	77

CAPÍTULO 1. SISTEMAS DE GENERACIÓN Y DISTRIBUCIÓN ELÉCTRICA

La corriente eléctrica es un fenómeno causado por el movimiento de los electrones libres de los átomos en un material conductor (material que permite el paso de la electricidad y oponen poca resistencia).

Un sistema de suministro eléctrico comprende la generación, el transporte y la distribución de la energía eléctrica.

Existen diversas formas de producir la energía eléctrica, sin embargo todas consisten en convertir otro tipo de energía en energía eléctrica, a continuación se mencionan algunas:

Centrales termoeléctricas

Producen energía eléctrica haciendo girar los generadores eléctricos, esto se puede hacer de diversas formas utilizando por ejemplo el vapor de agua es conducido a turbinas enlazada al generador para producir energía eléctrica.

Centrales hidroeléctricas

Consiste en la transformación de la energía producida por la caída de agua en energía mecánica que esta hace girar turbinas que enlazadas con un generador producen la energía eléctrica.

Centrales eólicas

Obtiene la energía eléctrica al convertir mediante una turbina la energía que porta el viento.

Después de ser generada la energía eléctrica debe ser transportada a los centros de consumo por medio de la red de transporte, ésta está compuesta por:

⇒ Torres

Soporte para las líneas de transmisión que con la ayuda de algunos materiales (botones de porcelana o silicón) evitan en contacto de las líneas de transmisión y las torres.

⇒ Líneas de transmisión

Dispositivos que se utilizan para transmitir la energía, se pueden utilizar diversos materiales (aluminio, cobre) para ello se debe tomar en cuenta las necesidades de la línea.

⇒ Subestaciones

Conjunto de elementos que interconectan las centrales generadoras de energía con las líneas de transmisión y son las encargadas de transformar el voltaje de la energía eléctrica generada a un voltaje adecuado para el consumo.

[ENRI02] clasifica las subestaciones en:

- ◆ Subestaciones elevadoras.
- ◆ Subestaciones reductoras.

- ◆ Subestaciones de enlace.
- ◆ Subestaciones en anillo.
- ◆ Subestaciones radiales.
- ◆ Subestaciones de switcheo.

Las subestaciones están compuestas por elementos primarios: transformador, interruptor de potencia, restaurador, cuchillas fusible, cuchillas desconectadoras, cuchillas de prueba, apartarrayos, tableros dúplex de control, condensadores, transformadores de instrumento y elementos secundarios: cables de potencia, cables de control, alumbrado, estructura, herrajes, equipo contra incendio, equipo de filtrado de aceite, sistemas de tierras, carrier, intercomunicación, trincheras, ductos, conductos, drenajes y cercas.

⇒ Equipos de protección, comunicaciones y control.

Distribución de la energía eléctrica

Para que la energía eléctrica llegue a los centros de consumo (hogares, industrias, etc.) las compañías eléctricas cuentan con un sistema de distribución. Un sistema de distribución es un conjunto de elementos que se encargan de conducir la energía eléctrica proveniente de una subestación para hacerla llegar al usuario, un aspecto importante a considerar es que dicha distribución debe ser ininterrumpida dado que el usuario no puede quedarse sin energía eléctrica.

Para la distribución de la energía eléctrica se debe tomar en cuenta los usuarios para suministrar el valor de la tensión adecuada para el correcto funcionamiento de sus aparatos eléctricos, las redes utilizadas para esta distribución deberán ser capaces de soportar cambios en el crecimiento de la carga.

Según [RAMIO4] Las redes de distribución se pueden clasificar en cuanto a su construcción en:

- Redes aéreas: son aquellas que se encuentran sostenidas por torres metálicas o postes y el material conductor no cuenta con protector alguno esto ocasiona que este tipo de redes sea más propenso a interrupciones por diversos factores como el clima. Sin embargo es la más utilizada por su bajo costo y su fácil mantenimiento.
- Redes subterráneas: son aquellas en las cuales los materiales conductores se encuentran enterrados y cubiertos por un aislante, aunque este tipo de redes es más seguro es menos utilizada dado que su costo es muy elevado además resulta difícil darles mantenimiento.

Clasificación en cuanto a voltajes nominales

- Redes de distribución secundarias: que va de 120 a 240 Volts.
- Redes de distribución primarias: desde 13.2 kV hasta 66 kV.

Clasificación en cuanto a la ubicación geográfica

- Redes de distribución urbana
- Redes de distribución rurales
- Redes de distribución suburbana
- Redes de distribución turística

Clasificación en cuanto al tipo de cargas

Carga es un conjunto de objetos o aparatos que consumen la energía eléctrica.

- Redes de distribución para cargas residenciales: comprende edificios de departamentos, condominios, residencias, entre otros.
- Redes de distribución para cargas comerciales: comprende a los que llevan a cabo actividades comerciales.
- Redes de distribución para cargas industriales: comprende a las industrias, esta red es muy importante dado que consume una considerable carga.
- Redes de distribución para cargas para alumbrado público: comprende a las lámparas que se encargan de brindar iluminación para la seguridad de la sociedad.

En cuanto a confiabilidad

- Cargas de primera categoría:
- Cargas de segunda categoría
- Cargas de tercera categoría

Un sistema de distribución consta de los siguientes elementos:

- Líneas de subtransmisión y distribución: las líneas subtransmisión operan entre 66 hasta 115 kV. Las líneas de distribución operan con niveles menores a 34.5 kV.
Según [CANT06] hay diversos tipos de estructuras para las líneas de distribución como:
 - Estructura de paso
 - Estructura de paso doble
 - Estructura semi-volada
 - Estructura de paso con cruceta volada
 - Estructura de paso doble cruceta semi-volada
 - Entre otras.

- Subestaciones de distribución: Su función consiste en modificar los valores de las líneas de subtransmisión a valores de distribución en mediana tensión, 13, 23 o 33 kV.
- Redes de distribución de media tensión: Son las líneas cuyo objetivo es conectar a los usuarios en mediana tensión a las subestaciones de distribución. También se les conoce como redes de distribución.
- Bancos de transformación: encargados de transformar los valores de media tensión a niveles utilizables para el consumo en baja tensión. Estos valores son 220/127 ó 240/120 V.
- Redes de distribución de baja tensión: líneas que unen los bancos de transformación con las acometidas de los usuarios en baja tensión.

Sin embargo en [CES98] y [RIV12] los componentes primarios de un sistema de distribución son:

- Alimentadores primarios de distribución
Elementos encargados de transportar la energía eléctrica desde las subestaciones hacia los transformadores. Sus componentes son: troncal y ramal.
- Transformadores de distribución:
Son los encargados de transformar la tensión de tal forma que pueda ser utilizada sin requerir de equipos o instalaciones especiales o peligrosas.
- Alimentadores secundarios
Su función consiste en distribuir la energía eléctrica desde los transformadores a las acometidas de los usuarios.
- Acometidas.
Encargadas de enlazar el sistema de distribución con las instalaciones del usuario, se pueden proporcionar a diferente tensión dependiendo de la carga que requiera el usuario.
- Equipo de medición
La medición se realiza dependiendo del tipo de acometida que el usuario necesite.

Los componentes secundarios del sistema de distribución:

- Cuchillas: permiten seleccionar, conectar o desconectar circuitos eléctricos que no tienen carga.
- Reactores: limitan la corriente que fluye en un circuito eléctrico bajo condiciones de cortocircuito.
- Interruptores: permiten abrir o cerrar circuitos, con o sin carga, o con corriente de falla también debe desconectar un elemento que se encuentre en cortocircuito.
- Capacitores: se oponen al cambio en la magnitud de tensión a través de un circuito.

- Fusibles: protegen a los elementos del sistema, interrumpen el paso de la corriente eléctrica cuando el amperaje pasa de niveles normales.
- Restauradores: dispositivos que permiten cerrar e interrumpir automáticamente circuitos con una secuencia de aperturas y cierres seguidos de una operación de cierre final.
- Seccionadores: encargados de abrir circuitos una vez que haya sido interrumpido un circuito por cortocircuito.

El sistema de distribución comprende diferentes niveles de tensión

- Tensión de seguridad: 24 voltios respecto a tierra
- Alta tensión: tensión por encima de 33,000 voltios
- Media tensión: tensiones entre 1,000 hasta 33,000 voltios
- Baja tensión: tensiones entre 50 hasta 1,000 voltios
- Muy baja tensión: tensión de hasta 50 voltios

A lo largo del tiempo una red de distribución puede presentar perturbaciones en la tensión estas pueden ser producidas durante la generación, la transmisión o distribución de la energía. Estas perturbaciones pueden ocasionar considerables pérdidas económicas dependiendo de la severidad de dichas perturbaciones, a continuación se mencionan algunas:

- Variaciones de tensión de larga duración.
- Sobretensión.
- Caídas de tensión.
- Variaciones de tensión de corta duración.
- Entre otros.

Como se mencionó anteriormente las cargas son aquellas que consumen energía eléctrica entonces cualquier variación en las cargas representa una alteración al comportamiento de los usuarios.

Parámetros De Un Sistema De Distribución.

Potencia

Velocidad a la que se produce y consume la energía eléctrica, se representa $P = J/\text{segundo}$, donde J es Joule. Su unidad de medida es el watt.

Carga Instalada

En [ACHU98] es la suma de las potencias máximas que demanda todas las máquina o aparato en condiciones de uso normales conectados a un sistema eléctrico sin importar que éste se encuentre en operación o no. Se expresa en kVA, kW.

Dependiendo del uso que le da usuario puede ser clasificada en:

- Cargas residenciales.
- Cargas de iluminación en predios comerciales.
- Cargas de fuerza en predios comerciales.
- Cargas industriales.
- Cargas gubernamentales.
- Cargas de hospitales.

Demanda

De acuerdo con [ENRI02] es el promedio de la carga eléctrica durante un lapso de tiempo.

Demanda Máxima

Es la demanda medida en kW (kilo Watts) durante cualquier periodo de tiempo, donde el consumo en ese periodo fue mayor al consumo en cualquier otro periodo.

Carga instalada

Es un conjunto de objetos o aparatos que consumen la energía eléctrica que se encuentran conectados a un sistema eléctrico ya sea que se encuentre o no se encuentre en funcionamiento.

Demanda Promedio

Es el promedio de la demanda en un determinado periodo de tiempo. Y se puede expresar:

$$Demanda Promedio = \frac{Energía consumida en un periodo de tiempo}{No. De horas del periodo}$$

Factor de carga

Relación entre la demanda promedio y la demanda máxima en un periodo de determinado. Y se expresa:

$$Factor de carga = \frac{demanda promedio}{demanda máxima}$$

Factor de demanda

Relación entre la demanda máxima y la carga instalada en un sistema eléctrico, su valor siempre debe ser menor a 1.

$$Factor de demanda = \frac{demanda máxima}{carga instalada}$$

Factor de utilización.

Es la relación entre la demanda máxima de un sistema y la capacidad normal del sistema.

$$\text{Factor de utilización} = \frac{\text{demanda máxima del sistema}}{\text{capacidad instalada}}$$

Factor de diversidad.

Es la relación entre la suma de las demandas máximas en distintas partes de un sistema o una instalación y la demanda máxima del sistema o instalación. Su valor siempre es mayor o igual a 1.

$$\text{Factor de diversidad} = \frac{\text{Suma de demandas máximas individuales}}{\text{demanda máxima del sistema}}$$

Factor de coincidencia.

Indica el grado de dispersión con que trabajan en el tiempo las cargas de un sistema. Su valor siempre es menor a 1.

$$\text{Factor de coincidencia} = \frac{1}{\text{factor de diversidad}}$$

Densidad de carga.

Indica la cantidad de carga que se encuentra en una superficie. Existen diferentes tipos:

$$\text{Densidad de carga} = \frac{\text{Carga instalada}}{\text{area de la zona}} = \frac{\text{kW}}{\text{km}^2}$$

Diversidad de carga.

Indica la diferencia entre la suma de las demandas máximas de las cargas individuales y la demanda máxima del grupo.

$$\text{Diversidad de carga} = \sum_i^n D_i - D$$

Donde $D = \sum_i^n C_i D_i$ siendo C_i las cargas y D_i las demandas.

Demanda no coincidente

Es la suma de todas las demandas máximas individuales sin tomar en cuenta el periodo de tiempo.

Planeación Del Sistema De Distribución

Existen diversos factores que se deben tomar en cuenta para el planteamiento de un sistema de distribución como son:

- Factores económicos ya que si no se cuenta con el capital necesario no podría expandir el sistema de distribución.

- Factores demográficos: para la visualización de las áreas que serán abandonadas por sus habitantes.
- Factores tecnológicos: ya que se puede visualizar el desarrollo de nuevas fuentes no convencionales.

En México existen 33 tarifas eléctricas, dichas tarifas son disposiciones que contienen cuotas y condiciones del suministro eléctrico. Se clasifican en:

- Domésticas.
- Servicio público
- Riego agrícola.
- Servicios generales, se dividen en alta, media y baja tensión.

Según [CANT06] en México la tensión de suministro eléctrico es la siguiente:

- Baja tensión: Menor a 1000 voltios.
- Mediana tensión: entre 1001 y 35000 voltios
- Alta tensión para nivel de subtransmisión: entre 35000 y 220000 voltios.
- Alta tensión para nivel de transmisión: voltaje mayor a 220000 voltios.

El consumo de energía es la energía eléctrica utilizada por una instalación en un lapso de tiempo y se puede clasificar en:

- Consumo base.
- Consumo intermedio.
- Consumo semipunta
- Consumo punta
- Consumo total.

Se debe adecuar al sistema eléctrico para poder atender a diferentes tipos de usuarios. Los usuarios por sector se clasifican en:

- Sector doméstico.
- Sector industrial.
- Sector comercial.
- Sector de servicios públicos.

Un problema de la energía eléctrica es que se debe producir la cantidad de energía que se consumirá ya que no hay forma de almacenarla.

Distribución De La Energía Eléctrica En México

En 1937 fue creada la Comisión Federal de Electricidad que es la empresa de gobierno encargada de generar, transmitir y distribuir la energía eléctrica en gran parte de México.

Para ello cuenta con más de 750 mil kilómetros de líneas de transmisión y distribución.

La CFE divide al país en 16 divisiones de distribución: Baja California, Noreste, Norte, Noroeste, Golfo Centro, Bajío, Jalisco, Centro Occidente, Centro Sur, Centro Oriente, Oriente, Sureste, Peninsular, Valle de México Centro, Valle de México Norte y Valle de México Sur.



Figura 1. Divisiones de distribución de CFE

Cada división de distribución se divide en zonas de distribución.

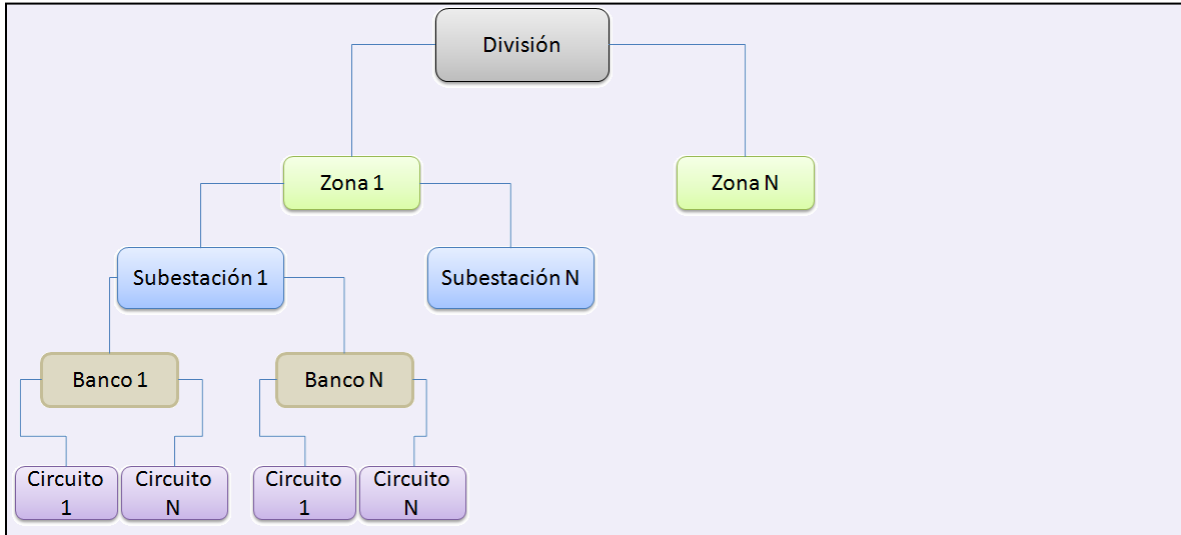


Figura 2. Elementos que forman una división de distribución

Mercado eléctrico de distribución

La CFE realiza anualmente un estudio del comportamiento del mercado eléctrico, dicho estudio se realiza a nivel de circuito de distribución este permite identificar el crecimiento de carga a nivel geográfico. La información que es considerada para este estudio son las demandas máximas por circuitos primarios y subestaciones de distribución.

Conforme va en aumento la población se va incrementando el consumo de la energía eléctrica.

El mercado eléctrico consiste en la comercialización con potencia y energía eléctrica, se muestra el organigrama del mercado eléctrico de CFE.

CAPITULO 2. PRONÓSTICOS Y SU APLICACIÓN EN LA ELECTRICIDAD

El pronóstico es una predicción o estimación de lo que puede ocurrir a futuro que se lleva a cabo a partir del análisis de conocimientos científicos y de los conocimientos empíricos. La predicción que se realiza es para un punto o periodo en un futuro, cuando se pronostica se debe tomar en cuenta que el pronóstico puede tener incertidumbre dado que no se tiene la certeza de las circunstancias que se pueden presentar en un tiempo dado.

Barry Render sugiere las siguientes formas de clasificar los pronósticos:

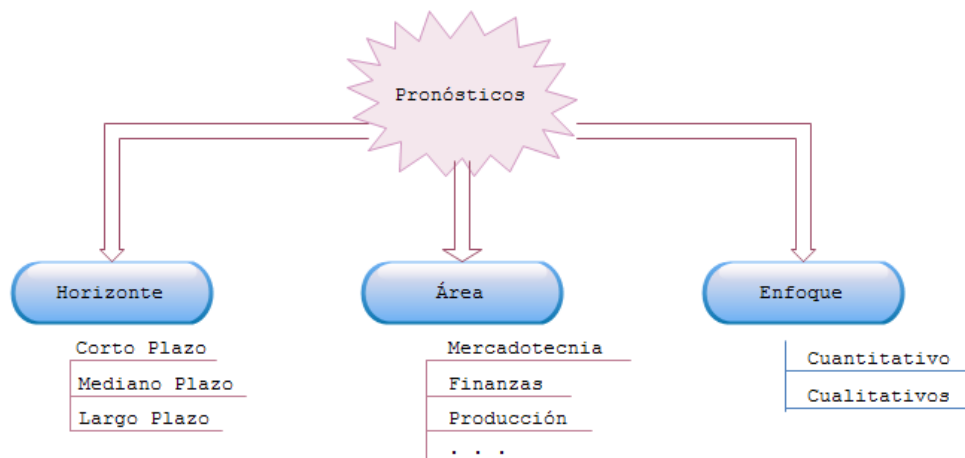


Figura 4. Clasificación de los pronósticos.

Respecto al horizonte de tiempo

- Pronóstico a corto plazo
Proporciona el pronóstico de un periodo de 0 a 3 meses.
- Pronóstico a mediano plazo
Proporciona el pronóstico de un periodo de 3 meses a 3 años.
- Pronóstico a largo plazo
Proporciona el pronóstico de un lapso de 3 a 5 años.

Respecto al área

- Mercadotecnia
Obtener una estimación de las ventas de un producto en determinado periodo de tiempo, para así preparar adecuadamente un plan de comercialización.
- Producción
Sirven como apoyo para la toma de decisiones para la planeación de productos.

- Finanzas
Pronostican inversiones a mediano y a largo plazo de una organización. Los pronósticos para esta área dan a conocer el valor monetario que se necesitara durante un periodo de tiempo por una organización.
- Recursos humanos
Pronostican el número de empleados que requerirá una organización de un periodo de tiempo, así mismo como las características que deberán cumplir dichos empleados.

Respecto al enfoque a pronosticar

- Pronósticos cualitativos
Son aquellos que por medios de la experiencia o intuición de los expertos proporcionan el pronóstico, utilizadas cuando se tienen pocos datos, se transforma la información cualitativa en estimados cuantitativos. Estas técnicas son utilizadas con mayor frecuencia en pronósticos de mediano y largo plazo, también se le conoce como técnicas subjetivas.
- Pronósticos cuantitativos
Son modelos matemáticos utilizados para analizar datos históricos, también se les conoce como técnicas objetivas.

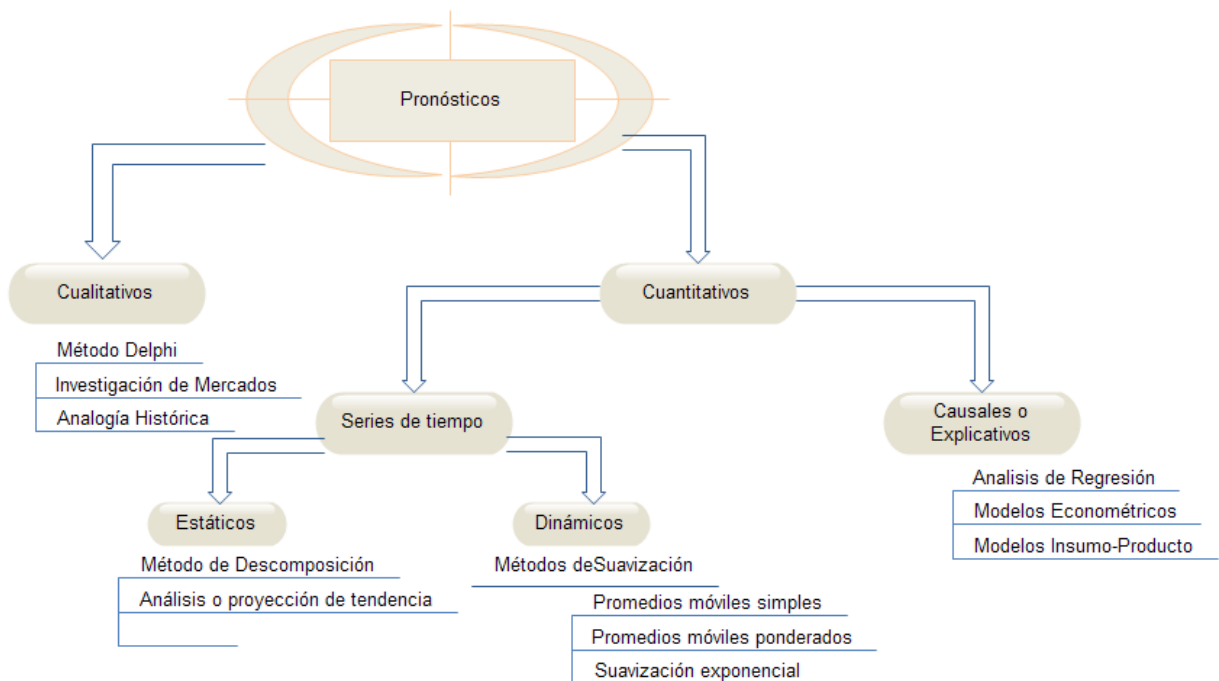


Figura 5. Clasificación de pronósticos de acuerdo al enfoque

Métodos cualitativos

Existen diversos métodos cualitativos entre los cuales destacan:

Método Delphi

Proceso mediante el cual se obtiene la opinión de un grupo de expertos sobre el futuro y se llega a un consenso. Tiene las siguientes características: el anonimato de los expertos, retroalimentación controlada y agregación de respuestas del grupo.

Este procedimiento no está diseñado para competir o desafiar procedimientos estadísticos o basados en modelos ya que este carece de datos históricos.

Investigación de Mercados

Este método permite la recolección y el análisis de información o datos relacionados con actividades relacionadas con la comercialización. Una investigación debe ser ordenada, emplear métodos científicos adecuados es decir que no deberá emplear razonamientos basados en prejuicios.

Analogía Histórica.

Este método identifica el producto o servicio del cual se obtendrá el pronóstico y liga la estimación de las ventas de dicho producto o prestación de servicios con su historia pasada de un producto o servicio similar.

Métodos cuantitativos

Se clasifican en las siguientes categorías:

Modelos de series de tiempo

Se le llama serie de tiempo a un conjunto de datos numéricos observados secuencialmente en intervalos de tiempo fijo.

El principal objetivo del análisis de las series de tiempo es descubrir un patrón en los datos históricos y así deducir los valores de la serie a futuro.

Estos modelos constan de cuatro componentes:

- Tendencia: Representa el cambio o movimiento a valores mayores o menores a lo largo de un periodo de tiempo de la serie.
- Cíclico: Es un movimiento cíclico solo si se produce en un intervalo de tiempo mayor a un año.
- Estacionalidad: Representa los efectos producidos por fenómenos que se repiten cada año con cierta constancia.

- Aleatoriedad: Caracteriza los movimientos imprevisibles.

Se conocen los siguientes modelos de series de tiempo:

- ⇒ Modelos Aditivo

Este modelo asume que el valor de la serie de Y es la suma de los cuatro componentes.

$$Y = \text{tendencia} + \text{estacionalidad} + \text{ciclicidad} + \text{aleatoriedad}.$$

- ⇒ Modelo Multiplicativo

Este modelo asume que el valor de la serie de Y es el producto de los cuatro componentes.

$$Y = \text{tendencia} * \text{estacionalidad} * \text{ciclicidad} * \text{aleatoriedad}.$$

Métodos de descomposición

Según [PARE03] el método clásico de descomposición descompone la serie de tiempo en tendencia, estacionalidad, ciclos y error.

$$\text{dato} = \text{funcion}(\text{estacionalidad}, \text{tendencia}, \text{ciclicidad}) + \text{error}$$

o bien

$$x_t = f(E_t, T_t, C_t) + Er_t$$

donde

x_t : El dato de la serie de tiempo en el periodo t.

E_t : La estacionalidad del dato x en el periodo t

T_t : La tendencia del dato x en el periodo t

C_t : El ciclo del dato x en el periodo t

Er_t : Error en el periodo t

Existen diversos modelos de descomposición como son:

- ⇒ Modelo multiplicativo $x_t = E_t * T_t * C_t * Er_t$

- ⇒ Modelo aditivo $x_t = E_t + T_t + C_t + Er_t$

Análisis o proyección de tendencia

Este método ajusta una línea de tendencia a una serie de puntos de datos históricos y posteriormente proyecta la línea hacia el futuro para pronosticar, para llevar a cabo pronósticos a mediano y largo plazo, para poder emplear este método la serie debe de carecer de un componente estacional. Existen diversos modelos que permiten encontrar la recta como son: modelos lineales, cuadráticos, exponenciales y de curva - S.

Promedios móviles simples

De acuerdo con R. Anderson consiste en tomar los n valores de los datos más recientes en la serie de tiempo, promediarlos y así obtener el pronóstico para el siguiente periodo. Se expresa de la siguiente forma:

$$\text{Promedio movil} = \frac{\sum(\text{valores de } n \text{ datos más recientes})}{n}$$

Se le denomina promedio móvil ya que conforme se dispone de nuevas observaciones en la serie de tiempo se utilizan los n valores de los datos recientes y se calcula el nuevo pronóstico. Para aplicar este método la serie de tiempo debe tener un patrón estacionario, se definirá la cantidad de datos de la serie de tiempo que se incluirán al promedio móvil y el cálculo del primer pronóstico no se realiza mediante la fórmula, es necesario efectuar el cómputo completo del promedio.

Promedios móviles ponderados

De acuerdo con [ANDE04] consiste en asignar un factor de ponderación distinto a cada valor de los n datos más recientes serie de tiempo, donde la suma de los factores de ponderación debe ser igual a 1, generalmente se le asigna mayor peso a los valores más recientes y menor a los valores más antiguos. Posteriormente se calcula el promedio ponderado de dichos valores. Expresión matemática:

$$\text{Promedio movil ponderado} = \frac{\sum(\text{peso para el valor } n \text{ de la serie de tiempo})(\text{valor } n \text{ de la serie de tiempo})}{\sum \text{pesos}}$$

Metodología Box-Jenkins

Esta metodología fue desarrollada por George Box y Gwilym Jenkins y consiste en proponer un modelo matemático que proporcione la mejor descripción de los datos de la serie y permita hacer las predicciones conociendo el periodo de tiempo, también es conocido como modelos integrados (I) autorregresivos (AR) y de medias móviles (MA). La notación ARIMA (p, d, q), donde p son términos autorregresivos, d es el número de diferenciaciones para que la serie sea estacionaria y q son términos de media móvil. También se encuentra la notación extendida para los modelos ARMA (p, q). El proceso consiste en identificar el proceso estocástico, estimar los parámetros del modelo, verificar que se cumplen las hipótesis de partida y para terminar se realizan los pronósticos.

Series de tiempo Shiskin

En 1954 Julius Shiskin presentó el método Census I para realizar ajustes estacionales por medio de computadoras; uno de sus principales objetivos fue aumentar la cantidad de series de tiempo que son ajustadas, posteriormente se fueron realizando mejoras a este método.

Suavización exponencial

Según Barry Render este método consiste en utilizar promedios ponderados de valores pasados como pronósticos, es similar a promedios móviles ponderados difiere en que solo se selecciona el factor de ponderación del valor más reciente y los demás factores con calculados de manera automática, el valor de los factores de ponderación se van haciendo más pequeños según las observaciones se alejan en el pasado. Para aplicar este método la serie de tiempo deberá tener un patrón estacionario. Se expresa matemáticamente de la siguiente forma:

$$F_t = F_{t-1} + \alpha (Y_{t-1} - F_{t-1})$$

donde:

F_t = Nuevo pronóstico (del periodo t)

F_{t-1} = Pronóstico anterior (del periodo t-1)

α = Constante de suavizamiento ($0 \leq \alpha \leq 1$)

Y_{t-1} = Demanda real del periodo anterior

Método lineal de Holt

Desarrollado por Charles C. Holt en 1957, este método asume que la serie tiene una tendencia lineal, suaviza el nivel y la pendiente utilizando constantes para el suavizamiento de cada una, el problema de este método es que al seleccionar los coeficientes para el suavizamiento no todos los valores resultan ser óptimos. También se le conoce como suavizamiento exponencial doble. Se puede expresar de la siguiente manera:

El estimado del nivel $L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$

El estimado de tendencia $T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$

Para obtener el pronóstico para cualquier instante de tiempo posterior:

$$\hat{Y}_{t+p} = L_t + pT_t$$

donde:

α = Constante de suavizamiento para el nivel $(0 < \alpha < 1)$

β = Constante de suavizamiento para la tendencia $(0 < \beta < 1)$

L_t = Estimación del nivel en el periodo t

T_t = Estimación de la tendencia en el periodo t

Y_t = Valor de la serie de tiempo en el periodo t

p = Periodo del futuro a pronosticarse

\hat{Y}_{t+p} = Pronóstico para el periodo p en el futuro

Método de Holt - Winters

Este método toma en cuenta la tendencia dado que es una extensión del método de suavizado exponencial simple (Holt) y la estacionalidad que es la extensión por Winters.

Entonces está formado por los siguientes componentes: nivel, estacionalidad y tendencia; también tienen α , β y γ como parámetros fijados por el usuario. Este método tiene dos modelos el modelo aditivo y el modelo multiplicativo.

Modelo Multiplicativo:

$$L_t = \alpha \left(\frac{Y_t}{S_{t-p}} \right) + (1 - \alpha)[L_{t-1} + T_{t-1}]$$

$$T_t = \gamma[L_t - L_{t-1}] + (1 - \gamma)T_{t-1}$$

$$S_t = \beta \left(\frac{Y_t}{L_t} \right) + (1 - \beta)S_{t-p}$$

$$\hat{Y}_t = (L_{t-1} + T_{t-1})S_{t-p}$$

Modelo Aditivo:

$$L_t = \alpha(Y_t - S_{t-p}) + (1 - \alpha)[L_{t-1} + T_{t-1}]$$

$$T_t = \gamma[L_t - L_{t-1}] + (1 - \gamma)T_{t-1}$$

$$S_t = \beta(Y_t - L_t) + (1 - \beta)S_{t-p}$$

$$\hat{Y}_t = L_{t-1} + T_{t-1} + S_{t-p}$$

donde:

L_t = Es el nivel en un tiempo t , α es el peso para el nivel

T_t = Es la tendencia en un tiempo t , γ es el peso para la tendencia

S_t = Es la estacionalidad en un tiempo t , β es el peso para el componente estacional.

p = Es el periodo estacional

Y_t = Es el valor de la serie de tiempo en el tiempo t

\hat{Y}_t = Es el valor ajustado, o un periodo de predicción en un tiempo t

Modelos causales o explicativos

Estos modelos pronostican el valor de una variable conociendo los valores de una o más variables.

A la variable que se desea pronosticar o predecir se le denomina variable dependiente, las variables que se utilizan para realizar dicha predicción se les llama variables independientes.

Para utilizar este modelo se deben hallar las relaciones matemáticas entre las variables independientes y dependientes, deben estar disponibles los valores de las variables para poder realizar en cualquier momento el pronóstico.

Según [EPPE00] supone que y indica el valor real de una variable de interés para el pronóstico, y que \hat{y} sea el valor pronosticado de esa variable. Entonces en un modelo causal se puede expresar:

$$\hat{y} = f(x_1, x_2, \dots, x_{n-1}, x_n)$$

Sea

f : una regla o función de pronóstico.

$x_1, x_2, \dots, x_{n-1}, x_n$: Conjunto de variables.

\hat{y} : Variable dependiente o de respuesta.

Existen diversos métodos causales que permiten crear un modelo de pronóstico entre los cuales se encuentran:

Análisis de regresión

Se le atribuye a Francis Galton el primer trabajo de regresión posteriormente Karl Pearson confirmó la Ley de regresión universal de Galton.

Este modelo permite pronosticar el valor de una variable a partir de los valores de otras variables. A la variable que se desea pronosticar o predecir se le denomina variable dependiente, las variables que se utilizan para realizar dicha predicción se les llama variables

independientes. Este método describe la relación matemática existente entre variables dependientes e independientes.

Regresión lineal simple: se utiliza el valor de la variable independiente x para pronosticar el valor de la variable Y .

$$Y = f(x)$$

Regresión lineal múltiple: permite utilizar más de una variable independiente para pronosticar el valor de una variable dependiente.

$$Y = f(x, w, z)$$

Modelos econométricos

De acuerdo con [CARI98] la econometría es una rama de la economía que utiliza métodos y técnicas de la estadística matemática para relaciones económicas. No hay un acuerdo general del nacimiento de la econometría, sin embargo la opinión más extendida vincula su origen al momento en que se manifiesta como movimiento organizado, concretamente, la constitución de la Econometrics Society en 1930.

Un modelo econométrico es la representación matemática de un conjunto de relaciones económicas reales. Un modelo econométrico consta de: variables, parámetros y relaciones. Se clasifican de distintas formas tomando en cuenta diversos criterios.

- De acuerdo al número de ecuaciones: modelo uniecuacional (una ecuación), modelo multiecuacional (una o más ecuaciones).
- De acuerdo a la forma de las ecuaciones: modelo lineal y modelo no lineal.
- De acuerdo con el número de variables endógenas y ecuaciones: completo si el número de ecuaciones es igual al número de variables endógenas (variables que se explican dentro del modelo) en otro caso es no completo.

Modelos insumo-producto

Este modelo fue desarrollado por Wassily Leontief está formado por un sistema de ecuaciones lineales que muestran la relación entre bienes y servicios de un espacio económico determinado. Dicho sistema también es conocido como matriz de insumo-producto. Estas matrices proporcionan un análisis del proceso de producción de bienes y utilización de bienes y servicios.

Medidas De Precisión Del Pronóstico

Un factor importante a considerar en los pronósticos es la precisión, es decir que tan exacto es un pronóstico, el error se obtiene del valor pronosticado y el valor real. Algunas medidas de error son las siguientes:

Error

$$E_t = Y_t - X_t$$

Y_t : es el valor de la serie de tiempo para el periodo t

X_t : es el valor pronosticado para el periodo t

Cuadrado Medio del Error (CME)

$$CME = \frac{\sum(Y_t - \hat{Y}_t)^2}{n}$$

donde:

Y_t : es el valor de la serie de tiempo para el periodo t

\hat{Y}_t : es el valor pronosticado para el periodo t

n : número de elementos de la serie.

Error cuadrático medio (MSE, Mean squared error)

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$$

Raíz del error cuadrático medio (RMSE, Root mean squared error)

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$

Error porcentual absoluto medio (MAPE, Mean absolute porcentaje error)

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}$$

Error porcentual medio (MPE, Mean porcentaje error)

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t}$$

Error absoluto medio (MAE, Mean absolute error)

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$

El valor del error puede ser positivo o negativo dependiendo si el pronóstico es demasiado bajo o demasiado alto.

Estado Del Arte

Desde las antiguas civilizaciones han utilizado diversos métodos para conocer el futuro, la mayoría de personas que realizaban esta actividad la hacían adivinando. Pero fue hasta principios del siglo XX cuando se realizaron formalmente trabajos que permiten hacer predicciones, en el transcurso de los años estos trabajos se han ido mejorando.

Según [BARR07] en ocasiones la información para realizar el pronóstico tiene un comportamiento complejo para poder ser modelada con métodos paramétricos, se realizó un análisis exploratorio el cual sugirió que existe una curva típica diaria de demanda por esta razón se optó por utilizar un modelo de regresión no paramétrica usando suavización spline para modelar y pronosticar la demanda horaria de energía eléctrica en una región del suroccidente de Colombia, tomando en cuenta que existen diversos factores que afectan el pronóstico. Se utilizó el lenguaje R (lenguaje para análisis estadístico) para realizar todos los cálculos.

En [ROJ] la predicción de la demanda de la energía eléctrica es muy importante para diferentes sectores como el sector industrial, comercial y habitacional, por la misma situación es que se deben tomar en cuenta diferentes variables que son utilizadas como entradas para un sistema de predicción de demanda de energía eléctrica utilizando redes neuronales artificiales. Las redes neuronales dado que son modelos matemáticos que permiten emular el procesamiento de la información realizado por el cerebro. El modelo de validación se realiza comparando los resultados de los modelos de regresión lineal y datos reales con el análisis de varianzas.

En [FERN07] se crearon cinco nuevos métodos avanzados para la predicción a corto plazo de la producción de energía eléctrica en parques eólicos, para los modelos de predicción se debe seleccionar adecuadamente las variables esto es porque dado una variable con una elevada correlación con la variable a predecir puede mejorar el resultado del modelo de predicción, en caso contrario si la correlación es muy baja entonces se realentará el modelo. Los modelos generados fueron los siguientes

- ☞ Correspondientes a series temporales: Modelos de predicción basada en descomposición de Fourier, Modelo de predicción basado series temporales fuzzy y Modelo de predicción basado en regresión lineal.
- ☞ Correspondientes a los que utilizan variables meteorológicas: Modelo de predicción con información limitada del modelo atmosférico y Modelo de predicción con información completa del modelo atmosférico.

En [VELA12] utiliza redes neuronales autorregresivas (ARNN), este modelo está compuesto por un modelo lineal autorregresivo con un perceptrón multicapa con una capa oculta. En el modelo ARRN la variable dependiente (variable a pronosticar) es obtenida después como una función no lineal de los valores pasados (valores conocidos de las variables). Se realizó una comparación entre el desempeño de un modelo ARIMA, un perceptrón multicapa y una red neuronal autorregresiva para general el pronóstico mensual de la electricidad de Colombia, los resultados que se obtuvieron de esta comparación es que la red neuronal autorregresiva es capaz de pronosticar con mayor precisión la demanda.

Un mapa auto-organizado (SOM) es un tipo de red neuronal artificial que se entrena mediante algoritmos de aprendizaje no supervisado [SENA], su objetivo es obtener la curva de demanda de energía eléctrica del día siguiente, día a predecir. Para validar los resultados obtenidos por dicho modelo se obtiene el porcentaje de error medio absoluto (MAPE) y así poder medir la exactitud de la curva de la demanda pronosticada con la demanda real.

CAPÍTULO 3. METODOLOGÍAS ÁGILES

Las metodologías tradicionales para el desarrollo de software tienen mayor énfasis en la definición de roles, de actividades, la arquitectura del software y una documentación detallada. Aunque estas metodologías son efectivas para algunos proyectos, no lo son para otros por ello surgieron las “metodologías ágiles” que son más livianas están centradas en la colaboración del cliente y el desarrollo incremental del software es decir, que deben responder a cambios que se presenten durante el desarrollo del software. La filosofía del desarrollo ágil se resume en el documento “El Manifiesto Ágil”.

Los principios del manifiesto ágil son:

- Satisfacer al cliente mediante la entrega temprana y continua de software con valor.
- Aceptar que los requisitos cambien, incluso en etapas tardías del desarrollo. Los procesos Ágiles aprovechan el cambio para proporcionar ventaja competitiva al cliente.
- Entregar frecuentemente software funcional, desde un par de semanas a un par de meses, con preferencia al periodo de tiempo más corto posible.
- Los responsables de negocio y los desarrolladores trabajan juntos de forma cotidiana durante todo el proyecto.
- Los proyectos son desarrollados por profesionales motivados. Para ello se les da el entorno y el apoyo que necesitan además se les confía la ejecución del trabajo.
- El método más eficiente y efectivo de comunicar información al equipo de desarrollo y entre sus miembros es la conversación cara a cara.
- El software funcionando es la medida principal de progreso.
- Los procesos Ágiles promueven el desarrollo sostenible. Los promotores, desarrolladores y usuarios debemos ser capaces de mantener un ritmo constante de forma indefinida.
- La atención continua a la excelencia técnica y al buen diseño mejora la Agilidad.
- La simplicidad o el arte de maximizar la cantidad de trabajo no realizado, es esencial.
- Las mejores arquitecturas, requisitos y diseños emergen de equipos auto-organizados.
- A intervalos regulares el equipo reflexiona sobre cómo ser más efectivo para a continuación ajustar y perfeccionar su comportamiento en consecuencia.

En la actualidad existen diversas metodologías ágiles entre las cuales se encuentran:

Programación Extrema (XP)

Uno de los creadores de esta metodología es Kent Beck. Esta metodología considera que las bases para éxito en el desarrollo del proyecto es la interacción entre el cliente y el equipo de desarrollo, la interacción entre todos los participantes del proyecto.

Según Sommerville los requerimientos se expresan como escenarios, que se implementan como una serie de tareas, los desarrolladores trabajan en parejas.

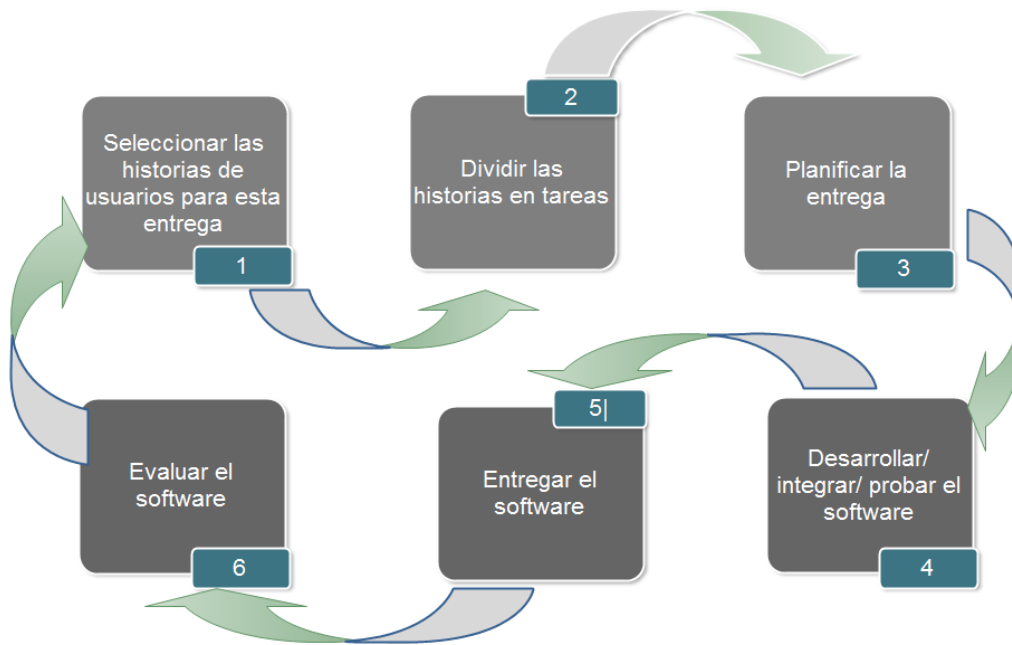


Figura 6. Ciclo de vida de la metodología Programación Extrema

En Ferrer en 2003 los principios fundamentales de la programación extrema son:

- Retroalimentación veloz
- Modificaciones incrementales
- Trabajo de calidad
- Asunción de simplicidad

Es apropiada cuando los requisitos del software cambian constantemente y existe un alto riesgo técnico.

SCRUM

Es una metodología iterativa e incremental, con ello se controlan los riesgos. Está compuesto por: equipos Scrum con sus respectivos roles, los bloques de tiempo, los artefactos

y las reglas. La parte fundamental de Scrum es el Sprint que es una iteración del producto en un lapso de tiempo.

Cada equipo Scrum debe tener los siguientes roles que realizan actividades diferentes: ScrumMaster, Propietario del producto y el equipo.

Los elementos basados en bloques de tiempo son: reunión de planificación de entrega, reunión de planificación de Sprint, el Sprint, el Scrum Diario, revisión del Scrum y retrospectiva del Sprint.

Los principales artefactos son: product backlog, sprint backlog, burndown de versión y sprint burndown.

Las reglas permiten unir los bloques de tiempo, los roles y los artefactos.

De acuerdo con Priolo es un proceso para el desarrollo de software que se enfoca al trabajo en equipo además de la aplicación de mejores prácticas para poder así obtener los resultados esperados del proyecto. Al igual que otras metodologías las ideas principales de Scrum son entregas periódicas o interacciones cortas con el objetivo de aceptar los cambios, la confección de equipos capacitados y la interacción con el cliente.

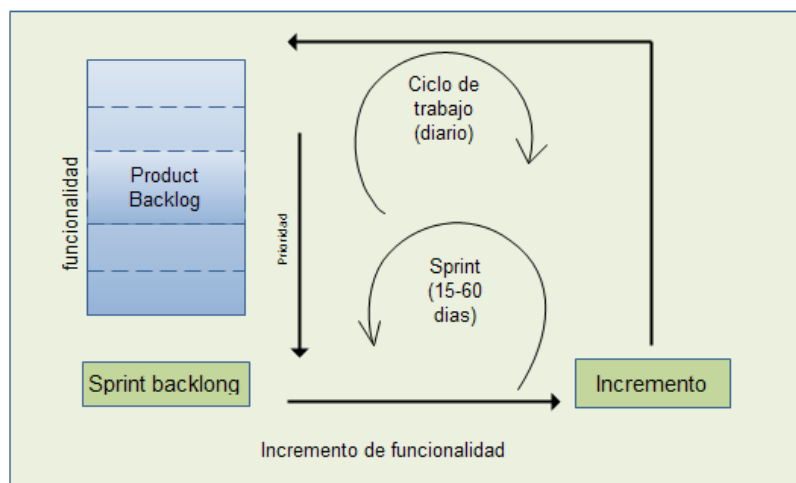


Figura 7. Ciclo de Vida de Scrum

Scrum es una metodología que no solo se puede aplicar en el desarrollo de software sino que se puede aplicar en cualquier área que tenga equipos de personas trabajando.

Desarrollo de Software Adaptable (ASD)

Fue desarrollada por Jim Highsmith y Sam Bayer, es iterativa, tolerante a los constantes cambios. Las fases que conforman dicha metodología son:

Especulación

Establece los principales objetivos y las metas del proyecto, se identifican los riesgos que puede tener el proyecto, posteriormente se determinan las iteraciones necesarias para culminar el proyecto. Esta fase dividida en los siguientes pasos: inicio, fijación, determinación de la iteración, objetivos, asignación de funcionalidad.

Colaboración

Se construye la funcionalidad del proyecto, dado que es una metodología iterativa en cada iteración se colabora para que se logre la funcionalidad planificada. Se deben realizar críticas del proyecto así mismo como comentar los problemas que se tienen con el proyecto, otro aspecto importante de esta etapa es establecer relaciones interpersonales dado que con ello permite llegar a un acuerdo entre los participantes cuando se presente una circunstancia compleja.

Aprendizaje:

Los participantes revisan los componentes del software desarrollados para mejorar la calidad se archiva o documenta el progreso del proyecto para posteriormente si es necesario identificar mejoras del proyecto.

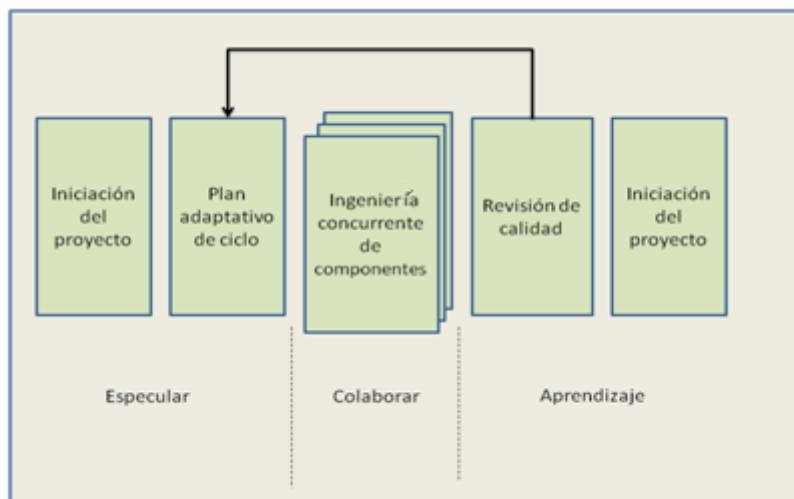


Figura 8. Ciclo de vida de ASD

Método de Desarrollo de Sistemas Dinámicos (DSDM)

Fue desarrollado en el Reino Unido en la década de los 90, tomando como fundamento las bases de RAD (Rapid Application Development).

Se han realizado diferentes versiones, en febrero de 1995 fue publicada la primer versión, pero ha sufrido diversos cambios, la que actualmente está en uso es la versión 4.2 publicada en abril del 2006.

Los principios en los que se basa son los nueve siguientes:

- La participación activa de los usuarios es imprescindible.
- Los miembros de los equipos de desarrollo deben tener la autonomía y poder para tomar las decisiones necesarias.
- Entrega periódica del producto.
- El principal criterio de prioridad, desarrollo y validación de las entregas incrementales son los propósitos del negocio.
- El desarrollo iterativo o incremental para satisfacer las necesidades del negocio a tiempo.
- Todos los cambios realizados en el desarrollo son reversibles.
- Los requisitos se establecen de manera general.
- Las pruebas forman parte del ciclo de desarrollo
- Cooperación y comunicación entre las partes interesadas del proyecto.

Ciclo de Vida:

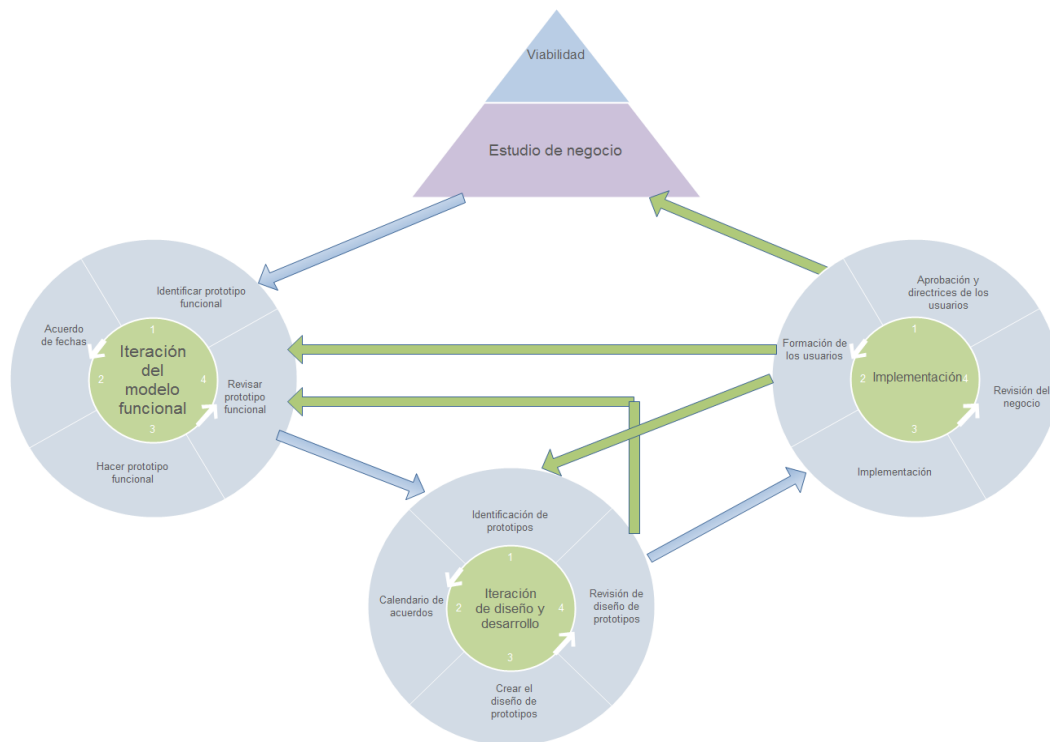


Figura 9. Ciclo de vida de DSDM

Estudio de viabilidad: se realiza un análisis de recursos humanos, materiales y financieros que se utilizarán para la realización del proyecto, así mismo se identifican los problemas que pueden causar un riesgo para la culminación del proyecto.

Estudio de negocio: se realiza un análisis para obtener los requisitos con mayor prioridad, para posteriormente determinar cuáles serán abordados primero.

Iteración funcional del modelo: se proponen modelos o esquemas que permitan dar una solución a los problemas, es decir se realiza el diseño del sistema.

Iteración de diseño y construcción: se lleva a cabo la creación de prototipos para posteriormente darle seguimiento.

Implementación: se realizan las pruebas pertinentes para verificar el correcto funcionamiento, se entrega el sistema al usuario final.

Esta metodología garantiza la viabilidad del proyecto antes de que se realice, además se asegura que los interesados en el proyecto tengan claro los diversos aspectos del proyecto para ello utiliza la creación de prototipos.

Su objetivo es el desarrollo rápido de aplicaciones.

CAPÍTULO 4. MARCO TEÓRICO DE HERRAMIENTAS ESTADÍSTICAS

La estadística inferencial utiliza datos de muestras para obtener mayor información de la población muestreada. Comprende diversos métodos estadísticos que permiten deducir como se distribuye la población bajo que se está analizando, nos enfocaremos en:

- Contraste de hipótesis
- Métodos no paramétricos

Contraste de hipótesis

También se le conoce como test de hipótesis. El contraste de hipótesis permite comprobar la veracidad de una hipótesis de una población, se distinguen dos tipos de hipótesis:

- Hipótesis nula es aquella negación de la hipótesis de investigación, se establece al inicio y cuya finalidad es anularla. Se denota por H_0 .
- Hipótesis alternativa es la contraria a la hipótesis nula

Se le atribuyen a Ronald Fisher algunos métodos de inferencia en las pruebas de contraste de hipótesis.

Métodos no paramétricos

También se les conoce como métodos de distribución libre. Estos métodos no dependen del conocimiento de cómo se distribuye la población, son utilizados para analizar datos cualitativos y cuantitativos. Estos métodos tienen diversas ventajas algunas de ellas son: se pueden utilizar en muestras pequeñas y se les puede aplicar a datos cualitativos. Aunque también presentan algunas desventajas entre ellas: no son muy eficientes y tienen una mayor probabilidad de no rechazar una hipótesis nula.

Los métodos no paramétricos en los que nos enfocaremos:

- Métodos de la Ji cuadrada o de la Chi cuadrada
- Coeficientes de correlación
- Prueba de Kruskal Wallis
- Prueba de la mediana
- Prueba X^2 de Pearson
- Prueba de Fisher

Coeficientes De Correlación

También se le conoce como coeficiente de correlación del producto-momento, fue desarrollado por Karl Pearson (1857- 1936). De acuerdo con Vargas Sabadías expresa la correspondencia o relación entre variables. Se expresa de la siguiente forma:

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

donde:

r_{xy} = Coeficiente de correlación

S_{xy} = Covarianza de (x, y)

S_x = Desviación estándar muestral de x

S_y = Desviación estándar muestral de y

Los valores del coeficiente de correlación están entre -1 y 1. Cuando el valor es 0 significa que no existe relación de dependencia lineal entre las variables, en caso que su valor sea 1 corresponde a una relación perfecta positiva, en el caso de que su valor sea -1 corresponde a una relación perfecta negativa entre las variables x y y .

Una relación perfecta positiva cuando conforme aumenta el valor de una variable aumenta el valor de la otra. La relación perfecta negativa ocurre cuando conforme disminuye el valor de una variable disminuye el valor de la otra.

Welkowitz et al (1981, 208) indica que el coeficiente de correlación posee las siguientes características:

- El valor cero indica ausencia de la relación lineal entre las variables.
- El valor numérico del coeficiente, independientemente del signo de la correlación, indica la fuerza de la relación.
- El signo indica la dirección de la relación.
- El valor máximo que puede tomar es 1 y el mínimo es -1.

Prueba Kruskal Wallis

Prueba propuesta por William Kruskal y W. Allen Wallis que es un método no paramétrico, es decir no dependen del conocimiento de la distribución de la población.

De acuerdo a Vargas Sabadías el procedimiento de la prueba Kruskal Wallis

Hipótesis a contrastar:

H0: las k muestras provienen de la misma población.

Ordenar las observaciones de las muestras en orden ascendente, asignarles el rango R_i correspondiente (1 para la observación más baja y N para la observación más alta, donde $N = N_1 + N_2 + \dots + N_k$), si hay observaciones repetidas se les asigna el rango medio. Se obtiene la suma de rangos por muestra i .

Se calcula la suma de cuadrados intragrupo de las muestras.

$$SCI = \sum_{i=1}^k \frac{1}{n_i} \left(R_i - \frac{n_i (n_T + 1)}{2} \right)$$

Se calcula en estadístico H de kruskal Wallis

$$H = \frac{\frac{12}{n_T(n_T + 1)} SCI}{1 - \frac{\sum_{j=1}^r (d_j^3 - d_j)}{n_T^3 - n}}$$

donde r : es el número de rangos con repetición.

d_j = Número de repeticiones del rango j -ésimo.

Nt = Total de observaciones

El estadístico H se distribuye aproximadamente como χ^2 con $k-1$ grados de libertad, por esta razón una vez obtenido el valor de H se debe comparar con $X_{\epsilon, k-1}^2$. Si el valor estimado por H es mayor que $X_{\epsilon, k-1}^2$ entonces la hipótesis se rechaza dado que hay por lo menos una muestra que se comporta de manera distinta a las otras, en caso contrario indica que las muestras provienen de la misma población es decir que son homogéneos.

Prueba De Fisher

Prueba DMS (Diferencia Mínima Significativa)

Cuando se tienen varias poblaciones con el análisis de varianza se puede determinar si las medias de dichas poblaciones son iguales o no. Esto se determina dado que si las medias son iguales entonces la varianza TOTAL disminuye, si son distintas aumenta. Cuando se requiere determinar donde están las diferencias se utilizan métodos de comparación múltiple, el método que se utilizará es DMS (Diferencia mínima significativa) o por

sus siglas en ingles LSD (Least Significant Difference) también se le conoce como <<Prueba t protegida>>.

En 1935 Ronald Aylmer Fisher propuso la prueba LSD que permite comparar pares de medias de grupos o poblaciones, esto se hace aplicando la <prueba de la t de Student> (ésta prueba permite comparar pares de medias provenientes de datos independientes con varianzas semejantes además de que el tamaño muestral es el mismo). El procedimiento consiste en:

Hipótesis a contrastar:

$$H_0 : \mu_i = \mu_j$$

$$H_a : \mu_i \neq \mu_j$$

$$t = \frac{\tilde{x}_i - \tilde{x}_j}{\sqrt{CME \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

donde CME es el error cuadrático medio, mencionado anteriormente.

Interpretación del resultado de DMS: Si la diferencia observada entre las medias muestrales es mayor que DMS entonces se toma que las diferencias son estadísticamente significativas.

Distribución Weibull

En 1939 Waloddi Weibull propuso la distribución Weibull para describir la duración de materiales, es utilizada en la fiabilidad de aplicaciones.

La distribución Weibull es un modelo estadístico con el cual se representa la probabilidad de fallo de un componente después de un lapso de tiempo t . Se obtiene la distribución de fallos de una población de donde proviene la muestra.

La función de probabilidad de fallo o función de fiabilidad se denota:

$$R(t) = \exp \left[- \left(\frac{t - \gamma}{\alpha} \right)^\beta \right]$$

donde

α es el parámetro de escala, es decir representa el tiempo.

β es el parámetro de forma, es decir representa la forma de la distribución.

γ es el parámetro de traslación, es utilizado cuando no se producen errores durante un periodo de tiempo t . Cuando $\gamma > 0$ la fiabilidad se puede aproximar a la distribución Weibull, cuando $\gamma < 0$ ocurrieron fallos antes de iniciar los ensayos.

La probabilidad de que un componente falle antes del instante t , se puede expresar en términos de la función de fiabilidad:

$$F(t) = 1 - R(t) \quad (\text{Función de distribución}).$$

Existen diversos métodos que permiten estimar los parámetros de la distribución Weibull, entre ellos se encuentran los métodos implícitos, métodos gráficos, método de la máxima verosimilitud, entre otros. El enfoque solo será al método de la máxima verosimilitud dado que es el que obtiene mejores aproximaciones:

Método de la máxima verosimilitud.

Método introducido por Ronal Fisher entre 1912 y 1922. De una forma general el método consiste en tomar una muestra aleatoria de una población determinada, dado que los valores de las variables aleatorias son obtenidos de una probabilidad.

Según López Casuso el método consiste en encontrar el valor del parámetro para el cual sea máxima la probabilidad de haber obtenido los valores muestrales encontrados.

La función de verosimilitud se define como

$$L(\alpha, \beta) = \prod_{i=1}^N f(t_i, \alpha, \beta)$$

Dada una muestra de tamaño N de una variable aleatoria T . Para obtener los valores de α y β son calculadas las derivadas parciales de la función de verosimilitud y se obtienen las siguientes ecuaciones:

$$\frac{N}{b} - N \frac{\sum_{i=1}^N t_i^b \ln t_i}{\sum_{i=1}^N t_i^b} + \sum_{i=1}^N \ln t_i = 0$$
$$\alpha = \left(\frac{\sum_{i=1}^N t_i^b}{N} \right)^{1/b}$$

Intervalos De Confianza

Cuando se estima un parámetro desconocido de una población es importante conocer un intervalo que nos permita conocer la certidumbre de dicha estimación, para ello se obtiene los intervalos de confianza.

El cálculo del intervalo de confianza depende de:

El tamaño muestral: la cantidad de datos influye dado que entre mayor sea la cantidad de datos se espera que sea menor el valor entre el valor estimado y el valor real.

El nivel o grado de confianza: la probabilidad de acierto en la estimación, es decir la probabilidad de que la estimación o predicción se encuentre en el intervalo de confianza.

La medida estimada de la muestra (media, porcentaje, varianza): los intervalos de confianza se forman por valores menores y mayores aproximados que la aproximación ofrecida por la muestra.

De acuerdo con Sánchez Algarra los intervalos de confianza proporcionan límites entre los que se puede encontrar el valor estimado del parámetro, esto es: a una estimación $\hat{\theta}$ de un parámetro θ se asocia un intervalo de la forma $(\hat{\theta} - d, \hat{\theta} + d)$, teniendo un grado de confianza conocido de contener a θ . Se expresa:

Sea $\vec{X} = (X_1, \dots, X_n)$ una muestra aleatoria simple de la población representada por la variable aleatoria X con una función de densidad $f(x, \theta)$ y sean $c_1(\vec{X})$ y $c_2(\vec{X})$ valores estadísticos tales que:

$$P(c_1(\vec{X}) < \theta < c_2(\vec{X})) = 1 - \alpha, \quad (0 < \alpha < 1)$$

Se tiene el intervalo $c_1(\vec{X}) < \theta < c_2(\vec{X})$ se toma que es un intervalo de confianza de $100(1 - \alpha)\%$. Donde c_1 y c_2 son los límites de confianza, $(1 - \alpha)$ es el coeficiente de confianza y $100(1 - \alpha)\%$ es el nivel de confianza.

Según Johnson la obtención de los intervalos de confianza se puede ver como un procedimiento de cinco pasos:

1. El planteamiento.
2. Especificación de los criterios del intervalo de confianza.
3. Recolección y presentación de hechos muestrales.
4. Determinación de los intervalos de confianza.
5. Presentación de los resultados.

CAPÍTULO 5. PLANTEAMIENTO DEL PROBLEMA Y ANÁLISIS

Planteamiento del problema

El pronóstico de la demanda eléctrica permite estimar cuanta energía será requerida por los usuarios en un instante de tiempo futuro y en un lugar determinado.

El pronóstico de demanda máxima de potencia eléctrica contribuye a la toma de decisiones adecuadas sobre optimización y expansión de la empresa Comisión Federal de Electricidad (CFE), por esta razón deberá ser preciso, ya que de no serlo puede ocasionar escasez de energía eléctrica así como también podría ocasionar que se trabaje con un mínimo de energía durante un lapso de tiempo.

Se cuenta con un módulo que permite obtener el pronóstico. Sin embargo se requiere tener mayor confiabilidad.

El Laboratorio de Visión por Computadora (LVC) del Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) desarrolló un prototipo de pronóstico de la demanda máxima de potencia eléctrica a nivel zona basada en un método de pronóstico multipaso y en modelos de estimación con algoritmos de series de tiempo y aprendizaje computacional, dicho prototipo fue desarrollado como una aplicación de escritorio, sin embargo se requiere la implementación del prototipo en una plataforma web.

Para obtener el pronóstico de la demanda máxima a nivel zona se requiere de diversas herramientas estadísticas, es decir se requiere de métodos que nos permitan obtener la demanda de potencia eléctrica. La finalidad del proyecto es la colaboración en el análisis e implementación de herramientas estadísticas que serán utilizadas para llevar a cabo el cálculo del pronóstico de la demanda máxima de potencia eléctrica para una plataforma web.

Antecedentes Del Proyecto

Actualmente la Comisión Federal de Electricidad (CFE) obtiene el pronóstico de la Demanda Máxima de Potencia Eléctrica por medio de un sistema denominado Desarrollo del Mercado Eléctrico (DME).

Objetivos

La Comisión Federal de Electricidad (CFE) requiere de una herramienta que permita obtener el pronóstico de la demanda máxima de potencia eléctrica además de que dicha herramienta debe tener mayor confiabilidad que el sistema DME.

Alcance

Se analizaron, diseñaron, implementaron y probaron las Herramientas Estadísticas requeridas para el pronóstico de la demanda máxima de potencia eléctrica a nivel zona en una plataforma Web.

Restricciones

El acceso al sistema estará restringido mediante autenticación mediante un usuario y una contraseña. Únicamente se le permitirá el acceso al personal autorizado por CFE, el pronóstico de la demanda máxima de potencia eléctrica solo lo pueden realizar responsables de zona.

Como se mencionó antes existen diversas metodologías ágiles sin embargo se optó por utilizar DSDM por las siguientes razones:

- ☞ Asegura la culminación exitosa del proyecto dado que antes de iniciar con el desarrollo se realizan los estudios pertinentes para determinar la factibilidad del proyecto.
- ☞ Asegura cumplir satisfactoriamente con las metas del proyecto debido a la amplia participación de los usuarios durante el desarrollo del proyecto.
- ☞ La calidad del software incrementa en cada iteración.

Estudio de viabilidad

El laboratorio de visión por computadora del INAOE realizó un análisis para determinar el número de etapas de las que constará el proyecto, así como una estimación de tiempo para cada una de ellas. Asimismo se realizó un estimado de los gastos. Durante estos análisis se identificaron los siguientes riesgos que se pueden presentar:

- No tener acceso a los sistemas de información que contienen información para la obtención del pronóstico.
- No contar con la disponibilidad de tiempo necesaria por parte del personal de CFE para llevar a cabo las entrevistas.

Basados en todo lo anterior se determinó que el sistema web para el Pronóstico de la Demanda Máxima (PDM) reúne características que aseguran la realización y culminación de las metas y objetivos.

Estudio de negocio

La estimación de la demanda máxima de potencial eléctrica es un problema que se tiene desde hace un tiempo, por lo que CFE se dio a la tarea de desarrollar una aplicación que lleve a cabo esta función. Sin embargo la aplicación que desarrollaron y utilizan tiene mu-

cha incertidumbre, lo cual es un inconveniente para CFE ya que puede ocasionar pérdidas económicas.

Por ello el sistema web PDM utilizará métodos de aprendizaje computacional con ello se reducirá la incertidumbre aumentando la confiabilidad de los pronósticos de demanda máxima de potencia eléctrica en diferentes horizontes dado que el sistema PDM web hará uso de diversas las herramientas estadísticas. Cabe mencionar que debido a la magnitud del sistema web PDM está dividido en diversos módulos sin embargo el módulo en el cual se enfoca ésta tesis es el de estadísticos.

Para obtener la versión final de las herramientas estadísticas utilizadas en el pronóstico de la demanda máxima de potencia eléctrica en una plataforma Web se realizaron varias iteraciones que a continuación se mencionan.

1. Definición de los casos de uso derivados de Crear Experimento.
2. Definición de los casos de uso para obtener la correlación de variables y refinamiento de los casos de uso derivados de Crear Experimento.
3. Realizar correcciones al caso de uso para la correlación de variables y se inicio la implementación del algoritmo Kruskal Wallis, teniendo como entrada una lista de modelos de aprendizaje en el proyecto propuestaPDM.
4. Implementación del algoritmo DMS teniendo como entrada una lista de modelos de aprendizaje en el proyecto propuestaPDM y realizando los cálculos con los pronósticos de los algoritmos de aprendizaje, se refinó la implementación del algoritmo Kruskal Wallis.
5. Se realizan refinamientos al algoritmo DMS y el algoritmo Kruskal Wallis se inicia la implementación del algoritmo para obtener el intervalo de confianza del pronóstico para modelos de aprendizaje computacional en el proyecto propuestaPDM.
6. Implementación de correlación de variables que recibe como entrada una lista de *doubles*¹ en el proyecto propuestaPDM, se refina el cálculo de intervalos de confianza para el pronóstico.
7. Implementación de la matriz de correlación recibiendo una matriz de doubles que representan los valores de entrada de las variables.
8. Traslado de Kruskal Wallis al proyecto PDMWEB.
9. Traslado de DMS al proyecto PDMWEB.
10. Traslado del algoritmo para intervalo de confianza del proyecto PDMWEB.
11. Traslado de correlación de variables al proyecto PDMWEB.
12. Traslado de matriz de correlación al proyecto PDMWEB.

¹ *double*: tipo de dato de java en coma flotante y doble precisión.

13. Adaptación de la matriz de correlación para que recibiera una lista de variables de entrada.
14. Implementación de las medidas de precisión del pronóstico: RMSE, MAPE, MAE.
15. Obtención del pronóstico extendido a partir de un conjunto de modelos de aprendizaje obtenidos por DMS, el conjunto que se utiliza es el que cumple con tener el menor promedio.
16. Obtención del pronóstico promedio
17. Adaptación de los algoritmos DMS para realizar los cálculos con los ajustes de los modelos de aprendizaje.
18. Adaptación del intervalo de confianza para Pronóstico Sugerido.
19. Adaptación del pronóstico extendido para que utilice el conjunto con mayor número de elementos y que además tiene en mayor promedio.

Análisis

En este capítulo se detallarán los casos de uso relacionados con el módulo de estadísticos.

Actores

- ⇒ Usuario divisional: Acceso a la visualización de la zonas
- ⇒ Usuario zonal: Acceso a su zona responsable, encargado de llevar a cabo el pronóstico.

Definición de casos de uso

1. Selección de métodos y sus modelos.
2. Selección de variables.
3. Obtener pronósticos de los modelos.
4. Obtener pronóstico sugerido.
5. Crear experimento.
6. Obtener correlación de Variables.

Diagrama de actividades correspondiente a Crear Experimento.

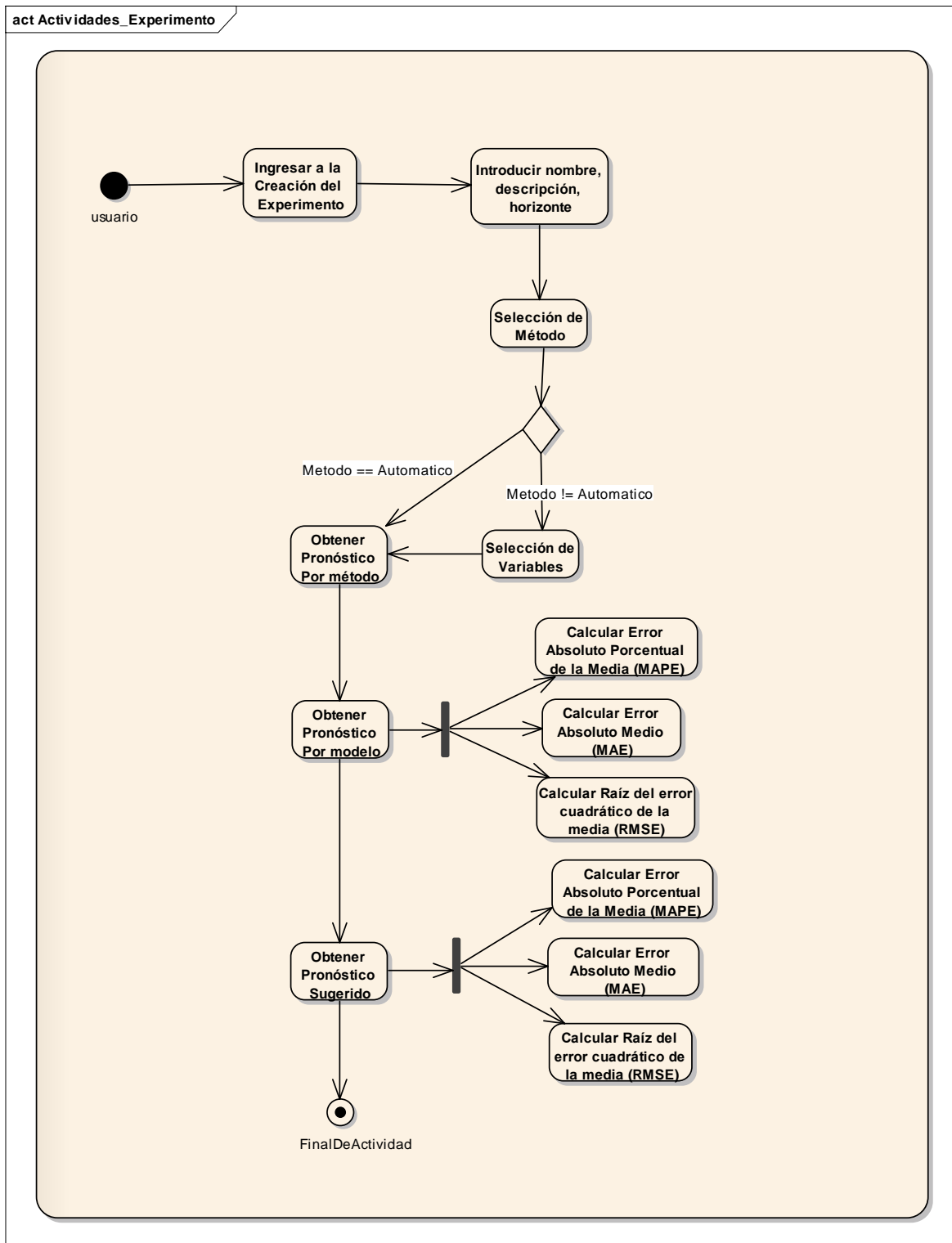


Figura 10. Diagrama de actividades de crear experimento

Diagrama de actividades para la matriz de correlación:

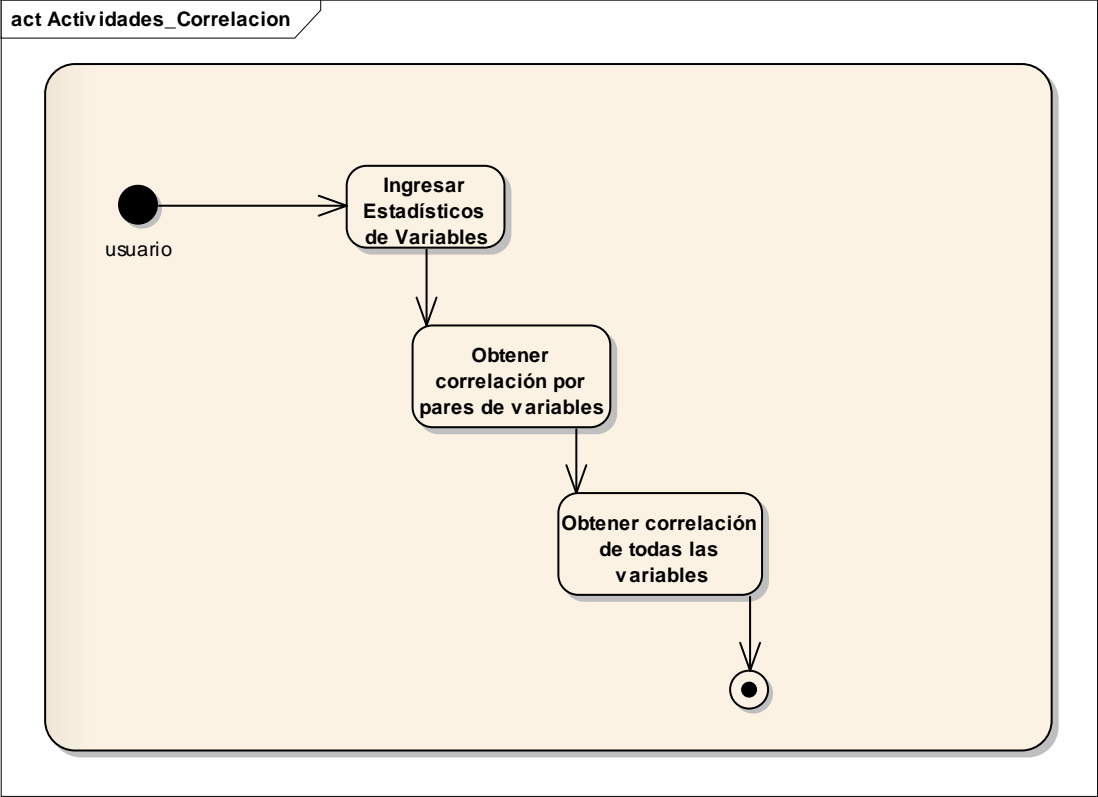


Figura 11. Diagrama de actividades para la correlación de variables

Descripción de los casos de uso

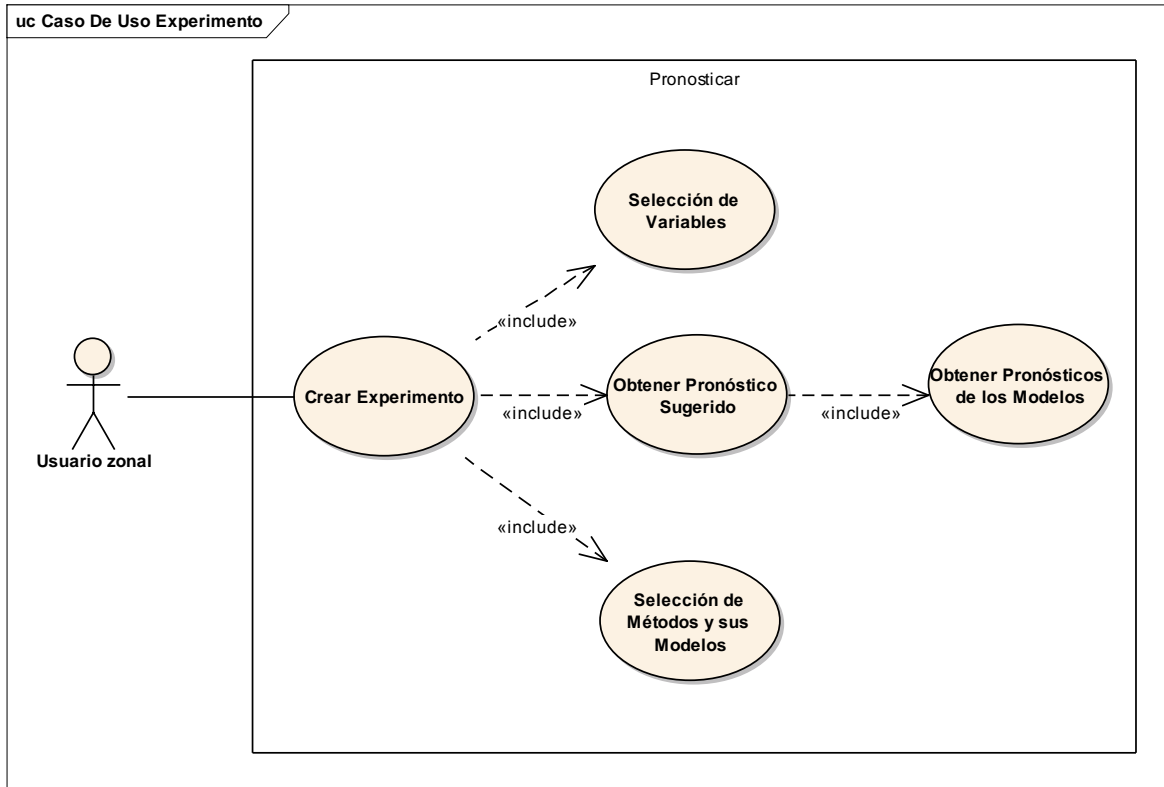


Figura 12. Caso de uso Crear Experimento

Caso de uso	Selección de métodos y sus modelos
Actores	Usuario zonal
Propósito	Seleccionar los métodos y sus modelos con los cuales se pronosticará la demanda máxima de potencia eléctrica.
Resumen	Determinar el o los métodos que se emplearán para obtener el pronóstico de la demanda máxima de potencia eléctrica.
Precondiciones	Autenticarse como usuario del sistema. Elegir la zona a pronosticar, estar en la vista Pronosticar.
Flujo Normal	
<ol style="list-style-type: none"> 1. El usuario zonal seleccionará métodos y modelos 2. El sistema carga desde un archivo de configuración los distintos métodos multipaso y los modelos de aprendizaje computacional que los componen y el método univariante. El sistema muestra: <ol style="list-style-type: none"> a. el método Automático. 	

<ul style="list-style-type: none"> b. Los distintos métodos multipaso con sus respectivos modelos de aprendizaje. c. Método univariante. <p>3. El usuario zonal selecciona el método automático.</p>
Flujo Alternativo
<p>3. El usuario zonal selecciona el o los métodos multipaso con los modelos de aprendizaje que lo conformarán.</p> <p>3. El usuario zonal selecciona el método univariante.</p>

Caso de uso	Selección de Variables
Actores	Usuario zonal
Propósito	Seleccionar diversas variables que se utilizarán para obtener el pronóstico de la demanda máxima de potencia eléctrica a nivel zona.
Resumen	Determinar las variables que intervendrán en la predicción de la demanda máxima de potencia eléctrica.
Precondiciones	Encontrarse en el vista Pronosticar. Selección de métodos y modelos.
Flujo Normal	
<ul style="list-style-type: none"> 1. El usuario zonal iniciará la selección de variables para el pronóstico. 2. El sistema muestra al usuario las diversas variables que afectan el comportamiento de la demanda máxima. 3. Si el método seleccionado es Automático entonces no se puede realizar la selección de variables dado que el sistema considerará todas las variables para la obtención del pronóstico. 	
Flujo Alternativo	
<ul style="list-style-type: none"> 3. El usuario zonal selecciona las variables que intervendrán para la obtención del pronóstico de una zona. 	

Caso de uso	Obtener pronósticos de los modelos.
Actores	Usuario zonal
Propósito	Obtener el pronóstico de la demanda máxima de potencia eléctrica por medio de los modelos de aprendizaje computacional.
Resumen	Obtener el pronóstico de la demanda máxima de potencia eléctrica para una zona determinada por medio de modelos de aprendizaje computacional, así mismo obtener las medidas de precisión del mismo.
Precondiciones	Haber realizado la selección de variables. Haber realizado la selección de métodos y sus modelos.
Referencias	Selección de variables. Selección de métodos y sus modelos
Flujo Normal	
<ol style="list-style-type: none"> 1. El usuario zonal selecciono el método para pronosticar la demanda máxima de potencia eléctrica. 2. El sistema obtiene las variables seleccionadas, pronostica el valor de las variables mediante el método univariante. Utiliza los valores predichos para pronosticar la demanda máxima de potencia eléctrica. De acuerdo al método se obtienen los modelos de aprendizaje computacional pronosticando con ellos la demanda máxima de potencia eléctrica, así mismo se obtiene las medidas de precisión y de confianza para cada pronóstico. El sistema muestra los resultados de los pronósticos. 	
Flujo Alterno	
<ol style="list-style-type: none"> 2. El sistema obtiene las variables seleccionadas, pronostica el valor de las variables mediante el método univariante. Utiliza los valores predichos para pronosticar la demanda máxima de potencia eléctrica. Pronostica la demanda máxima de potencia eléctrica mediante el método univariante, este método no cuenta con modelos de aprendizaje computacional. 	

Caso de uso	Obtener pronóstico sugerido.
Actores	Usuario zonal
Propósito	Obtener el pronóstico de la demanda máxima de potencia eléctrica para una zona.

Resumen	Obtener el pronóstico de la demanda máxima de potencia eléctrica que será almacenado por el sistema como el sugerido para una zona.
Precondiciones	Haber obtenido el pronóstico de los modelos de aprendizaje computacional
Referencias	Obtener pronóstico de modelos
Flujo Normal	
<ol style="list-style-type: none"> 1. El usuario zonal ejecuta el pronóstico. 2. El sistema por medio de métodos estadísticos obtiene los modelos de aprendizaje computacional cuyo resultado de pronóstico es similar y de ellos calcula el pronóstico promedio así como sus medidas de precisión (MAE, MAPE, RMSE) y confianza (intervalos de confianza). Muestra los resultados de cada modelo e incluye el pronóstico promedio. 3. El usuario selecciona el pronóstico que de acuerdo a sus conocimientos sea el más aceptable. 	
Flujo Alternativo	

Caso de uso	Crear experimento
Actores	Usuario zonal
Propósito	Obtener el pronóstico sugerido para una zona
Resumen	Se crea un experimento con el cual se obtiene el pronóstico sugerido de la DMPE para un periodo de tiempo determinado, así mismo se obtiene las medidas de precisión del pronóstico sugerido.
Precondiciones	Autenticarse como usuario del sistema.
Referencias	Selección de variables. Selección de métodos y modelos. Obtener pronóstico sugerido.

Flujo Normal

1. El sistema muestra las zonas que se pueden pronosticar.
2. El usuario zonal selecciona la zona que desee pronosticar y pulsa pronosticar.
3. El usuario zonal selecciona Crear experimento.
4. El sistema pide al usuario introducir el horizonte, descripción.
5. Seleccionar método y modelos (ver Selección de métodos y modelos).
Selección de variables (ver Selección de variables).
6. El usuario proporciona al sistema el horizonte y la descripción.
Oprime el botón Nuevo.
7. El sistema obtiene los valores de las variables seleccionadas que intervendrán en el pronóstico de la DMPE, almacena todos los parámetros de entrada como parte del experimento. (Ver Obtener pronóstico sugerido)
8. El sistema almacena el pronóstico sugerido como parte de un experimento.

Flujo Alterno

uc Caso De Uso Correlacion

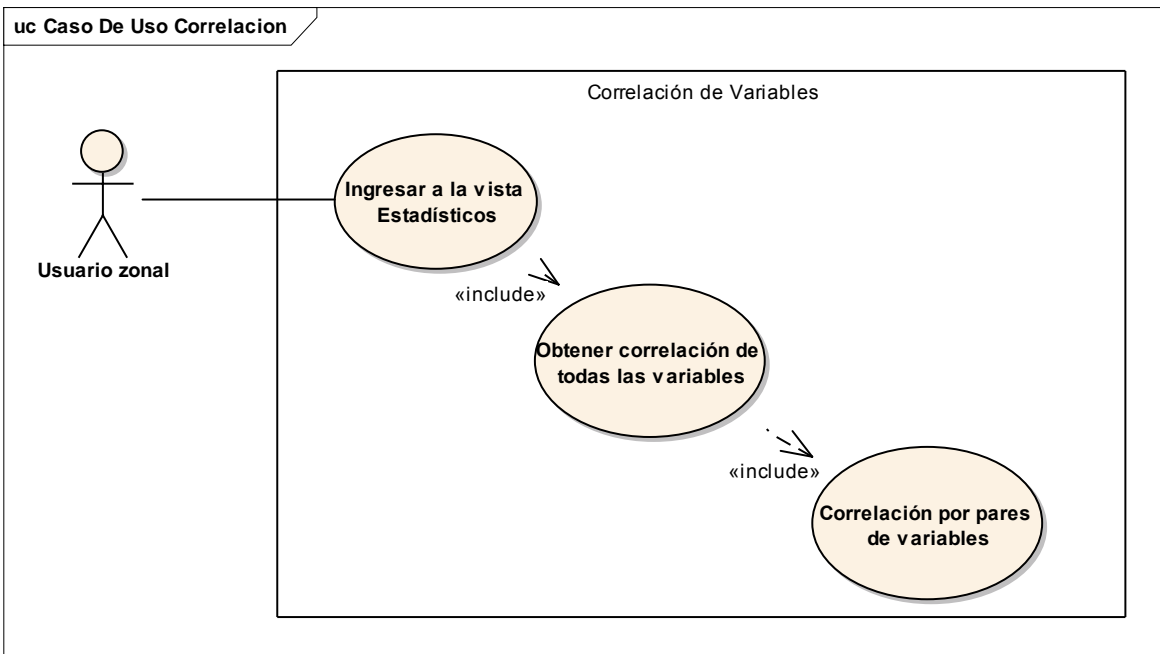


Figura 13. Caso de uso obtener Correlación de Variables

Caso de uso	Obtener Correlación de Variables
Actores	Usuario zonal
Propósito	Obtener la correlación de todas las variables involucradas en el pronóstico de la demanda máxima de potencia eléctrica.
Resumen	Obtener correlación de variables.
Precondiciones	Autenticarse como usuario del sistema
Flujo Normal	
<ol style="list-style-type: none"> 1. El usuario ingresa a la vista Estadísticos. 2. El sistema obtiene las variables que intervienen el pronóstico de la demanda máxima, posteriormente obtiene los valores de dichas variables. El sistema calcula la correlación de Pearson para cada par de variables, almacenando los valores obtenidos por de la correlación. 	
Flujo Alternativo	

Diagrama de secuencia correspondientes a los casos de uso

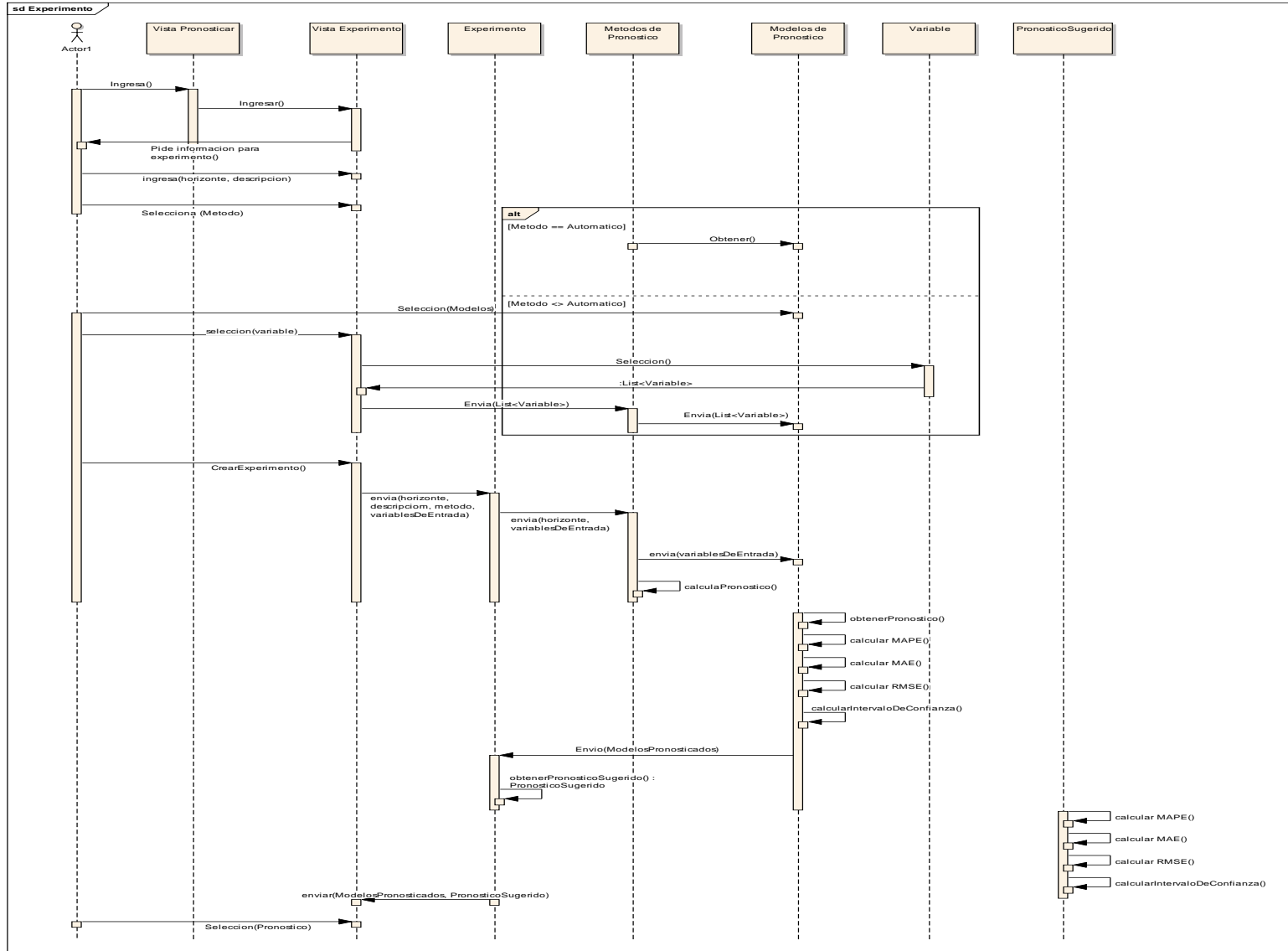


Figura 14. Diagrama de secuencia de Crear Experimento

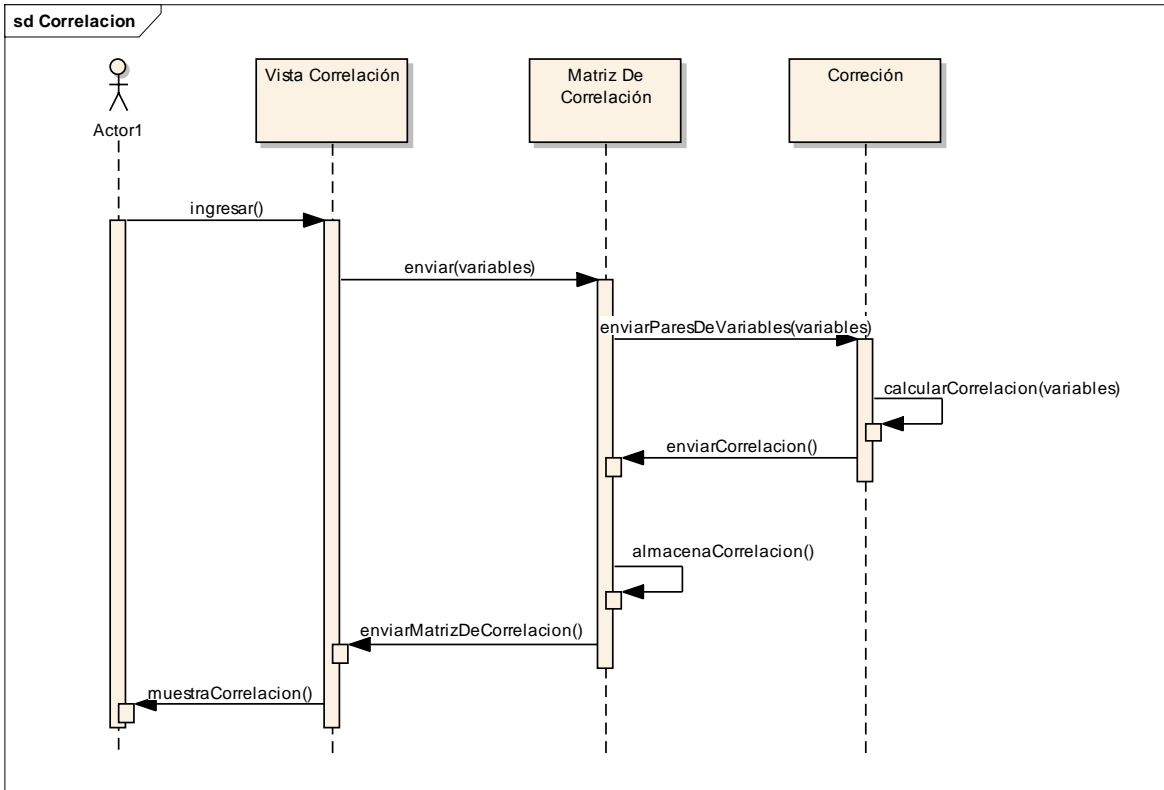


Figura 15. Diagrama de secuencia de obtener Correlación de Variables

CAPÍTULO 6. DISEÑO Y ARQUITECTURA

Diseño

Para poder obtener el pronóstico promedio se debe llevar a cabo el procedimiento que a continuación de muestra

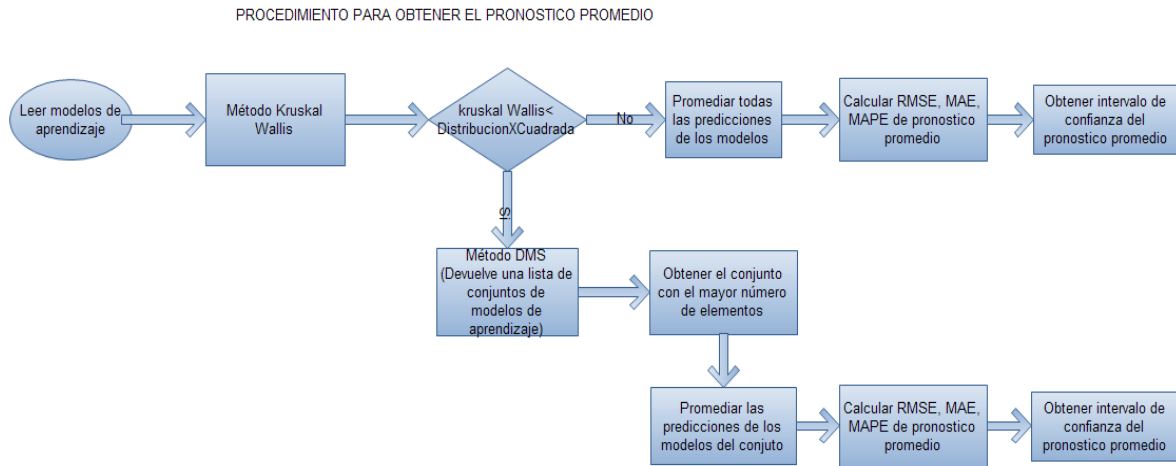


Figura 16. Procedimiento para obtener el pronóstico promedio

Modelo de datos

A continuación se muestra el diagrama de clases, las clases que se muestran son las relacionadas con el módulo herramientas estadísticas.

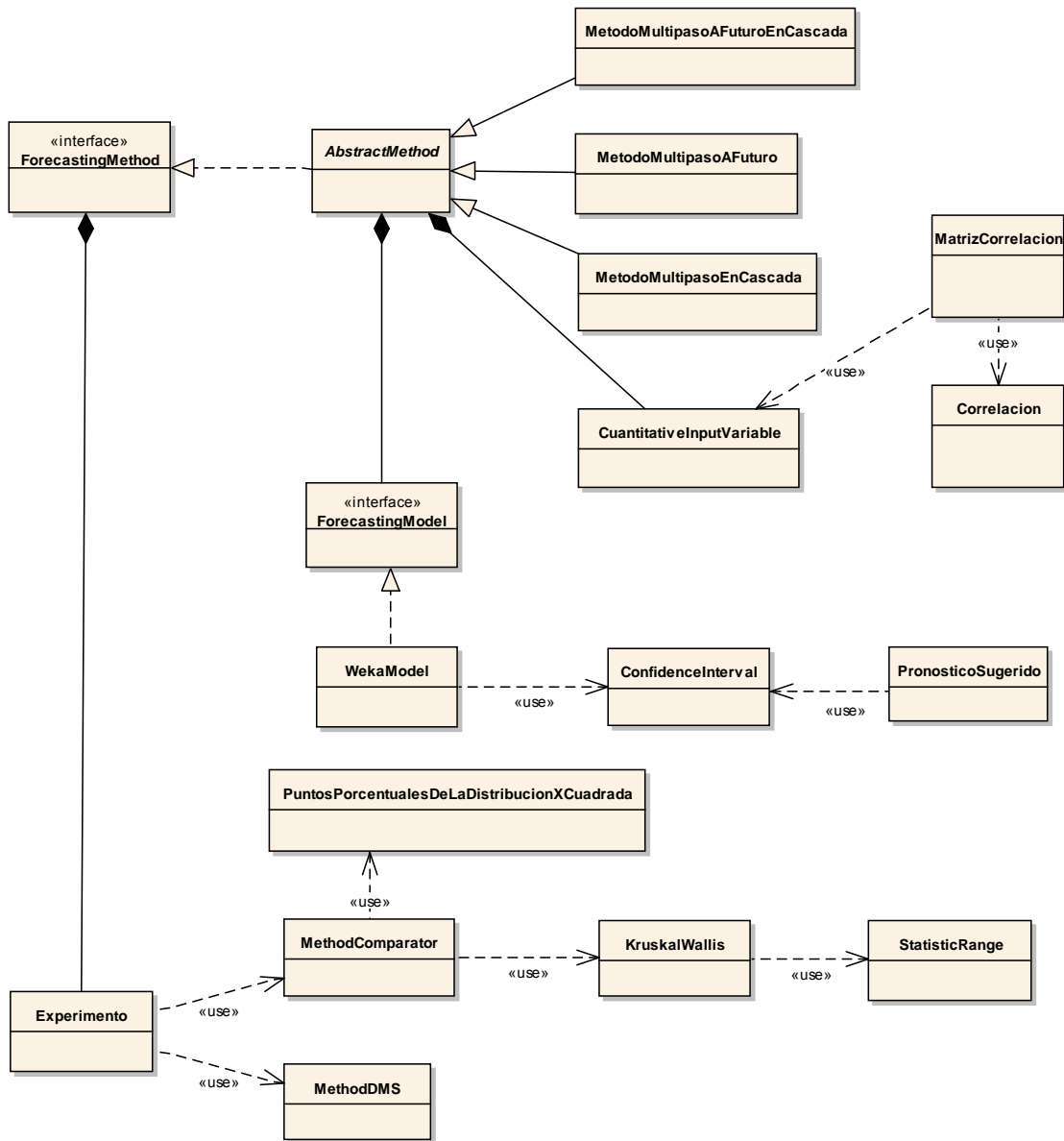


Figura 17. Diagrama de clases utilizadas para el módulo Herramientas Estadísticas

Clases que intervienen para el cálculo del pronóstico promedio.

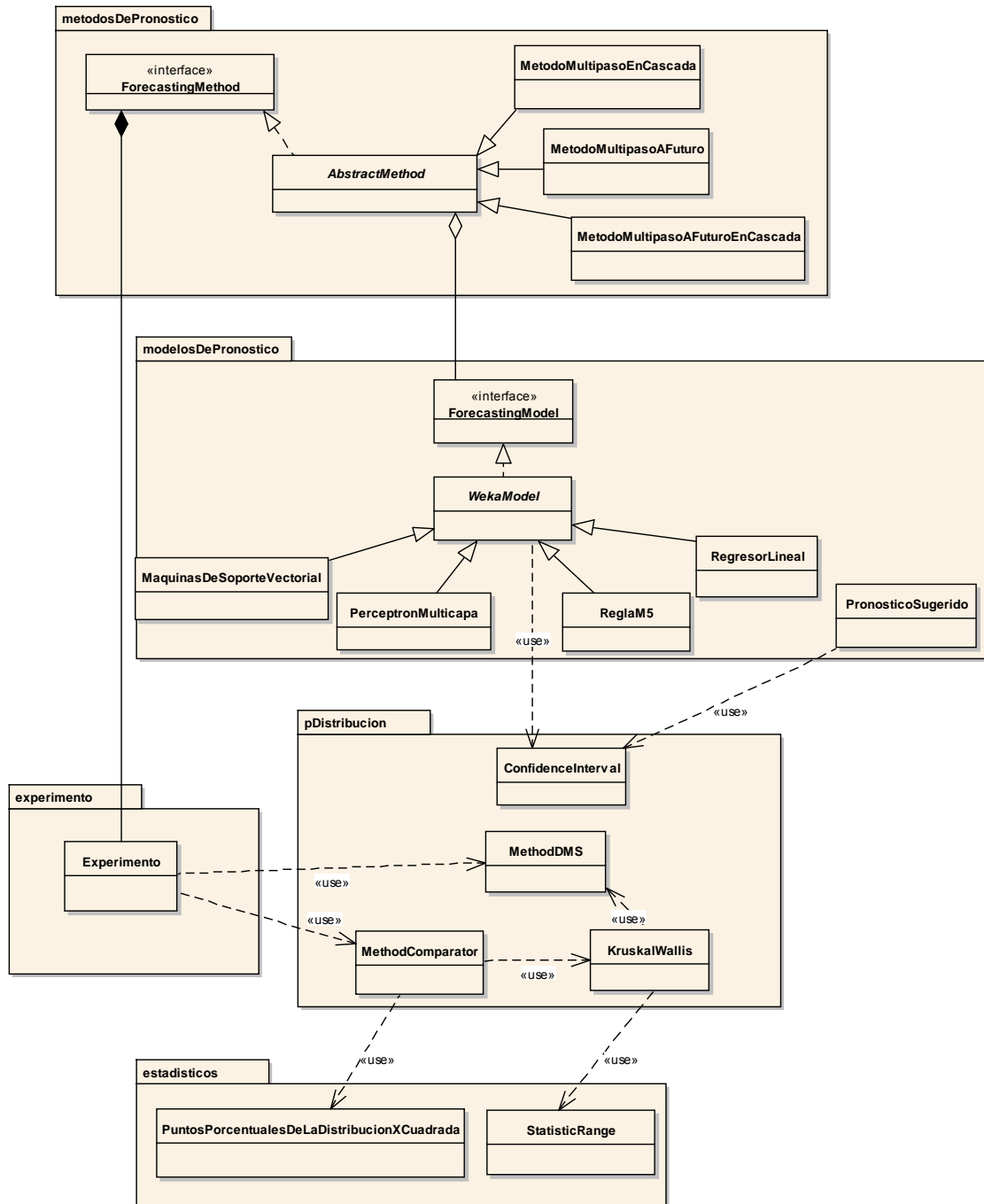


Figura 18. Diagrama de paquetes correspondiente al pronóstico

Clases que intervienen en la obtención de la correlación de las variables de entrada.

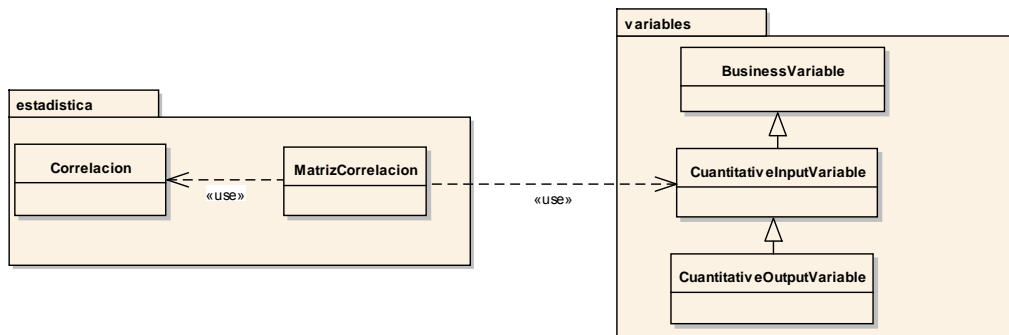


Figura 19. Diagrama de componentes correspondientes a Correlación

Dado que una parte fundamental de esta de DSDM son las iteraciones, en esta parte se mostrará la organización final de las herramientas estadísticas:

- Matriz de correlación
- Método Kruskal Wallis
- Método DMS
- Obtención del pronóstico promedio
- Obtención del pronóstico extendido
- Medidas de precisión del pronóstico: RMSE, MAPE, MAE.
- Intervalo de Confianza

ForecastingModel es la interface que declara los métodos que utilizarán los modelos de aprendizaje computacional. WekaModel implementa ForecastingModel los métodos que se implementaron a lo largo de este trabajo son: getMAPE(), getMAE(), getRMSE(). La clase PronosticoSugerido no implementa ForecastingModel dado que se hace una adaptación para otro tipo de modelos que no son de aprendizaje computacional, sin embargo también se implementaron los métodos getMAPE(), getMAE(), getRMSE().

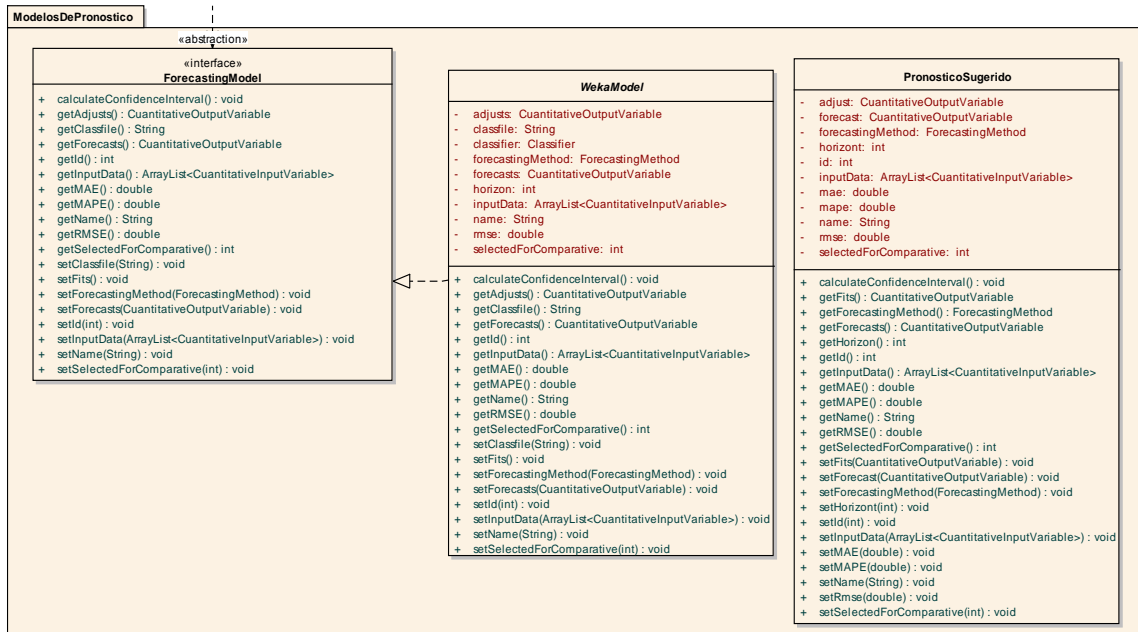


Figura 20. Interface ForecastingModel, clase WekaModel y PronosticoSugerido

La clase MatrizCorrelacion recibe como parámetros una lista de variables de entrada y permite obtener la correlación entre dichas las variables, el retorno es una matriz de doubles que representan la correlación del conjunto las variables.

La clase Correlacion recibe una pareja de listas de doubles y con ellas obtiene la Correlación de Pearson. Cada lista de doubles representa los valores de una variable de entrada. El retorno es un double que representa la correlación entre dos variables.

Ambas clases fueron totalmente implementadas durante el desarrollo de este trabajo.

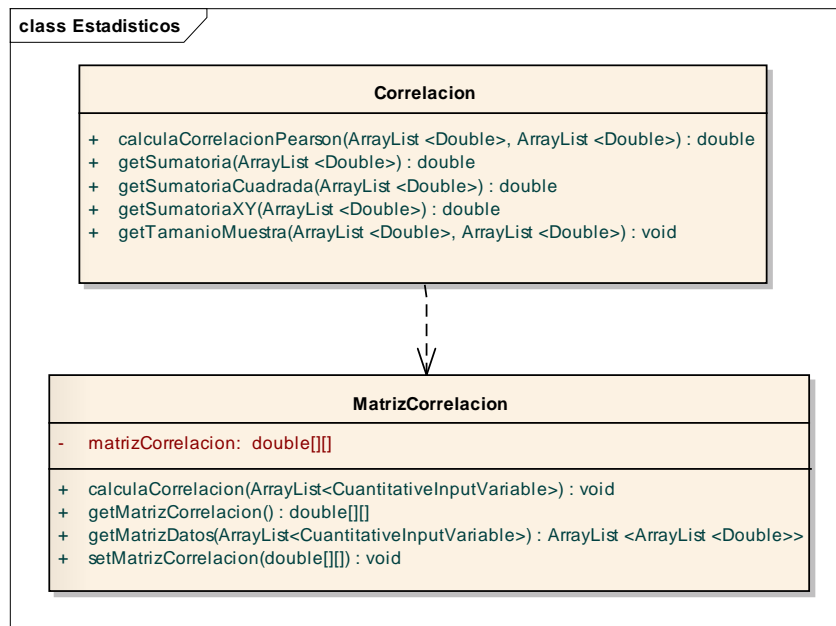


Figura 21. Diagrama de las clases utilizadas para la correlación de variables

La clase MethodComparator recibe una lista de métodos multipaso, utiliza la clase KruskalWallis que en conjunto con la clase PuntosPorcentualesDeLaDistribucionXCuadrada determina si los modelos de aprendizaje son similares. Retorna un valor booleano que determina la similitud de modelos.

Clase KruskalWallis recibe una lista de métodos multipaso y con ellos obtiene el valor de una constante, esta es utilizada para determinar si son o no similares los modelos de aprendizaje.

Clase ConfidenceInterval obtiene el intervalo de confianza tanto de los WekaModels como de PronosticoSugerido.

Clase MethodDMS recibe todos los modelos de aprendizaje de los métodos y construye una lista de conjuntos de modelos de aprendizaje, los conjuntos son formados por los modelos que son similares.

Las clases que se acaban de mencionar fueron totalmente implementadas durante el desarrollo de este trabajo.

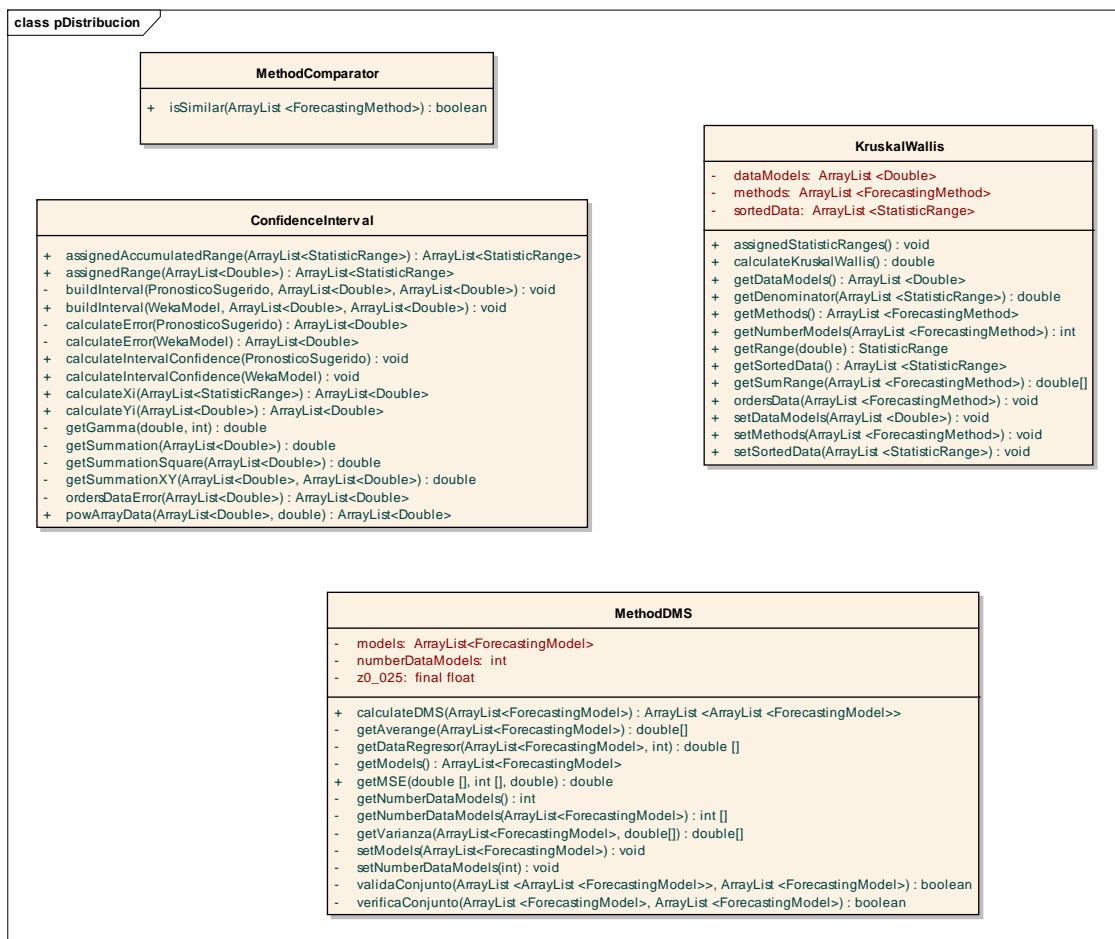


Figura 22. Diagrama de las clases que intervienen para obtener el pronóstico promedio

La clase Experimento se puede considerar como la más importante dado que en ella se realiza todo el procedimiento de obtener el pronóstico promedio. Para ello utiliza el método: `calcularPronosticoSugerido()`, en el se determina de cuales modelos se obtendrá el pronóstico promedio, para llevar a cabo este cálculo utiliza las clases `KruskalWallis`, `MethodDMS`, `MethodComparador`.



Figura 23. Diagrama de la clase Experimento

Arquitectura

Se utilizó Java como lenguaje para el desarrollo del módulo Herramientas Estadísticas dado que fue requerimiento del usuario.

Java: es un lenguaje multiplataforma y de programación de alto nivel, orientado a objetos. Tiene diferentes distribuciones para distintos fines, entre las que destacan:

- La plataforma Java 2 Standar Edition J2SE:
Consiste en un entorno de tiempo de ejecución y un conjunto de diversos API's para crear una variedad de aplicaciones utilizando principalmente las siguientes funcionalidades: conectividad con bases de datos, entrada y salida, gráficos libre-rías como AWT, SWING, procesamiento de XML.
- La plataforma Java 2, Enterprise Edition (J2EE):
J2EE es una plataforma que trabaja con J2SE permite la creación de diversas aplicaciones principalmente con el uso de los siguientes servicios: Servlets, JavaServerPages, tecnología XML, conectividad a BD.
- La plataforma Java 2, Micro Edition (J2ME):
Permite la creación de aplicaciones Java para pequeños dispositivos móviles como teléfonos, tabletas, etc.

El sistema Pronóstico de la demanda maxima de potencia electrica estará situada en el servidor del SIAD.

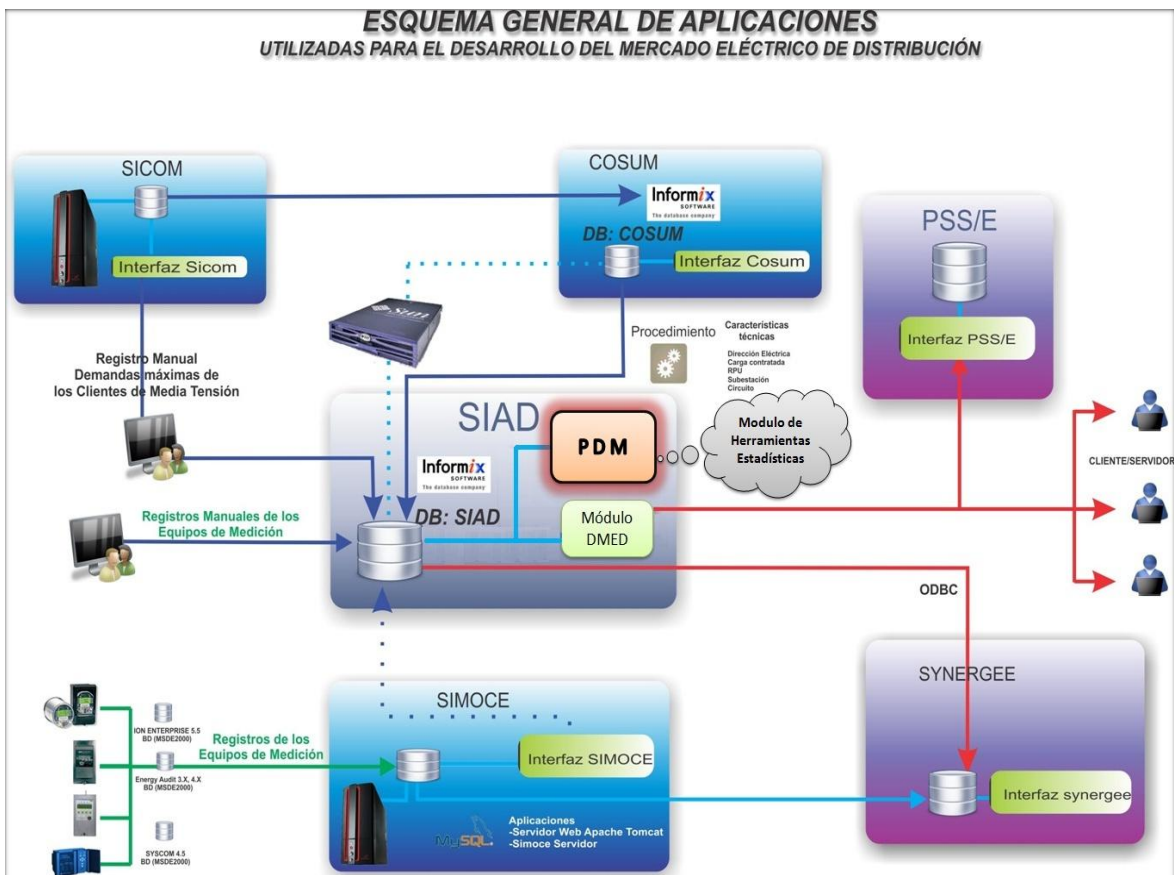


Figura 24. Esquema de aplicaciones que intervienen para el desarrollo del mercado eléctrico de distribución

CAPÍTULO 7. PRUEBAS Y RESULTADOS

Pruebas

Una vez terminada la codificación del software se tiene que realizar una serie de pruebas que permitan comprobar el correcto funcionamiento del módulo Herramientas Estadísticas. Existen las pruebas de unidad, de integración y de validación.

Las pruebas de unidad consisten en verificar el correcto funcionamiento de las unidades más pequeñas del sistema, las de integración comprueban el adecuado funcionamiento de un conjunto de elementos o módulos del sistema. Cabe mencionar que hay dos formas distintas de realizar las pruebas: de caja negra cuando se realizan sobre la interfaz para demostrar el adecuado funcionamiento del software y cuando se requiere un detallado análisis de la estructura del código se realizan las pruebas de caja blanca.

Un experto se encargó de realizar las pruebas de unidad verificando que los resultados arrojados por el módulo implementado tuvieran un error aceptable, para ello utilizó la aplicación que fue desarrollada como primer parte del sistema de pronóstico de la DMPE.

El RMSE y el MAE se mide en Megawatts, de acuerdo con el experto se considera una desviación aceptable hasta 0.30 entre los resultados arrojados por ambas aplicaciones.

El MAPE indica el porcentaje de desviación de 3.5 % en cuanto a valores reales.

Posteriormente el experto realizó las pruebas de integración es decir que al incorporar el módulo a la parte web continúe funcionando correctamente.

Se realizaron diferentes pruebas en la que se obtiene el pronóstico promedio, encontrando un error aceptable de acuerdo a los criterios del experto.

	RMSE	MAPE	MAE	
Con todas las variables	4.41454037	39.6977369	3.19086517	PDMWEB
	4.16098682	36.4676539	2.89312667	PROTO
Sin Salario Mínimo	3.86882099	34.9942583	2.79214935	PDMWEB
	3.9690795	34.3832905	2.71106	PROTO
sin UAgricola	3.84531137	38.7749733	3.06250765	PDMWEB
	4.16098682	36.4676539	2.89312667	PROTO
Sin Población	4.20140738	43.3032786	3.49362792	PDMWEB
	4.5834248	43.5497576	3.5157	PROTO

Tabla 1 Comparación de errores de los pronósticos obtenidos con el sistema PDMWEB y el prototipo de MatLab

Resultados

Para fines de ilustrar el funcionamiento de las clases diseñadas se construyeron los siguientes escenarios con datos correspondientes a la zona Xalapa:

Escenario Pronosticar.

Para obtener el pronóstico promedio se requiere pronosticar los metodos multipaso con sus respectivos modelos de aprendizaje, para ello se accede a la interfaz gráfica Pronosticar.

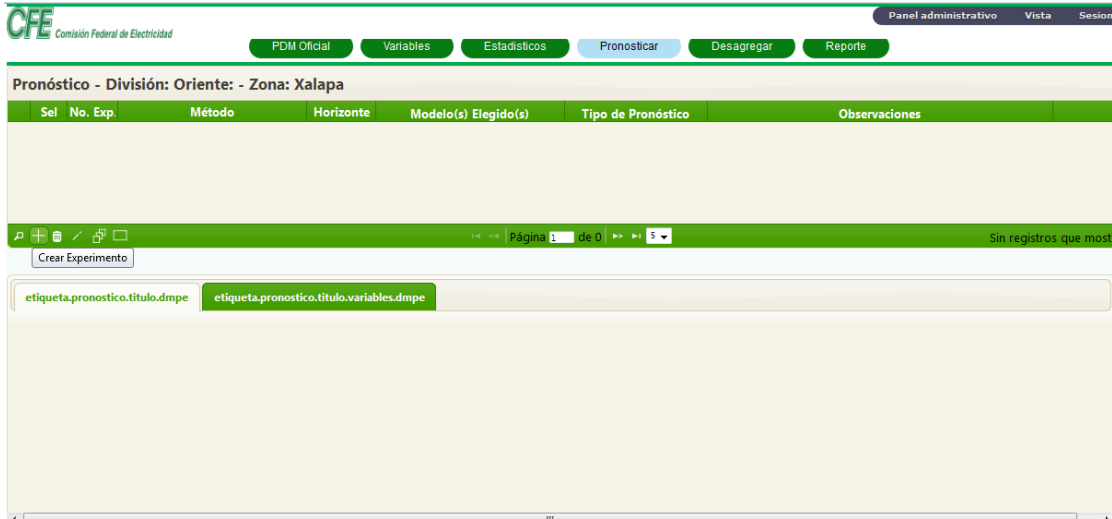


Figura 25. Interfaz de Pronosticar

Posteriormente se pulsa el boton Crear Experimento, el experimento se creará con el método Automático, este implica que participen todas las variables.

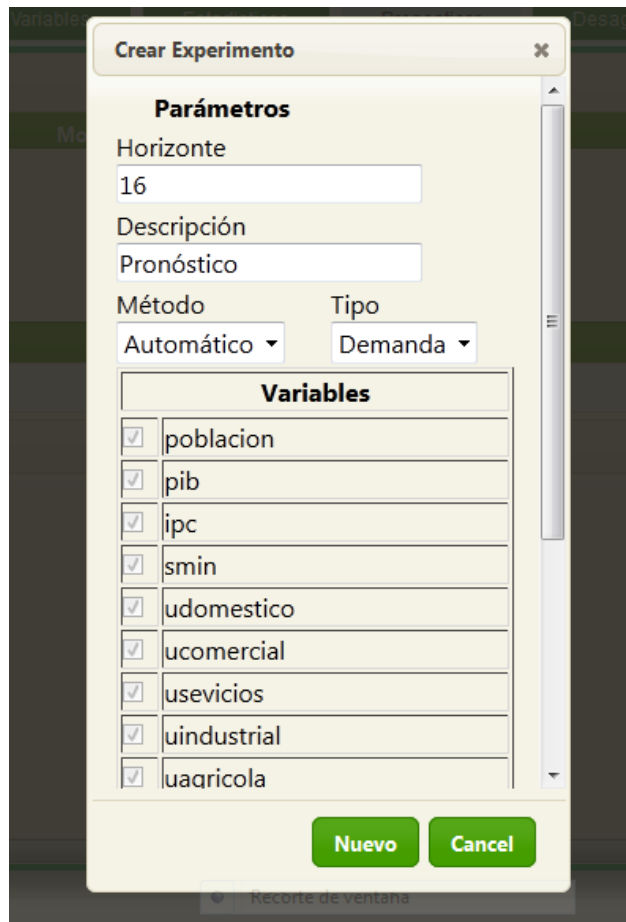
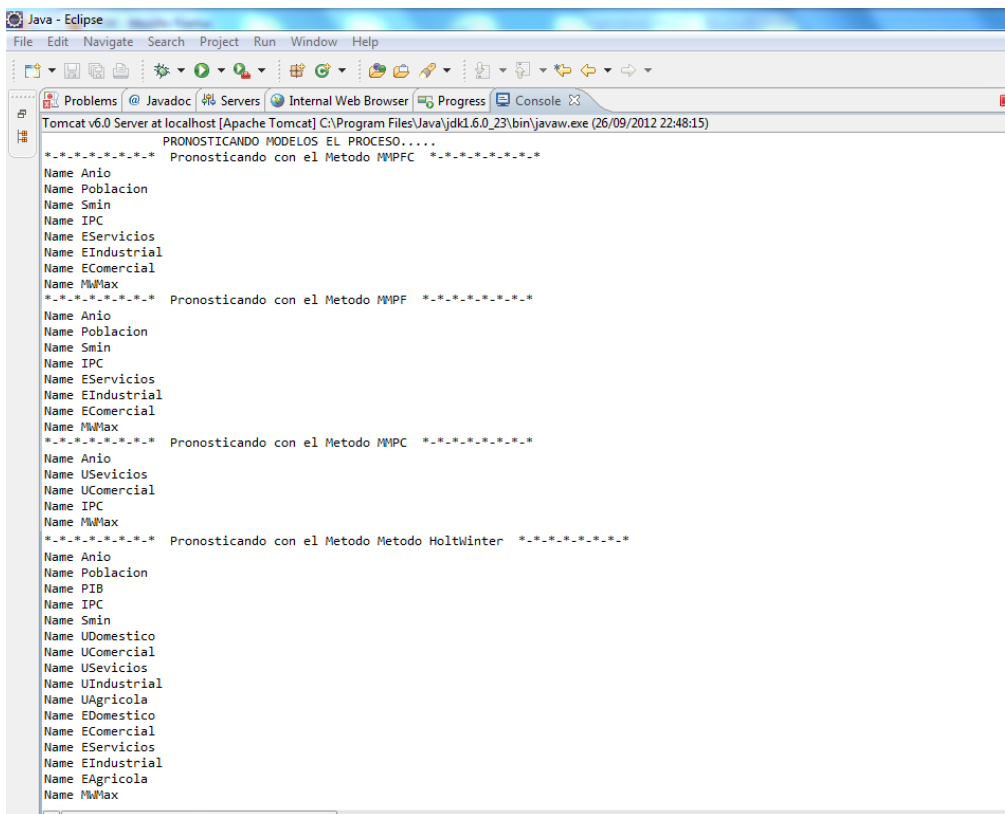


Figura 26. Interfaz de Crear Experimento

Cuando se presiona el boton Nuevo, se realiza todo el procedimiento descrito anteriormente en Figura 16. Procedimiento para obtener el pronóstico promedio.

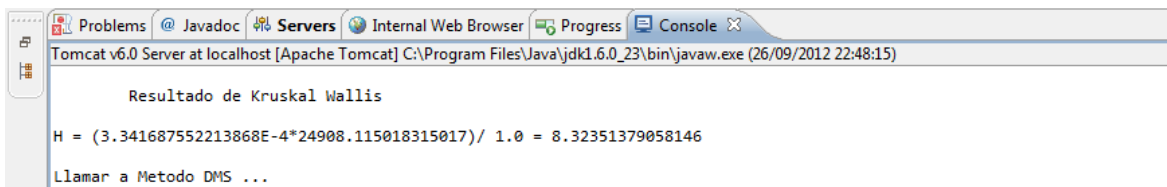
Se pronostican los métodos multipaso con sus respectivos modelos.



```
Tomcat v6.0 Server at localhost [Apache Tomcat] C:\Program Files\Java\jdk1.6.0_23\bin\javaw.exe (26/09/2012 22:48:15)
PRONOSTICANDO MODELOS EL PROCESO...
*-*-*-*-*
Name Anio
Name Poblacion
Name Smin
Name IPC
Name EServicios
Name EIndustrial
Name EComercial
Name MIMax
*-*-*-*-* Pronosticando con el Metodo MWPF *-*-*-*-*
Name Anio
Name Poblacion
Name Smin
Name IPC
Name EServicios
Name EIndustrial
Name EComercial
Name MIMax
*-*-*-*-* Pronosticando con el Metodo MWPC *-*-*-*-*
Name Anio
Name UServicios
Name UComercial
Name IPC
Name MIMax
*-*-*-*-* Pronosticando con el Metodo Metodo HoltWinter *-*-*-*-*
Name Anio
Name Poblacion
Name PIB
Name IPC
Name Smin
Name UDomestico
Name UComercial
Name UServicios
Name UIndustrial
Name UAgricola
Name EDomestico
Name EComercial
Name EServicios
Name EIndustrial
Name EAgricola
Name MIMax
```

Figura 27. Pronosticando métodos multipaso

Se utilizan todos los modelos de aprendizaje computacional de los métodos para calcular Kruskal Wallis, con ello se determina si se utiliza el método DMS para obtener los modelos que formarán el pronóstico promedio, sino se utiliza DMS todos los modelos formarán parte del pronóstico promedio.



```
Tomcat v6.0 Server at localhost [Apache Tomcat] C:\Program Files\Java\jdk1.6.0_23\bin\javaw.exe (26/09/2012 22:48:15)
Resultado de Kruskal Wallis
H = (3.341687552213868E-4*24908.115018315017)/ 1.0 = 8.32351379058146
Llamar a Metodo DMS ...
```

Figura 28. Resultado de Kruskal Wallis

Conjuntos de modelos de aprendizaje formados por el método DMS.

```

Tomcat v6.0 Server at localhost [Apache Tomcat] C:\Program Files\Java\jdk1.6.0_23\bin\javaw.exe (26/09/2012 22:48:15)
.....Se han formado los siguientes conjuntos.....
Conjunto 0
MMPFC -> Reglas M5
MMPFC -> Perceptron multicapa
MMPFC -> Regresor Lineal
MMPFC -> Maquinas de Soporte Vectorial
MMPF -> Reglas M5
MMPF -> Perceptron multicapa
MMPF -> Regresor Lineal
MMPF -> Maquinas de Soporte Vectorial
MMPFC -> Reglas M5
MMPFC -> Perceptron multicapa
MMPFC -> Regresor Lineal
MMPFC -> Maquinas de Soporte Vectorial
Metodo HoltWinter -> Modelo HoltWinter
Conjunto 1
MMPFC -> Reglas M5
MMPFC -> Perceptron multicapa
MMPFC -> Regresor Lineal
MMPFC -> Maquinas de Soporte Vectorial
MMPF -> Reglas M5
MMPF -> Perceptron multicapa
MMPF -> Regresor Lineal
MMPF -> Maquinas de Soporte Vectorial
MMPFC -> Reglas M5
MMPFC -> Perceptron multicapa
MMPFC -> Regresor Lineal
MMPFC -> Maquinas de Soporte Vectorial
Metodo HoltWinter -> Modelo HoltWinter
Conjunto 2
MMPFC -> Perceptron multicapa
MMPFC -> Reglas M5
MMPFC -> Regresor Lineal
MMPFC -> Maquinas de Soporte Vectorial
MMPF -> Reglas M5
MMPF -> Perceptron multicapa
MMPF -> Regresor Lineal
MMPF -> Maquinas de Soporte Vectorial
MMPFC -> Reglas M5
MMPFC -> Perceptron multicapa

```

Figura 29. Resultados de DMS

Selección del conjunto del cual se formará el pronóstico promedio.

Tomcat v6.0 Server at localhost [Apache Tomcat] C:\Program Files\Java\jdk1.6.0_23\bin\javaw.exe (26/09/2012 22:48:15)							
No de conjuntos 14							
Mayor promedio 785.7805760991669							
Indice del mayor promedio 11							
Modelo Regresor Lineal	com.cfe.pdm.modelosdepronostico.RegresorLineal	MMPFC					
-32.11564248396371	4685.564030867268	79.3171684697381	90.71306134584765	118.54416865438975	117.0550909410593	-0.5218601472277093	
Modelo Reglas M5	com.cfe.pdm.modelosdepronostico.ReglasM5	MMPFC					
149.32562240577056	155.47269179144848	158.10639612324212	161.44273141192446	164.4352251006029	165.6292204613938	170.40526472054506	
Modelo Perceptron multicapa	com.cfe.pdm.modelosdepronostico.PerceptronMulticapa	MMPFC					
159.80423501950295	162.85774047237402	162.76395835998284	165.59345345117524	166.1935524525704	166.80317996144126	169.42384923931724	
Modelo Regresor Lineal	com.cfe.pdm.modelosdepronostico.RegresorLineal	MMPFC					
159.9573954050634	162.05867731856756	168.47257915822945	168.2150305334004	175.39250435711244	177.05946947414074	181.1231037967218	
Modelo Maquinas de Soporte Vectorial	com.cfe.pdm.modelosdepronostico.MaquinasDeSoporteVectorial	MMPFC					
158.3528935143654	160.64661498751548	165.3728558387395	168.18242358952608	173.05598910298005	175.27894888849335	180.6339644036252	
Modelo Reglas M5	com.cfe.pdm.modelosdepronostico.ReglasM5	MMPFC					
149.32562240577056	150.38048592838777	153.17246961679245	154.22733313940967	157.01931682781446	158.07418035043167	160.86616403883636	
Modelo Perceptron multicapa	com.cfe.pdm.modelosdepronostico.PerceptronMulticapa	MMPFC					
159.80423501950295	161.75500828727833	164.50062410854326	168.89373817306353	167.87286182742722	168.76591705510515	170.1027730353334	
Modelo Regresor Lineal	com.cfe.pdm.modelosdepronostico.RegresorLineal	MMPFC					
159.9573954050634	166.03344140433973	169.42515879110798	173.5350512917458	179.80193719575166	183.12521642199886	187.04603202951995	
Modelo Maquinas de Soporte Vectorial	com.cfe.pdm.modelosdepronostico.MaquinasDeSoporteVectorial	MMPFC					
158.3528935143654	160.7930322953666	165.5434114173214	168.2463523294099	173.24999470795257	175.44033361430425	180.78604766421535	
Modelo Reglas M5	com.cfe.pdm.modelosdepronostico.ReglasM5	MMPFC					
141.12987584880102	-7369.196491079061	164.71121193915891	166.67258351409222	0.13530277257826562	173.03275599246118	162.53406107585528	
Modelo Perceptron multicapa	com.cfe.pdm.modelosdepronostico.PerceptronMulticapa	MMPFC					
16.26042097249487	-31.147444482538106	33.02104352505344	40.9301050966569	35.02969653005465	31.029155549971463	26.38747517304138	
Modelo Maquinas de Soporte Vectorial	com.cfe.pdm.modelosdepronostico.MaquinasDeSoporteVectorial	MMPFC					
46.55444131250953	860.990406298614	21.345804387156996	35.009171806142795	27.33158650585358	30.042280160457118	35.823370115364526	
Modelo Modelo HoltWinter	com.cfe.pdm.modelosdepronostico.ModeloHoltWinter	Metodo HoltWinter					
163.69942053878074	162.065197595766	173.84986746396856	172.2156445209538	184.0003143891564	182.3660914461416	194.1507613143442	

Figura 30. Modelos que forman el conjunto para obtener el pronóstico

Se obtiene promedio de los pronósticos y de los ajustes del conjunto seleccionado.

Tomcat v6.0 Server at localhost [Apache Tomcat] C:\Program Files\Java\jdk1.6.0_23\bin\javaw.exe (26/09/2012 22:48:15)	

PRONOSTICOS	
2010	---122.339138504013
2011	---31.67818261411022
2012	---136.89250378454113
2013	---140.3751292394883
2014	---132.46634234032646
2015	---146.43860310133843
2016	---139.90469313168882
2017	---141.4255052958014
2018	---148.63600331290706
2019	---147.46989616202657
2020	---153.97774807705517
2021	---85.29528187102136
2022	---159.63655667234596
2023	---170.71080199354958
2024	---170.1239760646925
2025	---169.95285975797034
AJUSTES	
1995	---NaN
1996	---NaN
1997	---77.37706807131192
1998	---85.11031055079224
1999	---93.08592373864053
2000	---102.47541102152589
2001	---106.8820171745714
2002	---113.92446081931236
2003	---118.28273486482406
2004	---124.29342016244776
2005	---127.42756732386428
2006	---134.62911021457546
2007	---138.42112246020054
2008	---143.98296584366597
2009	---146.8091127960851

Figura 31. Pronóstico Promedio

Se calcula el intervalo de confianza para determinar la confiabilidad del pronóstico obtenido en el paso anterior.

```

Tomcat v6.0 Server at localhost [Apache Tomcat] C:\Program Files\Java\jdk1.6.0_23\bin\javaw.exe (26/09/2012 22:48:15)
* * * * * INTERVALO DE CONFIANZA DE PRONOSTICO * * * * *
122.339138504013 [103.85286328200814, 144.11605358691156]
-31.67818261411022 [NaN, NaN]
136.89250378454113 [118.24874038786237, 158.4757480792932]
140.3751292394883 [121.6982929160962, 161.9182688338003]
132.46634234032646 [113.86699695021015, 154.10375546039356]
146.43860310133843 [127.70776889254122, 167.91667933933965]
139.90469313168882 [121.23223267996067, 161.45312783229332]
141.4255052958014 [122.73899881320908, 162.95695534075233]
148.63600331290706 [129.88663951956732, 170.09187059232747]
147.46989616202657 [128.73029988467786, 168.9374630022701]
153.97774807705517 [135.1854507569226, 175.3823859751931]
85.29528187102136 [67.43398913868594, 107.88750899037731]
159.63655667234596 [140.80178170101905, 180.99082212124222]
170.71080199354958 [151.8007305544964, 191.97653272702055]
170.1239760646925 [151.21765820102476, 191.39409759662522]
169.95285975797034 [151.0476411145608, 191.2242675673798]

```

Figura 32. Resultado del intervalo de confianza

Escenario Correlación

La matriz de correlación también forma parte del módulo herramientas estadísticas, para visualizar los resultados de ésta, se accede a la interfaz gráfica Estadísticos.

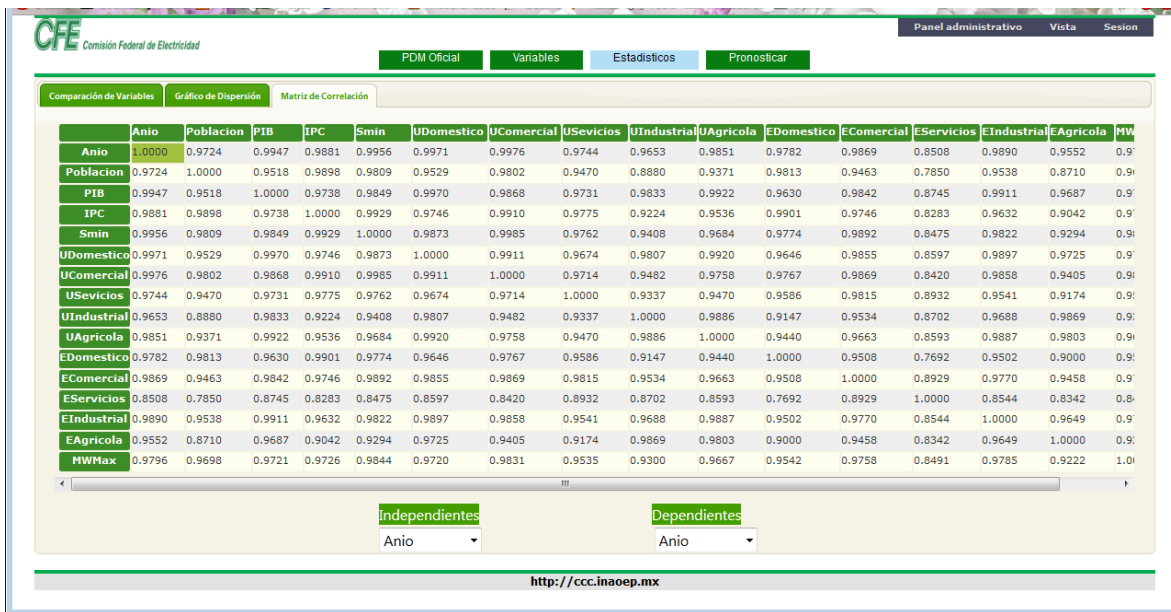


Figura 33. Interfaz de Matriz De Correlación

A continuación se muestran los resultados obtenidos:

```

Problems @ Javacod Servers Internal Web Browser Progress Console
Tomcat v6.0 Server at localhost [Apache Tomcat] C:\Program Files\Java\jdk1.6.0_23\bin\javaw.exe (26/09/2012 22:48:15)
* * * * * Variables para la Correlación * * * * *

Anio
Poblacion
PIB
IPC
Smin
UDomestico
UComercial
USevicios
UIndustrial
UAgricola
EDomestico
EComercial
ESevicios
EIndustrial
EAgricola
Máx

* * * * * MATRIZ DE CORRELACION * * * * *

1.0      0.9724391186478564    0.9947282039865192    0.9881302842882114    0.995559612129968    0.9970670631605687    0.997569279628559    0.974
0.9724391186478564    1.0      0.9518109616417373    0.9898475613643373    0.9809170637122112    0.9529170545047732    0.9801742358272302    0.947
0.9947282039865192    0.9518109616417373    1.0      0.9737813159797574    0.9848805629970281    0.9970321474735168    0.9868303823409268    0.973
0.9881302842882114    0.9898475613643373    0.9737813159797574    1.0      0.9929282406346178    0.9745620376290044    0.9909716310538249    0.977
0.995559612129968    0.9809170637122112    0.9848805629970281    0.9929282406346178    1.0      0.9873403788559717    0.9985266817154377    0.976
0.9970670631605687    0.9529170545047732    0.9970321474735168    0.9745620376290044    0.9873403788559717    1.0      0.9911257573835764    0.967
0.997569279628559    0.9801742358272302    0.9868303823409268    0.9909716310538249    0.9985266817154377    0.9911257573835764    1.0      0.971
0.974391376093292    0.9470162249723552    0.9730561596520764    0.9775049099713295    0.9762125265266459    0.967428418838616    0.97141472240
0.9652563388186117    0.8880479870751127    0.9833468022208305    0.9224002149248968    0.9408183769591655    0.9806537372914087    0.94818824312
0.9851371820434045    0.9371258427251554    0.9921989565031797    0.9535656241565704    0.9684292894279365    0.9920232555106471    0.97578129285
0.9781982547533816    0.9813192401117854    0.9630138040565697    0.9900933790015417    0.9774018464765735    0.9645924187469647    0.97671548701
0.9869364564349759    0.9462541687735231    0.9842219343050694    0.974542664547255    0.9891971679820822    0.9854870692933254    0.98693562776
0.8508422614942539    0.784991644470317    0.87452111935939    0.8282702630199833    0.8474829694706192    0.8506947462131664    0.84197997571
0.989037865094182    0.9538054239511866    0.9911213420784574    0.963167858812031    0.9822029034688021    0.9897276922842912    0.98577331864
0.9551776012469003    0.8718383364757136    0.9686997864622496    0.904166640614347    0.929425815624858    0.9725286086908224    0.94052161436
0.9796363190186499    0.9698410410239733    0.972085844573119    0.9725881153532117    0.9843961854940113    0.9719875190235998    0.98312071563

```

Figura 34. Resultados obtenidos por la Matriz de Correlación

Las interfaces gráficas mostradas anteriormente fueron implementadas por el equipo de desarrollo de INAOE-LVC.

CONCLUSIONES.

Para llevar a cabo este trabajo se recabó información sobre los distintos métodos de obtener pronósticos, las diversas medidas de error que permiten determinar qué tan preciso resulta el pronóstico y de los métodos que permiten determinar las relaciones que existen en las variables de entrada.

Se optimizó la forma de obtener el pronóstico para la CFE ya que la forma en que obtienen el pronóstico promedio es mediante un proceso que involucra el Desarrollo de Mercado Eléctrico sin embargo es poco confiable, con la mejora se redujo la incertidumbre dado que se utilizó un algoritmo de optimización con el cual se optimizan los parámetros de Holt-Winters y un algoritmo de selección de variables encargado de seleccionar las variables más relevantes para obtener el pronóstico, con ello se reduce el riesgo de ocasionar pérdidas económicas a la empresa.

La utilización de una metodología ágil DSDM durante el desarrollo del módulo Herramientas estadísticas permitió culminar el módulo satisfactoriamente, dado que se determinó las prioridades de cada parte del módulo para posteriormente realizar entregas periódicas y funcionales del mismo, además de que permitió que si en algún momento no se estaba de acuerdo con la funcionalidad del entregable se podía regresar a la versión anterior.

Se cumplió con el objetivo de implementar el módulo de Herramientas Estadísticas con ello se permite obtener el pronóstico promedio de la demanda máxima de potencia eléctrica.

Trabajo a futuro

Se agregará mayor funcionalidad al sistema pronóstico de demanda máxima de potencia eléctrica, se obtendrá el pronóstico a un menor nivel. Es decir actualmente se obtiene el pronóstico a nivel zona con este resultado por medio de diversos métodos se desagregará y obtendrá el pronóstico a nivel circuito y banco.

REFERENCIAS

- [BARR07] A. F. Barrientos, J. Olaya, V. M. González, “Un modelo spline para el pronóstico de la demanda de energía eléctrica”. Revista Colombiana de Estadística, Vol. 30, No. 2, Diciembre 2007.
- [FERN07] L. A. Fernández Jiménez, “Modelos Avanzados para la predicción a corto plazo de la producción eléctrica en parque eólicos”, Universidad de la Rioja. Junio de 2007.
- [ROJ] J.L. Rojas Rentería, R. Luna Rubio, “Predicción del consumo de energía eléctrica usando redes neuronales”. Universidad Autónoma de Querétaro
- [VELA12] J. D. Velásquez, C. J. Franco, H. A. García, “Un modelo no lineal para la predicción de la demanda mensual de electricidad en Colombia”. Vol. 25, No. 112, Septiembre 2012
- [SENA] C. Senabre, S. Valero, A. Gabaldon, M. Ortiz, M. González. “Predicción de la Demanda Total de Energía Eléctrica Española utilizando mapas auto-organizados como modelo de red neuronal”, Universidad Miguel Hernández de Elche.
- [RAMI04] Ramírez Castaño Samuel. “Redes de distribución de energía”. Universidad Nacional de Colombia. Tercera edición. 2004
- [ENRI02] Enríquez Harper Gilberto. “Elementos de diseño de subestaciones eléctricas”. Editorial Limusa. Segunda Edición, 2002
- [CANT06] Cantú Gutiérrez Paz Vicente, Jiménez Meza Obed Renato. “Líneas de transmisión y distribución”. Departamento de iluminación y alta tensión. 2006
- [ACHU98] Achuri Holguin Sixto Humberto. “Apuntes generales sobre las redes eléctricas de distribución” Tesis. Universidad Pontificia Bolivariana. 1998
- [DIDD03] Didden Marcel. “Guía de calidad de la energía eléctrica”. Katholieke Universiteit Leuven, 2003
- [EPPE00] G.D Eppen, F.J. Gould, C.P. Schmidt, J.H. Moore, L.R. Weatherford. “Investigación de operaciones en la ciencia administrativa”. Editorial Prentice Hall. Quinta Edición, 2000
- [ANDE04] David Ray Anderson, Dennis J. Sweeney. “Métodos cuantitativos para los negocios”. Editorial Prentice Hall. Novena Edición, 2004.

- [PARE03] Vélez Pareja Ignacio. “Decisiones empresariales bajo riesgo e incertidumbre”. Grupo editorial norma. 2003.
- [CARI98] Caridad y Ocerín José María. “Econometría: Modelos econométricos uniecuacionales”. Editorial Reverté. 1998
- [CFE12] Comisión Federal de Electricidad, Generación de electricidad, [en línea]. Disponible en:
<http://www.cfe.gob.mx/sustentabilidad/publicaciones/genElectricidad/Paginas/Generaciondeelectricidad.aspx> , [Fecha de consulta: Marzo 2012].
- [CES98] Cesar, La energía eléctrica, [en línea]. Disponible en:
<http://www.escolar.com/article-php-sid=31.html>, [Fecha de consulta: Marzo 2012]
- [RIV12] Rivera Francisco, Sistemas de distribución eléctrica, [en línea]. Disponible en:
http://issuu.com/artureh/docs/revista_sistema_de_distribucion_francisco_rivero, [Fecha de consulta: Febrero 2012]
- [ICF12] Ingeniería Responsable -Comisión Federal de Electricidad, La importancia del capital intelectual en el desarrollo de infraestructura del país, [en línea]. Disponible en:
<http://www.ingenieriaresponsable.mx/Ponencias/Tema%202/CONFERENCIA%2004.pdf>, [Fecha de Consulta: Abril 2012]

Apéndice A. Reglas de Usabilidad

Datos técnicos en cuanto al módulo desarrollado, con ello facilitar las actualizaciones del mismo en caso de requerirlas.

Se utilizara Eclipse JEE para la ejecución del módulo, servidor Apache Tomcat.

Para obtener el pronóstico promedio recibe como parámetros una lista de métodos de pronóstico, independientemente del procedimiento que se utilice para obtener el pronóstico estos métodos deben de tener pronósticos y ajustes, estos métodos deben tener modelos dichos modelos no necesariamente deben ser regresores de Weka, sin embargo también deben tener pronósticos y ajustes.

Para calcular las medidas de precisión de cada modelo de pronóstico únicamente se requiere tener los ajustes que es un List de QuantitativeOutputVariable de dicho modelo y los valores del histórico de la demanda máxima de potencia eléctrica, de estos cálculos solo se obtiene un valor double que corresponde al tipo de error calculado.

Para obtener la correlación de las variables se recibe como parámetros una lista de QuantitativeInputVariable que contiene el año y valor correspondiente de esa variable en ese año, se obtiene una matriz con los valores de la correlación entre todas las variables.