



San Marcos

MIEMBRO DE LA RED
ILUMINO

MECANISMOS DE TOMA DE DECISIONES



San Marcos

MIEMBRO DE LA RED
ILUMNO

MECANISMO DE TOMA DE DECISIONES

MÉTODOS DE ANÁLISIS DE *DATA MINING* SUPERVISADOS Y NO SUPERVISADOS

Los procesos completos de extracción de información de los datos que recogen un conjunto de hechos (una base de datos) y los patrones son expresiones que describen un subconjunto de los datos (un modelo aplicable a ese subconjunto).

Las datos supervisadas involucran un proceso iterativo e interactivo de búsqueda de modelos, patrones o parámetros. Los patrones descubiertos han de ser válidos y novedosos para el sistema (para el usuario siempre que sea posible) y potencialmente útiles; ya que es una técnica que permite simplificar o resumir la información para llevar a cabo una evaluación inicial, más bien generalizada. Se basa en la combinación de objetos con rasgos similares dentro de un único objeto de la búsqueda de determinar el proceso de mercadeo de una empresa. (Macao, 2016, párr. 5)

Esta técnica permite procesar mayor cantidad de datos y de información, pero tiene como desventaja que el nivel de detalle decrece y pueden no contemplarse patrones importantes durante el proceso de recolección de datos.

De acuerdo con Macoe (2016),

“

En la abstracción, el análisis se ejecuta sobre los datos resumidos, pero debe disponerse siempre de la información original del mercado y de la industria. Este método puede aplicarse no solo a datos numéricos, sino también a datos en forma cualitativa. (párr. 5)

”



LAS CONSIDERACIONES CON LAS CUALES SE ESTABLECE ESTOS MECANISMOS SON:

- **La búsqueda de no perder dinero dentro de la empresa:** Que siempre el costo de implementación sea menor que las mejoras obtenidas, produciéndose así un retorno de la inversión del mercado e industria.
- **Obtención de la información de los resultados de forma ágil:** Esto quiere indicar que generalmente se espera obtener resultados dentro de un período de tiempo razonable para la toma de decisiones. Si este se hace muy extenso debería retornarse al inicio y realizar los cambios que se consideren necesarios.
- **El acceso y la implementación de los datos:** En este proceso no es imprescindible un acceso *On-line* a las fuentes de datos obtenidas, ya que *Data Mining* no se realiza en tiempo real, pero sí se hace necesario el acceso a toda la información para la ejecución del análisis.
- **El empleo del sistema de información recolectado:** En la mayoría del tiempo nunca se utiliza una simple aplicación, sino más bien una combinación de técnicas y metodologías para la toma de ediciones dentro del periodo necesarios para la tomad de decisiones. (Morineau, 2000, p. 323)
- **Las tomas de decisiones en las empresas o en las organizaciones:** Como una de las reglas generales, nunca una herramienta sirve solo para proporcionar la solución buscada, sino simplemente ayuda a encontrarla, como tal lo indica la palabra herramienta. Encontrar la solución al problema propuesto y la total responsabilidad de la toma de decisiones se deposita sobre el o los profesionales que realizan el análisis, esto quiere decir que es para toma de decisiones de la empresa.



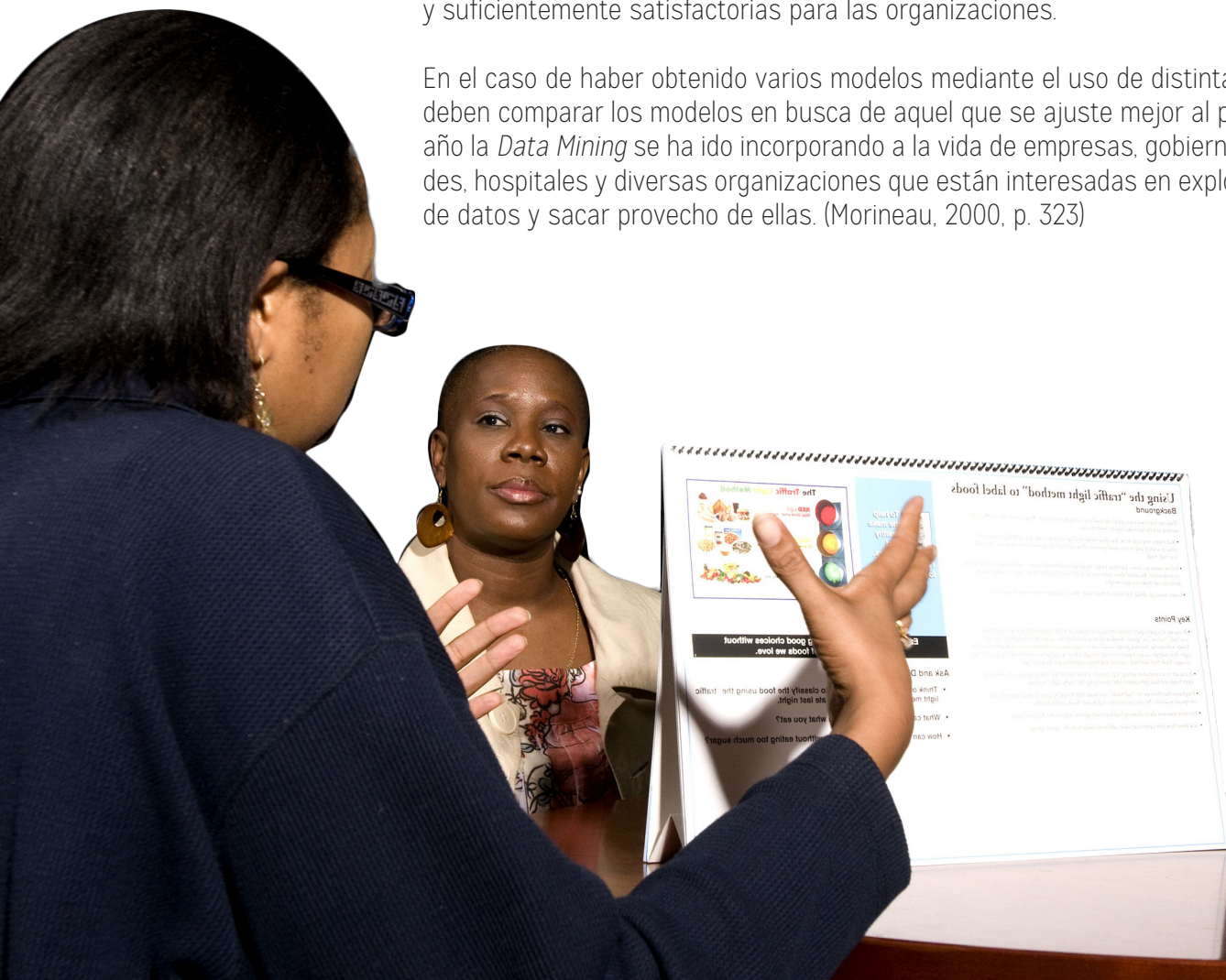
Los datos como tales no pueden ser manejados directamente por *Data Mining*, es necesario modelarlos llevándolos a un formato tal que sí pueda ser empleado, y el desarrollo de dicho modelo es decisivo, ya que determina los tipos de resultados que se pueden obtener.

ESTACIONALIDAD CONSISTE EN FLUCTUACIONES PERIÓDICAS, POR EJEMPLO, EN UNA TIENDA DE PRODUCTOS VARIOS CUANDO HAY PICOS DE VENTAS EN LA NAVIDAD Y DESPUÉS DECLINAN.

Este modelado de datos asume una estructura orientada a objetos, donde la información está representada por objetos, sus atributos descriptivos y las relaciones entre las clases de objetos. Una

vez obtenido el modelo que se desea dentro de la empresa, se debe proceder a su validación de la información, comprobando que las conclusiones que arroja sean válidas y suficientemente satisfactorias para las organizaciones.

En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Cada año la *Data Mining* se ha ido incorporando a la vida de empresas, gobiernos, universidades, hospitales y diversas organizaciones que están interesadas en explorar sus bases de datos y sacar provecho de ellas. (Morineau, 2000, p. 323)



FORECAST

En toda la empresa tiene que hacer siempre una estimación de mercado de sobre cuáles serán sus ventas durante un periodo. Estos mecanismos son la principal fuente de ingresos de las compañías y por lo tanto los que determinan su futura viabilidad del negocio. Además, que la compañía este basada en la previsión de ventas que van a depender de todas la estructura de los costos. Esto va a implicar entre más ventas esperadas, mayor será la inversión que se querrá hacer durante un futuro cercano.

De acuerdo con Zhao, Zhu, Zhang, Zhang, Li y He (2011) se deben tomar en cuenta unos procesos:

- En los casos de estimar unas ventas superiores a las finalmente reales del mercado o la industria, las empresas se encontrarán que ha gastado más de lo que debería de gastar durante ese período. Si la diferencia es muy grande, hay el peligro de incurrir en pérdidas durante el período o el futuro cercano de las empresas.
- Si se estiman ventas inferiores durante un periodo determinado a las que acaban ocurriendo, la empresa habrá perdido la oportunidad de ganar las cuotas de mercado y habrá dejado sin abastecer sus clientes y socios. Antes esta falta de manejo, puede que muchos compradores potenciales opten por no esperar y elijan un producto de la competencia lo cual desbalancea la compañía a un futuro corto.

Por estos motivos el *forecast* consiste en la estimación y monitorización de las ventas futuras para un producto dentro de un mercado bien organizado, utilizando diferentes herramientas matemáticos y de mercadeo, como los que se sacan con los datos históricos de venta y las previsiones del equipo de ventas y del departamento de *marketing*. Lo cual tiene como objetivos:

1. Descubrir los problemas con anterioridad para tomar una decisión.
2. Buscar soluciones rápidamente para restablecer la calma a la empresa.
3. Preparar las áreas de la organización para las acciones a emprender durante un período o la continuidad de un nuevo mecanismo de tomas de decisiones.
4. Estar al tanto en el corto y medio plazo (no se hacen estimaciones más allá del año debido al alto grado de cambio en los mercados en la actualidad).
5. Perfeccionar el flujo de información en la cadena de suministro.
6. Rebajar inventarios durante periodos de la empresa.
7. Ofrecer una atención eficiente al cliente durante la vida productiva de la compañía.

SE DIFERENCIAN DOS MÉTODOS BÁSICOS PARA REALIZAR EL ANÁLISIS:

En la búsqueda de los procesos cuantitativos que se basan en datos históricos de la empresa de las ventas que ya se disponen durante un periodo. (p. 147)

Hay que tener presente que ventas pasadas no presuponen ventas futuras, pero pueden ser una base para predecir lo que pasará en un futuro próximo. Igualmente hay que tener en cuenta los siguientes factores:

- **Cambios en el mercado:** puede que en el mercado haya aparecido alguna oportunidad o amenaza que afecta a todo el sector y que afecte las ventas.
- **Cambios en la propia empresa:** en la misma compañía pueden haber sucedido determinados hechos que la hayan fortalecido o, por el contrario, debilitado respecto la competencia.
- **Contexto económico:** hay que tener en cuenta la situación general del mercado al cual dirigirse para saber si disponen de la misma capacidad adquisitiva que antes.
- **Cambios en las preferencias:** las modas y preferencias son muy cambiantes en la actualidad, y aquello que antes era valorado, hoy ya no. Por lo tanto, es imprescindible estar atentos a estos cambios.

Al hablar de procesos cualitativos se pueden definir que según evaluaciones subjetivas

EL DATA MINING SON HERRAMIENTAS QUE PUEDE SER APLICADAS EN DISTINTOS ENTORNOS COMO: EL GOBIERNO, LA EMPRESA COMERCIALES, LA UNIVERSIDAD, LOS DEPORTES, OTROS.

de todas aquellas personas que conforman la empresa que han tenido algún tipo de contacto con el consumidor final o en alguno de los procesos necesarios para hacer llegar el producto al punto de venta. En este sentido, toma mucha importancia la figura del comercial. El comercial, la principal fuente de información. (Hernández, Ramírez y Perri, 2004, p. 245)

La mayoría de veces se tiende a valorar a una persona que trabaja como comercial solamente por sus resultados en ventas. Esta valoración tan simplista provoca que se dejen muchos cabos sueltos respecto a la que tendría que ser la totalidad de sus tareas. El comercial es la persona que tiene el contacto más directo con el consumidor final. Por lo tanto, tiene que conseguir el máximo de información de él para conocer cuáles son sus preferencias y qué puntos valora más y menos del producto que se le está ofreciendo. De este modo, la empresa podrá:

- 1 PODER SOLUCIONAR AQUELLOS PUNTOS QUE HA PASADO DESAPERCIBIDOS.**
- 2 ADAPTARSE A LAS NECESIDADES DEL CONSUMIDOR.**
- 3 TENER UNA BASE DE DATOS ACTUALIZADA QUE PUEDE SER VITAL PARA EL LANZAMIENTO DE FUTUROS PRODUCTOS.**

Razón por la cual es imprescindible pedir a los vendedores que, de cada contacto con el público que realicen, realicen un informe. Ellos son la principal fuente de conocimiento del cliente y es básico hacerles entender de la importancia de esta tarea, no solamente para la empresa, también para ellos mismos. Si la compañía consigue enfocar correctamente el producto, este será más atractivo y la tarea de venta, más sencilla. A más ventas, más ingresos para la empresa y, a la vez, más comisiones para los comerciales. Por lo tanto, todas las partes salen ganando con la aplicación del forecast.

La minería de datos incluye áreas de conocimiento tales como Estadística, Inteligencia Artificial y Bases de datos; el análisis de estos datos pueden surgir ventajas competitivas o novedosas soluciones a antiguos problemas, la minería de datos es una de gran actividad a nivel académico como lo demuestran el gran número de eventos científicos relacionados, como así también laborales.

En estos procesos existen problemas en la recolección de la información, los cuales pueden ser de muchas naturalezas.

PROBLEMAS QUE PUEDEN SURGIR

Es muy importante analizar correctamente los datos una vez agrupados ya que aun así, se pueden dar los siguientes problemas o desviaciones.

- **Desviaciones respecto la media general en algunos resultados:** puede que los datos presentados por algún comercial disten, en relación con las preferencias de los consumidores finales, de los que ha conseguido la mayoría. Es imprescindible analizar a qué se debe tal desviación. Si se trata de un caso aislado, no se toma en consideración y se aparta del análisis.
- **Diferentes formas de realizar el *forecast*:** cada comercial y cada área de la empresa tiene que enfocar la realización de esta tarea siguiendo unos mismos patrones y parámetros. En caso contrario, es imposible realizar la extrapolación y análisis de los datos.
- **Existe poca comunicación entre departamentos:** es imprescindible que todas las áreas que conforman la empresa tomen como suya la realización del forecast. La información que dispone un área puede afectar y ser importante para otras, por lo que tienen que existir canales de comunicación que permitan una circulación ágil y fiable de toda la información y conocimiento que hay en la empresa.
- **Inadecuada o poca atención al cliente:** es importante disponer de un servicio completo capaz de recopilar los datos referentes a los clientes. Los comerciales los transmiten y tiene que haber un grupo capaz de agruparlos y de ordenarlos. Del mismo modo, el servicio post venta es clave en el conocimiento de las preferencias de los clientes, ya que nos permite saber en qué falla el producto y solucionarlo.



ANÁLISIS DE SERIES DE TIEMPO, TÉCNICAS Y MODELOS

En cada una de las instituciones, ya sea las familias, las empresas o los gobiernos a nivel mundial, en todos se tienen que hacer planes para el futuro si ha de sobrevivir y progresar en las empresas y de organizaciones. En este mundo globalizado en las diversas instituciones requieren conocer el comportamiento futuro de ciertos fenómenos con el fin de planificar, prever o prevenir posibles riesgos en sus transacciones.

SERIE DE TIEMPO SON LOS DATOS ESTADÍSTICOS QUE SE RECOPILAN, OBSERVAN O REGISTRAN EN INTERVALOS DE TIEMPO REGULARES, COMO LOS DIARIOS, SEMANALES, SEMESTRAL, ENTRE OTROS.

El análisis de tiempo es un conjunto de mediciones de cierto fenómeno o experimento que registradas secuencialmente en el tiempo, por ejemplo, se ve cuando se establece por cada hora, mensualmente, trimestralmente, semestralmente, y otras medidas.

Un supuesto en muchas técnicas de series de tiempo es que los datos son estacionarios, donde su media, variancia y auto correlación no cambia en el tiempo, tampoco se presentan patrones de estacionalidad; sin embargo, en la práctica sí se presentan estos patrones de tendencia y de estacionalidad y es necesario contar con modelos que las consideren.

Los principales atributos que tienen la minería de datos son:

- Los atributos definen los comportamientos de una clase de objeto ya determinada organización.
- Los valores de los atributos es una medida de la cantidad que dicho atributo se refiere a un objeto.
- Colaboran a diferenciar los objetos individuales dentro de una clase o determinado proyecto.
- Para caracterizar y desarrollar un objeto cualquiera, con cada atributo deber tener un único valor dentro del mercado. Si en el set o portafolios de datos hay más de un valor, no se puede determinar cuál es el correcto. Esto generalmente ocurre cuando dichos valores se toman en distintos intervalos de tiempo.
- Y para los atributos que cambian con el tiempo, se deben plantearse un modelo diferente en la marca y esta situación se conoce como “análisis basado en estados” (*state-based*).

Para la planificación racional exigen prever los sucesos del futuro que probablemente vayan a ocurrir dentro de la industria. La previsión se suele basar en lo que ha ocurrido en la organización. Se tienen un nuevo tipo de inferencia estadística que se hace acerca del futuro de alguna variable o compuesto de variables basándose en sucesos pasados para tomas de decisiones. La técnica más importante para hacer inferencias sobre el futuro con base en lo ocurrido en el pasado, es el análisis de series de tiempo con sus herramientas matemática nos colaboran en sus interpretaciones y análisis. (Vinnakota, 2006, p. 45)

Unos de los problemas que intenta resolver las series de tiempo es el de predicción del mercado bajo las condiciones reales del mercado. Esto es dado una serie $\{x(t_1), \dots, x(t_n)\}$ y nuestros objetivos de interés son describir el comportamiento de la serie y sucesiones matemáticas, investigar el mecanismo generador de la serie temporal, buscar posibles patrones temporales que permitan sobrepasar la incertidumbre del futuro.

El primer paso en el análisis de series de tiempo consiste en graficar la serie. Esto nos permite detectar las componentes esenciales de la serie.

Cuando se entregan los resultados es importante que mediante este proceso las gráficas de la serie permitan:

- **Detectar en la operativa *Outlier***, en este proceso que se refiere a puntos de la serie que se escapan de lo normal del mercado. En realidad es una observación de la serie que corresponde a un comportamiento anormal del fenómeno (sin incidencias futuras) o a un error de medición de los datos obtenidos. Se debe determinar desde fuera si un punto dado es outlier o no. Si se concluye que lo es, se debe omitir o reemplazar por otro valor antes de analizar la serie. (Kulldorff, 1995, p. 800)
- **Detectar las tendencias del mercado y de los consumidores:** la tendencia representa el comportamiento predominante de la serie y las influencias del mercado. Esta puede ser definida vagamente como el cambio de la media a lo largo de un periodo. (Kulldorff, 1995, p. 800)
- **Las principales variaciones estacionales de los productos de la industria:** las variaciones estacionales representan un movimiento periódico de la serie de tiempo de cada uno de los productos dentro del mercado. La duración de la unidad del periodo es generalmente menor que un año (corto plazo). Puede ser un trimestre, un mes o un día. (Kulldorff, 1995, p. 800)

A nivel matemático lo podemos decir que la serie representa variación estacional si existe un número s tal que $x(t) = x(t + ks)$. Donde las principales fuerzas que causan una variación estacional son las condiciones del tiempo.



San Marcos

MIEMBRO DE LA RED
ILUMNO

PREDICCIÓN BASADA EN MÉTODOS LINEALES Y NO LINEALES

PREDICCIÓN BASADA EN MÉTODOS LINEALES

De este modo, la operación del sistema hace posible que el mercado funcione y aporta una liquidez absoluta a distribuidores y consumidores.

Por lo cual las herramientas para las decisiones tecnológicas tales como los mode-

“UN PROCESO NO TRIVIAL DE IDENTIFICACIÓN VÁLIDA, NOVEDOSA, POTENCIALMENTE ÚTIL Y ENTENDIBLE DE PATRONES COMPENSIBLES QUE SE ENCUENTRAN OCULTOS EN LOS DATOS” (FAYYAD, ET ÁL. 1996, P. 121)

los matemáticos han sido aplicadas a una amplia gama de situaciones en la toma de decisiones dentro de diversas áreas de la gerencia y en el mercadeo de productos. En la toma consciente de decisiones bajo incertidumbre, siempre realizamos pronósticos o predicciones. Podríamos pensar que no estamos

pronosticando, pero nuestras opciones estarán dirigidas por la anticipación de resultados de nuestras acciones o inacciones. Este sitio tiene el objetivo de ayudar a los gerentes y administradores a hacer un mejor trabajo al momento de anticipar hechos, y por lo tanto, un mejor manejo de la incertidumbre mediante el uso de técnicas de predicción y pronóstico efectivas. (Kulldorff, 1995, p. 817)

El esquema de funcionamiento seguido por el operador serial el siguiente. En primer lugar, debería averiguar el valor que toma la demanda en cada momento, tratando de llegar a la denominada curva de carga. Una vez conocido este dato, el operador se dirigía a los centros de producción, indicadores que deben de introducir en la red en cada momento para garantizar la cobertura de la demanda y, al mismo tiempo, minimizar la sobreproducción y sus costes de eficiencia. Sin embargo la demanda es muy variable a lo largo de los años, los meses, los días e incluso las horas.



Estas características aportan una incertidumbre al sistema que afecta de forma significativa al coste final del suministro. Para reducir este factor de costo, el gestor necesita conocer con precisión los mecanismos que hacen variar la demanda. (Benzecri, 2015, p. 221)

Los métodos de predicción lineales son aquellos que intentan modelar el comportamiento de una serie temporal mediante una función lineal. Entre estos métodos se destacan los modelos como AR, MA, ARMA y ARIMA. (Peña, 1989, p. 123)

La de un proyecto consiste en familiarizarse con los mercados financieros a nivel globalizado o nacional, en concreto con el mercado de renta variable, y analizar y comprender su funcionamiento.

Con lo anterior, el objetivo es saber aplicar los diferentes modelos, así como conocer cuáles son las variables explicativas más representativas en función de cada uno de los conjuntos estudiados..

Los diferentes métodos de predicción existentes se dividirán en dos grupos, modelos lineales y estrategias no lineales. Así mismo, se estudiarán los diferentes métodos para analizar la estacionalidad de las series, las funciones de auto correlación, las diferentes medidas de error y los diferentes procedimientos de comparación entre métodos de predicción.

El uso de modelos matemáticos ha sido incrementado para interpretar y predecir las dinámicas y controles en la toma de decisiones gerenciales. Dichas aplicaciones incluyen pronósticos de ventas, predicciones del impacto y efecto de campañas publicitarias, estrategias para proteger desabastecimiento de inventarios y para determinar estrategias óptimas de inversión de portafolios





Cuando la serie presenta un comportamiento estacionario, es decir, no tiene tendencia y puede ser modelizada como $X_t = a + u_t$ (donde u_t es un término de perturbación aleatorio, con valor esperado cero y varianza constante para todo t , e independiente de X_t para todo t), el método de predicción adecuado es el alisado exponencial simple. Este método estima para cada período T el parámetro a como suma ponderada de todas las observaciones anteriores, dando mayor importancia a las observaciones más recientes que a las más antiguas. La expresión de cálculo es: $F_{t+1} = a D_t + (1 - a) F_t$, de donde:

- D_t e el valor actual
- F_t es el valor pronosticado
- a es el factor de ponderación, el cual oscila entre 0 y 1
- t es el período de tiempo actual

Note que el valor atenuado se convierte en el pronóstico para el período $t + 1$ y obtenida en el período $T-1$ y α es la constante de alisado que toma valores entre 0 y 1." Tomado de (Kulldorff, 1995, p. 817)

PREDICCIÓN BASADA EN MÉTODOS NO LINEALES

La tendencia no lineal cuando la serie de tiempo presenta un comportamiento curvilíneo se dice que este comportamiento es no lineal.

Dentro de las tendencias no lineales que pueden presentarse en una serie se encuentran la polinomial, logarítmica, exponencial y potencial, entre otras.

MODELOS Y ANÁLISIS DE DECISIONES CON PERIODOS DE TIEMPO CRÍTICO DEL MERCADO Y LA ALTA GERENCIA

La capacidad de modelar y realizar modelos de decisión y análisis es un rasgo esencial entre las muchas aplicaciones reales que van desde los tratamientos médicos de emergencia en las unidades de cuidados intensivos hasta en los sistemas de control de los comandos militares. Los formalismos y los métodos de inferencia existentes no han sido eficaces en las aplicaciones de tiempo real, donde las compensaciones entre la calidad de decisión y manejabilidad computacional son esenciales.



En la práctica, un acercamiento eficaz al modelamiento de decisión de tiempo dinámico crítico debería proporcionar el apoyo explícito al modelamiento de procesos temporales y al manejo de situaciones críticas de tiempo. (Lebart, 2000, p. 45)

El objetivo de las técnicas de predicción no causal es obtener estimaciones o pronósticos de valores futuros de una serie temporal a partir de la información histórica contenida en la serie observada hasta el momento actual. Estas técnicas no requieren la especificación de los factores que determinan el comportamiento de la variable, sino que se basan únicamente en la modelización del comportamiento sistemático de la serie. Se consideran tres modelos posibles del comportamiento sistemático de una serie temporal: modelo estacionario (sin tendencia), modelo con tendencia lineal y modelo con estacionalidad. La técnica de predicción adecuada dependerá del modelo de comportamiento de la serie.

El filtraje de datos es una herramienta efectiva y eficiente para el modelamiento de series de tiempo cuando se aplican las técnicas de transformación apropiadas. La mayoría de las técnicas de análisis de series de tiempo envuelven algunas formas de filtraje de ruido con el objetivo de hacer los patrones de comportamiento más obvios.

DIFERENCIACIÓN. Un tipo especial de filtraje, el cual es particularmente especial para remover tendencias, es simplemente diferenciar una serie de tiempo dada hasta que se convierta estacionaria. Este método es útil en el modelamiento de la Box-Jenkins. Para datos no estacionales, la diferenciación de primer orden es normalmente suficiente para alcanzar una estabilidad aparente, de manera tal que las nuevas series estén formadas de las series originales. (Lebart, 2000, p. 48)

FILTRAJE ADAPTATIVO. Cualquier técnica de atenuación tal como la de promedios móviles la cual incluye el método de aprendizaje por errores pasados que pueden responder a cambios en la tendencia, estacionalidad y factores aleatorios de relativa importancia. En el método adaptativo de atenuación exponencial se podría ajustar a para permitir los cambios en patrones de comportamiento. (Lebart, 2000, p. 49)

FILTRO DE HODRICK-PRESCOTT. Este es un mecanismo de atenuación utilizado para obtener los componentes de tendencia a largo plazo en las series de tiempo. Esta es una manera de descomponer una serie de tiempo dada sus componentes estacionarios y no estacionarios de tal manera que la suma de los cuadrados de la serie de los componentes no estacionarios sea mínima con una penalidad sobre los cambios derivativos de los mismos." (Lebart, 2000, p. 50)



MODELOS DE REGRESIONES LOGÍSTICAS

Se dice que un proceso es binomial cuando solo tiene dos posibles resultados: "éxito" y "fracaso", siendo la probabilidad de cada uno de ellos constante en una serie de repeticiones. Al variable número de éxitos en n repeticiones se le denomina variable binomial. Al variable resultado de un sólo ensayo y , por tanto, con solo dos valores: 0 para fracaso y 1 para éxito, se le denomina binomial puntual. (Maoe, 2016, párr. 3)

El objetivo de este modelo es:

- Ampliar los conceptos de regresión al caso en que la variable dependiente sea categórica.
- Conocer el modelo de probabilidad lineal y los problemas que este presenta a la hora de explicar el comportamiento de una variable dependiente binaria.
- Entender los modelos Logit y Probit como modelos que permiten superar las dificultades del modelo de probabilidad lineal.
- Aprender a realizar regresión logística binaria con ayuda de Minitab, interpretando correctamente los resultados generados por el programa.

Cuando se pretende explicar mediante un modelo de regresión, el comportamiento de una variable (llamada variable endógena o dependiente) en función de los valores que tomen otras (llamadas variables exógenas o explicativas), suele utilizarse un modelo de regresión lineal múltiple (MRLM o MRLG). Ahora bien, como veremos en este *math-block*, el modelo lineal presenta ciertos problemas serios cuando la variable dependiente es binaria (y, en general, categórica), lo cual nos llevará a usar modelos de regresión no lineales específicamente pensados para realizar regresión con variables categóricas. Los modelos que analizaremos aquí serán el Logit y el Probit.

EJEMPLO DE LA APLICACIÓN DEL MODELO

De acuerdo con Garrido (2012),

“

Todo proyecto de *Data Mining* tiene unas fases bien definidas que van desde la definición del problema hasta la ejecución y evaluación del modelo, pasando por el estudio de los datos y la creación de dicho modelo. Dichas fases quedarán ilustradas en los dos ejemplos que se darán a continuación. (p. 56)

”

Una empresa en posesión de unas bases de datos de calidad y tamaño suficiente puede emplear el *Data Mining* en regresiones para generar nuevas oportunidades de negocio, dada su capacidad para proporcionar:

1. PREDICCIÓN AUTOMÁTICA DE COMPORTAMIENTOS

Generalmente se trata de problemas de clasificación. Como ejemplo podemos citar el *marketing* dirigido. *Data Mining* en regresiones lógicas usa los resultados de campañas de marketing realizadas anteriormente para identificar el perfil de los clientes que son más propensos a comprar el producto y de este modo permitimos sustituir el correo masivo por el correo dirigido.

2. PREDICCIÓN AUTOMÁTICA DE TENDENCIAS

Basándonos en bases de datos históricas, *Data Mining* en regresiones creará un modelo para predecir las tendencias. Como ejemplos podemos citar la predicción de ventas en el futuro o la predicción en mercados de capitales.

3. DESCUBRIMIENTO AUTOMÁTICO DE COMPORTAMIENTOS DESCONOCIDOS ANTERIORMENTE

Las herramientas de *Data Mining* en regresiones de visualización y *clustering*, permiten «ver» nuestros datos desde una perspectiva distinta y por ello descubrir nuevas relaciones entre ellos. (Garrido, 2012, p. 57)





PREDICCIONES CON ÁRBOLES DE DECISIÓN

De acuerdo con Johnson (2013)

(...) esta técnica se encuentra dentro de una metodología de aprendizaje supervisado. Su representación es en forma de árbol en donde cada nodo es una decisión, los cuales a su vez generan reglas para la clasificación de un conjunto de datos.

Los árboles de decisión son fáciles de usar, admiten atributos discretos y continuos, tratan bien los atributos no significativos y los valores faltantes. Su principal ventaja es la facilidad de interpretación. El análisis de la crisis empresarial se configura como un aspecto clave en el análisis de estados financieros al permitir a la empresa disponer de un sistema de diagnóstico que abarque los aspectos de solvencia y de rentabilidad. La preocupación por el cumplimiento del objetivo de solvencia ha originado la elaboración de distintos modelos de predicción del fracaso empresarial obtenidos mediante el empleo de diferentes técnicas estadísticas como el análisis univariante o el multivalente. (p. 34)

El algoritmo de árboles de decisión es un algoritmo de clasificación y regresión proporcionado para el modelado de predicción de atributos discretos y continuos.

Para los atributos discretos, el algoritmo hace predicciones basándose en las relaciones entre las columnas de entrada de un conjunto de datos. Utiliza los valores conocidos como estados de estas columnas para predecir los estados de una columna que se designa como elemento de predicción. Específicamente, el algoritmo identifica las columnas de entrada que se correlacionan con la columna de predicción. Por ejemplo, en un escenario para predecir qué clientes van a adquirir probablemente una bicicleta, si nueve de diez clientes jóvenes compran una bicicleta, pero solo lo hacen dos de diez clientes de edad mayor, el algoritmo infiere que la edad es un buen elemento de predicción en la compra de bicicletas.

El árbol de decisión realiza predicciones basándose en la tendencia hacia un resultado concreto.

Para los atributos continuos, el algoritmo usa la regresión lineal para determinar dónde se divide un árbol de decisión.

Si se define más de una columna como elemento de predicción, o si los datos de entrada contienen una tabla anidada que se haya establecido como elemento de predicción, el algoritmo genera un árbol de decisión independiente para cada columna de predicción.

REFERENCIAS BIBLIOGRÁFICAS

- Alegret, M., Herrera, M. y Grau, R. (2008). Las técnicas de estadística espacial en la investigación salubrista. Revista Cubana Sal Públ [revista en la Internet]. Recuperado de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-34662008000400003&lng=es
- Barcellos, C. y Buzai, G. (2006). La dimensión espacial de las desigualdades sociales en salud: aspectos de su evolución conceptual y metodológica. México: Universidad Nacional de Luján: Anuario de la División Geografía.
- Benzecri, J. (2015). Gestión del conocimiento y minería de datos. París: Dunod.
- Berry, M. y Linoff, G. (2014). Data Mining Techniques for Marketing Sales and Customer Support. USA: John Wiley & Sons
- Chatfield, C. y Collins, A.J. (1999). Introduction to multivariate analysis. London: Chapman and Hall.
- Kamber M. (2006). Data mining: concepts and techniques. Morgan Kaufmann;
- Hernández R. (2004). USA: Centricity Solution for Marketing.
- Hoaglin D.C., Mosteller F., Tukey J.W. (2005). N.Y: Exploring Data tables. Trends and Shapes, Wiley.
- Kuldorff, M. y Nagarwalla, N. (1995). Spatial disease clusters: detection and inference. Customer Support. USA: John Wiley & Sons
- Ester, M., Kriegel, H., y Sander, J (1999). Knowledge discovery in spatial databases. KI-99. Advanc Artif Intellig.
- Jambu, M. (2000). Classification Automatique pour l'Analyse des données. París: Dunod.
- Johnson Dallas, E. (2013). Métodos multivariados aplicados al análisis de datos. México: Thomson editores.





- Johnson, R.A. y Wichern Dean, W. (2010). Applied Multivariate Statistical Analysis. 3rd De. USA: Prentice Hall Inc.
- Lebart, L., Morineau, A. y Tabard, N. (2000). Techniques de la description statistique. París: Dunod.
- Lebart, L., Morineau, A. y Piron, M. (2005). Statistique exploratoire multidimensionnelle. París: Dunod.
- Maoe, J. (2016). 5 de los mejores software de minería de datos de Código Libre y Abierto. Recuperado de: [Http://blog.jmaoe.com/gestion_ti/base_de_datos/5-mejores-software-mineria-datos-codigo-libre-abierto/](http://blog.jmaoe.com/gestion_ti/base_de_datos/5-mejores-software-mineria-datos-codigo-libre-abierto/)
- Monografía N° 27 Serie de matemática. USA: O.E.A.Washington.
- Tecnologías-información.com. (2015). Minería de Datos. Recuperado de: <http://www.tecnologias-informacion.com/mineria-de-datos.html>
- Tecnologías-información.com. (2015). Minería de Datos. Recuperado de: <http://www.tecnologias-informacion.com/mineria-de-datos.html>
- Vinnakota, S. y Lam, N. (2006). Socioeconomic inequality of cancer mortality in the United States: a spatial data mining approach. Internat J Heal Geogr.
- Zhao, F., Zhu, R., Zhang, L., Zhang, Z, Li, Y. y He. M. (2011). Application of satscan in detection of schistosomiasis clusters in marshland and lake areas. Zhongguo xue xi chong bing fang zhi za zhi. Alemania: Chin J Schistosom Contr.

