



Universidad Nacional de Asunción
Facultad Politécnica
Ingeniería en Informática
Trabajo Final De Grado

TPDBM: Técnica de pronóstico de la demanda basada en Business Intelligence
y Machine Learning

Julio 2017

1. SÍNTESIS DEL PROYECTO

1.1 TÍTULO INICIAL DEL PROYECTO

TPDBM: Técnica de pronóstico de la demanda basada en Business Intelligence y Machine Learning

SINTETIZAR EL PROYECTO EN TRES PALABRAS CLAVE

Pronóstico de la Demanda

Business Intelligence

Machine Learning

1.2 DURACIÓN DEL PROYECTO

12	MESES
----	-------

1.3 DATOS DE LOS ESTUDIANTES Y DEL ASESOR DEL PROYECTO

NOMBRE DEL ASESOR PRINCIPAL:

Ph.D. Diego Pinto

Email: dppintoroa@gmail.com

Teléfono: 0984 401 410

CARGO:

Docente

FIRMA:

NOMBRE DEL ASESOR SECUNDARIO:

Ing. Aditardo Vázquez

Email: vazquez.aditardo@gmail.com

Teléfono: 0981 358 294

CARGO:

Docente

FIRMA:

NOMBRE DEL ALUMNO:

Alberto Gabriel Garcete González

CI: 3.676.206

PLAN DE ESTUDIOS: 2008

Email: albertogarcetepy@gmail.com

Teléfono: 0971 494427

AÑO DE INGRESO: 2008

FIRMA:

NOMBRE DEL ALUMNO:

Raúl Alberto Benítez Martínez

CI: 4.089.446

PLAN DE ESTUDIOS: 2001

Email: raulkvd@gmail.com

Teléfono: 0984 196134

AÑO DE INGRESO: 2007

FIRMA:

2. RESUMEN DEL PROYECTO

El pronóstico de la demanda supone uno de los procesos más importantes y de mayor impacto económico en empresas relacionadas a las ventas minoristas. Dentro de la gestión de compras, un cálculo que regularmente debe hacerse es este pronóstico de la demanda mencionado, para estimar el volumen de productos de la orden de compra y con esto realizar la reposición de stock. En este trabajo se propone una técnica que ayuda a tomar decisiones acerca del volumen de la orden de compra para la reposición de stock. Esta técnica está basada en Business Intelligence y Machine Learning integradas, se trata de un modelo que pronostica la demanda de productos para el siguiente período de ventas, a su vez que también incluye decisiones de si realmente hay que seguir adquiriendo un determinado producto.

3. Marco Teórico

Pronóstico de la Demanda

En las empresas retail o de ventas minoristas uno de los principales desafíos que afrontan es el manejo eficiente del stock. La eficiencia se busca de manera a evitar tener excesivas cantidades de productos en los depósitos, o evitar la falta de productos denominada ruptura de stock. Los excesos incurren en sobrecostos, lo contrario conlleva a pérdida de oportunidades de ventas al no disponer del producto en el momento oportuno. Ambos repercuten directamente en las utilidades de la empresa y en la satisfacción de los clientes. El problema a resolver consiste en estimar o predecir la demanda que luego se traduce en el volumen de productos a comprar para la reposición de stock para las ventas del siguiente período [11].

Business Intelligence (KPI)

Con el presente trabajo se elabora un nuevo modelo de estimación de cantidades eficientes en las órdenes de compra de productos para la reposición de stock del siguiente periodo de venta. En este nuevo modelo se integran técnicas de Business Intelligence y Machine Learning.

En la etapa de Business Intelligence el objetivo principal es calcular los Indicadores Claves de Rendimiento (KPI - Key Performance Indicators) de los productos en base a los datos históricos obtenidos de la base de datos transaccional. Luego cada serie de KPI obtenidos pasan por un proceso de etiquetado, donde el experto en compras los analiza y determina qué nivel de compra conviene para cada serie de KPI.

Machine Learning

En la etapa de Machine Learning se utilizan como entrada las series de KPI obtenidas en la etapa de Business Intelligence. Estas entradas constituyen las instancias que alimentan los distintos algoritmos de clasificación del Machine Learning supervisado. Luego tienen lugar los procesos propios de esta etapa que son el entrenamiento y testeo para finalmente evaluar los distintos desempeños a fin de determinar los algoritmos más adecuados. Los algoritmos elegidos serán utilizados para estimar las cantidades de las órdenes de compra por cada producto.

Weka es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Los algoritmos pueden ser aplicados directamente a un conjunto de datos o llamados desde código Java. Weka contiene herramientas para pre-procesamiento de datos, clasificación, regresión, clustering, reglas de asociación y visualización. También es adecuado para desarrollar nuevos esquemas de aprendizaje automático [6]. Los algoritmos de clasificación de Weka que se utilizarán son los siguientes [7]: BayesNet, NaiveBayes, NaiveBayesUpdateable, Logistic, MultilayerPerceptron, SimpleLogistic, SMO, OneR, DecisionTable, JRip, PART, ZeroR, DecisionStump, J48, LMT, RandomForest, RandomTree, REPTree.

La evaluación es la clave para lograr avances reales en el aprendizaje automático. Una de ellas es la Validación Cruzada (Cross-Validation) que consiste en dividir los datos en un número de pliegues o particiones. Si por ejemplo elegimos cuatro, entonces cada partición se utiliza para las pruebas y las demás para el entrenamiento. Al repetir este proceso 4 veces se consigue que cada partición se haya utilizado una vez como conjunto de pruebas. La técnica estándar para predecir la tasa de error es la Validación Cruzada Estratificada (Stratified k-fold Cross-Validation). La estratificación se refiere al proceso de reorganizar los datos de tal manera a asegurar que cada pliegue sea una buena representación del

conjunto. Comúnmente se acepta que 10 es el número de pliegues con el que se obtiene la mejor estimación de error, idea basada en diversas pruebas sobre conjuntos de datos diferentes y para distintas técnicas de aprendizaje [8].

Otra técnica es el Porcentaje de División (Percentage Split) con el que puede retener para la prueba un determinado porcentaje de los datos. Es una alternativa utilizar un conjunto de pruebas separado o una división porcentual de los datos de entrenamiento. Si elegimos 60% como porcentaje de división, entonces el conjunto de prueba se constituirá con el 40% de las instancias y el conjunto de entrenamiento con el 60% de las instancias.

Experimentos

Cada producto está asociado a un conjunto de instancias de BI mensuales, quincenales y semanales. Es decir, cada producto está asociado a tres conjuntos de instancias diferentes.

Luego, cada conjunto de instancias se entrena con todos los algoritmos de clasificación WEKA posibles y la evaluación se hace tanto por el método Stratified K-fold Cross Validation. Finalmente se obtienen métricas de evaluación del aprendizaje automático con el método Stratified K-fold Cross Validation para los periodos mensuales, quincenales y semanales.

Por cada modelo procesado, luego de construir su clasificador y evaluarlo se obtienen las métricas Cantidad de Aciertos o el Porcentaje de Acierto, la estadística Kappa; y por cada clase (Nada, Poco, Medio, Mucho) se obtienen las métricas Recall (RCALL), Precision (PREC), F-Measure (FMEA).

El análisis global de los resultados se basa en la métrica Kappa. Por cada producto analizado se elige como clasificador aquel que haya alcanzado el mayor valor de Kappa, para luego calcular tres porcentajes de aciertos promedios globales: la tasa de aciertos para periodos mensuales, quincenales y semanales.

En cuanto a las limitaciones de este nuevo modelo se puede mencionar que no toma en cuenta los costes relacionados al inventario: como costes de mantenimiento, de personal, seguros, etc. El modelo planteado se aplica a empresas retail dedicadas a la venta de productos terminados.

4. JUSTIFICACIÓN

Actualmente, en el proceso de gestión de compras para la reposición de stock se utilizan técnicas de pronósticos para determinar las cantidades de las órdenes de compra. Estas técnicas pueden estar basadas en pronósticos cuantitativos o cualitativos. Independientemente de la técnica elegida el problema real con los pronósticos es su falta de confiabilidad, ya que por lo general no son precisos. La interrogante que siempre surge en estos modelos es si sus resultados serán superiores o inferiores a la demanda real y en qué medida.

En este contexto, este trabajo apuesta a desarrollar una solución que automatice la toma de decisión de la reposición de stock utilizando técnicas basadas en Business Intelligence y Machine Learning en dos fases. En la primera fase, a partir del histórico del movimiento del stock y detalle de ventas el módulo Business Intelligence obtiene los KPI asociado al producto y periodo. En la segunda fase, estos datos son utilizados como entrada al módulo de Machine Learning para determinar el volumen de compra para el siguiente periodo de ventas.

5. OBJETIVOS DEL PROYECTO

5.1.1 OBJETIVOS GENERALES

- Diseñar una nueva técnica para la toma de decisión en la reposición de stock utilizando técnicas basadas en Business Intelligence y Machine Learning en dos fases.

5.1.2 OBJETIVOS ESPECÍFICOS

- Realizar el proceso de Business Intelligence, donde dos de sus etapas claves son: modelar un datawarehouse y obtener los KPI (Key Performance Indicators).
- Realizar el etiquetado con los KPI obtenidos, con la colaboración del experto en compras. Al final de este proceso se obtendrán las instancias que se necesitan en el proceso de aprendizaje.
- Realizar el proceso de Machine Learning con las instancias proveídas en la etapa anterior. Al final del proceso se obtendrán los algoritmos de clasificación utilizados en la toma de decisión.

6. FORMULACIÓN GENERAL DEL PROYECTO

PROBLEMA E HIPÓTESIS

¿CUÁL ES EL PROBLEMA U OPORTUNIDAD?

El problema en general de cualquier técnica actual de pronóstico de la demanda, es que por ejemplo en los pronósticos de demanda cuantitativos se depende en gran medida de la variable cantidad de ventas. En los pronósticos de demanda cualitativos se depende en gran medida de la opinión y experiencia del experto, que si bien pueden ser válidos se utilizan principalmente en casos donde la proyección es a largo plazo o ante carencia de datos históricos.

También podemos mencionar otra debilidad de las técnicas de pronósticos actuales, y es que no son adaptativas, en el sentido de que como se tratan de fórmulas genéricas los resultados son uniformes y no tienen en cuenta la evolución del mercado ni aspectos particulares como el tipo de producto, tipo de empresa, ubicación geográfica, etc.

La oportunidad que surge es poder implementar una técnica de toma de decisión de reposición automática o con una mínima intervención manual. Esta técnica permitiría paliar las debilidades de los modelos actuales, y a su vez que integraría las características esenciales de las técnicas cualitativas y cuantitativas existentes.

¿CUÁL ES LA SOLUCIÓN PROPUESTA POR EL PROYECTO DE INVESTIGACIÓN?

Realizar primeramente un proceso de Business Intelligence, donde se pretenden obtener los valores de los KPI considerados para la solución. Una vez obtenidos los conjuntos de KPI por producto, estos son analizados por un experto en compras, quien luego asigna una etiqueta a cada subconjunto de valores.

Los resultados obtenidos del proceso de Business Intelligence constituyen la entrada para el proceso de Machine Learning. Luego se realizarán el entrenamiento y testeo de las instancias, para obtener por cada producto el mejor algoritmo de clasificación que posteriormente se utilizará para pronosticar la demanda de períodos futuros.

¿CUÁL ES LA HIPÓTESIS DEL PROYECTO?

La hipótesis de este trabajo es que la nueva técnica de toma de decisión alcanzará altas tasas de aciertos en comparación a las indicadas por el experto. También será una técnica adaptativa, permitirá el aprendizaje continuo, integrará aspectos esenciales de los modelos de pronósticos cualitativos y cuantitativos, y paliará ciertas debilidades de los modelos mencionados.

7. ANTECEDENTES.

Antecedentes de Pronóstico de la demanda

El pronóstico de la demanda es una predicción de eventos que ocurrirán, es una técnica utilizada por las empresas para anticiparse al futuro y elaborar las estrategias apropiadas [1]. En las empresas, durante décadas han intentado optimizar la cantidad de inventario por medio de una predicción más efectiva de la demanda de sus clientes, sin embargo la naturaleza incierta de las predicciones muchas veces ha contribuido más a la incertidumbre en lugar de minimizarlo. *“Los pronósticos jamás son perfectos y serán menos confiables mientras mayor sea el lapso que se pronostique hacia el futuro”* [2]

Las técnicas empleadas para la realización de pronósticos de la demanda se clasifican en [12]:

- Técnicas Cualitativas: las técnicas cualitativas se basan principalmente en el conocimiento humano, en el juicio de un experto para elaborar las estimaciones futuras. En ocasiones son conocidas como técnicas

subjetivas y son utilizadas cuando no se tienen información histórica sobre la variable que se está pronosticando. Se clasifican en: Método Delphi, Juicio experto, Redacción de escenarios, Enfoques intuitivos.

- Técnicas Cuantitativas: las técnicas cuantitativas se utilizan cuando se dispone de información histórica de las variables a estudiar, el pronóstico se basa en modelos matemáticos principalmente los estadísticos. Los métodos cuantitativos se dividen en a) *métodos de serie de tiempo* (Métodos de suavización, Proyección de la tendencia, Proyección de la tendencia ajustada por influencia estacional) se utilizan cuando la variable a pronosticar únicamente hace relación a la información histórica que se posee observados en periodos de tiempo iguales y b) *métodos de pronóstico causal* están basadas en el supuesto de que la variable que se quiere pronosticar tiene relación de causa y efecto con una o más variables se trata de determinar exactamente cuáles son dichas relaciones.

Antecedentes de Business Intelligence

La mayoría de las organizaciones contaban con un sistema de información donde se guardan las actividades diarias realizadas en la empresa, con el transcurso del tiempo estas informaciones van creando un histórico de las actividades que generalmente son almacenadas en una base de datos [3] pero que no disponían de un mecanismo que permita un rápido y fácil acceso a dicha información.

En el año 1989, Howard Dresner consultor de Gartner Group acuñó por primera vez el término de Business Intelligence, que abarca un conjunto de conceptos, técnicas y herramientas que se utiliza para la transformación de datos transaccionales en información útil y significativa para el análisis de negocios, tiene como objetivo principal apoyar a las organizaciones a mejorar su competitividad, facilitando la información necesaria para la toma de decisiones [4]. Entre las muchas áreas donde implementan soluciones de Business Intelligence podemos citar: Ventas Minoristas en empresas retail, Inventario, Gestión de Pedidos, Contabilidad, Gestión de Relación con el Cliente, Gestión de recursos humanos, Servicios financieros, Telecomunicaciones, Educación, Medicina, Comercio Electrónico, Seguros entre otros [5].

Antecedentes de Machine Learning

En 1959 Arthur Samuel en una publicación escribió: “Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort” [9]. Lo que nos lleva a pensar que uno de los pioneros del aprendizaje automático ya dejaba visualizar que los programas, a partir del aprendizaje sobre los datos históricos (la experiencia), podrían efectuar tareas de toma de decisiones sin ser programadas explícitamente dichas decisiones. Samuel define al aprendizaje automático como sigue: “El aprendizaje automático es un campo de estudio que da a las computadoras la capacidad de aprender sin ser explícitamente programadas”.

Otro investigador de aprendizaje automático Tom Mitchell propuso en 1998 la siguiente definición: “Well posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E”. Donde se nos indica que el aprendizaje en las máquinas deberá ser parecido al aprendizaje en los humanos, por ejemplo cuando una criatura comienza a hablar a través de la experiencia de pronunciar las palabras y de su interacción con otras personas, entonces sucede que su capacidad de hablar se va perfeccionando o mejorando.

“The purpose of machine learning is to learn from training data in order to make as good as possible predictions on new, unseen, data” [10]. La dificultad radica en que debemos construir modelos que nos acerquen a una buena predicción sobre datos aún no conocidos o imprevistos.

Publicaciones que comprenden los últimos diez años, presentan trabajos muy interesantes en el ámbito de Machine Learning asociado al Demand Forecasting (Pronóstico de la Demanda). Se citan a continuación algunos de los problemas afrontados en publicaciones, que dan una idea del estado del arte en este tema:

- Mejorar la precisión de la previsión de demanda de agua urbana para la ciudad de Montreal – Canadá (2017).
- Proponer un método de control inteligente para sistemas de calefacción y refrigeración. (2017).
- Proponer un modelo predictivo probabilístico de consumo de energía, basado en datos, para la predicción del consumo en edificios residenciales (2017).

- Revisión de diferentes modelos de predicción de la carga eléctrica con un enfoque particular en modelos de regresión (2017).
- Aplicación de Machine Learning en la nube para encontrar conversaciones de los consumidores que influyen en las decisiones de compras (2016).
- Modelar la demanda turística de España (2016).
- Predicción del Mercado de Valores (2016).
- Análisis para un minorista en línea: Previsión de la demanda y optimización de precios (2016).
- Demanda de calefacción residencial basado en el consumo total mensual de gas natural (2015).
- Predicción de la demanda de importación de crudo en Taiwán (2014).
- Predicción del desempeño de las estrategias de pronóstico para la demanda de repuestos navales (2012).

8. METODOLOGÍAS

- Investigación del estado del arte de:
 - Pronóstico de la demanda.
 - Business Intelligence.
 - Machine Learning.
- Implementación de un datawarehouse, a partir de una base de datos real de una empresa retail.
- Generación de los valores de KPI y emulación del experto en compras.
- Implementación de un programa que utiliza la API de Weka para el entrenamiento y testeo del dataset.
- Evaluación de los resultados para seleccionar el mejor algoritmo de clasificación por cada producto.
- Redacción del documento científico.

9. RESULTADOS ESPERADOS DEL PROYECTO

Un sistema de toma de decisión en la reposición de stock adaptativo en los distintos productos de un establecimiento comercial.

10. PLANEAMIENTO DEL PROYECTO

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
Investigación del Estado del Arte.												
Diseño del modelo propuesta.												
Implementación.												
Pruebas y Ajustes												
Escritura del Libro												
Evaluación de Resultados y elaboración de las conclusiones												
Defensa del trabajo												

11. TEMARIO TENTATIVO DEL REPORTE TÉCNICO FINAL

1. Introducción
2. Pronóstico de la demanda
3. Business Intelligence
4. Machine Learning
5. Modelado de Business Intelligence
6. Modelado de Machine Learning
7. Experimentación
8. Conclusiones y trabajos futuros
9. Referencias

REFERENCIAS BIBLIOGRÁFICAS

- [1] Anderson, David R., Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm y Kipp Martin. Métodos cuantitativos para los negocios, 11a ed.
- [2] Narasimhan, S., McLeavey, D.W., Billington, P. Planeación de la producción y control de inventarios. 1996. Prentice-Hall Hispanoamericana
- [3] Business Intelligence: State of the art. Alveiro Rosado, Dewar Rico. 2010 <https://dialnet.unirioja.es/descarga/articulo/4564348.pdf>
- [4] Josep Luis Cano. Business Intelligence: Competir con información. ESADE, Banesto, Banesto Pyme, 2007.
- [5] “The datawarehouse Toolkit.” Ralph Kimball. Wiley, 1996.
- [6] Weka 3: Data Mining Software in Java. <http://www.cs.waikato.ac.nz/ml/weka/>
- [7] Witten, I. H.; Frank, E.; Hall, M. A. & Pal, C. J. Data Mining - Practical Machine Learning Tools and Techniques - Fourth Edition, 2016.
- [8] Witten, I. H.; Frank, E. & Hall, M. A. Data Mining - Practical Machine Learning Tools and Techniques - Third Edition, 2011.
- [9] Samuel, A. Some Studies in Machine Learning Using the Game of Checker IBM Journal 3, 211-229, 1959.
- [10] Puget, J. F. What Is Machine Learning?, 2016.
- [11] Johnson, P. F.; Leenders, M. R. & Flynn, A. E. Administración de compras y abastecimientos McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V, 2012
- [12] Anderson, D. R.; Sweeney, D. J.; Williams, T. A.; Camm, J. D. & Martin, K. Métodos cuantitativos para los negocios © D.R. 2011 por Cengage Learning Editores, S.A. de C.V., una compañía de Cengage Learning, Inc, 2011