

# Herramienta web de selección de productos en tiendas de conveniencia para su promoción utilizando analítica de datos.

**Trabajo Terminal No. \_\_\_\_\_ - \_\_\_\_\_**

*Alumnos: Aguilar Reyes Abraham Zaid, \*Franco Ake Alan, Medero Luján Alejandro.*

*Directores: Araujo Díaz David, Méndez Segundo Laura.*

*Email: afrancoal400@alumno.ipn.mx*

**Resumen** – Las tiendas de conveniencia día a día están en constante cambio, durante un periodo generan ganancias que les permite hacer crecer sus ventas y otros en los que sufren pérdidas y que en el peor caso, puede ocasionar su cierre. Esto no sólo se debe al alta o a la baja demanda que pueden llegar a tener cada uno de los productos que ofrecen, si no que también influye el comportamiento de los clientes, su actividad por periodo y sus razones de compra. En este Trabajo Terminal se propone elaborar una herramienta web intuitiva de analítica de datos en tiempo real que permita dotar a las tiendas de conveniencia con nuevas sugerencias de qué productos conviene vender según las compras de los clientes en cada época del año.

**Palabras clave** – Análisis Predictivo, Ciencia de Datos, Desarrollo de aplicaciones Web y Minería de datos.

## 1. Introducción

Las tiendas de conveniencia nos facilitan la adquisición de productos de la canasta básica. Mismos de los cuales existen una gran variedad de opciones a escoger, es tanta la competencia en el mercado que nosotros como clientes frecuentes, factores como el precio, la cantidad y la presentación influyen en nuestra compra final, sin embargo, conforme ha avanzado el comercio electrónico, han surgido nuevos métodos de compra y distribución, nuevas formas para promocionar productos, motivos por el cual la Ciencia de Datos ha tomado un papel muy importante. [1]

Analizar los datos de un ticket de compra supone la oportunidad de hacer recomendaciones personalizadas, basadas en compras o consultas anteriores del propio cliente o de clientes que hayan hecho compras similares.[2] Con base en esta información, se puede determinar qué productos tienen más ventas, cuáles no y en qué época del año tuvieron mayor demanda.

Analizar los datos y aplicar técnicas como el árbol de decisión o los bosques aleatorios pueden ayudar a la tienda de conveniencia a averiguar cuánto más va a vender de un producto de baja rotación si saca una nueva promoción de 3x2 o de 2ª unidad al 50% de descuento. [3]

La Ciencia de Datos es una disciplina científica y de negocios que ha adquirido gran notoriedad a medida que las tendencias de *Big Data* se han ido desarrollando y desplegando [4]. La persona que se encarga de encontrarle valor significativo a los datos recibe el nombre de “Científico de Datos” y para lograrlo suele utilizar una caja amplia de herramientas. Cassie Kozyrkov (científica y estadística de datos) resume la caja de herramientas en tres subáreas que son [5] :

- La minería de datos (o *Data Mining*), permite encontrar patrones y resumir grandes volúmenes de información.
- La inferencia estadística, conjunto de técnicas explicativas que utilizan muestras representativas de una población para comprobar la certeza de nuestras afirmaciones (llamadas hipótesis).
- Y el aprendizaje automático (o *Machine Learning*), basado en la experiencia, estudia y diseña algoritmos para la toma de decisiones.

De las subáreas anteriormente mencionadas, partiremos de la minería de datos, a una aplicación en particular que es el análisis predictivo.

El análisis predictivo puede generar conocimientos futuros con un grado significativo de precisión. Con la ayuda de herramientas sofisticadas y modelos de análisis predictivo, cualquier organización ahora puede usar datos pasados y actuales para pronosticar de manera confiable tendencias y comportamientos de sus clientes, además de detectar riesgos y oportunidades para el crecimiento del negocio. [6]

El análisis predictivo en supermercados tiene 5 áreas de aplicación que surgen a partir de los hábitos de compra de los clientes, estas son:

1. Promociones.
2. Orientación al comprador.
3. Gestión de campañas de marketing.
4. Precios.
5. La gestión del inventario.

### **Herramientas similares.**

Realizando una investigación sobre productos similares que actualmente se encuentran al acceso del público pudimos encontrar lo siguiente:

#### **iStratega**

Se trata de una herramienta web, que se encarga de desarrollar páginas web para el *eCommerce*. Dentro de las características principales, está que al cotizar tu proyecto, decides que tan compleja será dicha página, al igual de las características generales de la página a desarrollar. Además te ofrece la previsualización del costo estimado del proyecto y finalmente aclara que de ser necesario, se llega a una atención personalizada, para declarar que se espera realmente de la página y del costo exacto que tendrá la creación de la página. [7]

#### **Express Analytics**

Se encarga del análisis de los clientes, segmentar el mercado y directamente promocionar productos a potenciales compradores. Es una herramienta en inglés, así que si no se cuenta con el conocimiento en el idioma, es una barrera bastante importante a considerar. [8]

#### **Syte**

Una herramienta de *eCommerce* que implementa una inteligencia artificial para analizar los datos de entrada, clicks y la interacción que tienen los clientes con tus productos, más que un análisis para promoción de productos, es un análisis de promoción de negocio, ya que te permite observar cómo se está observando tu empresa a los usuarios y te da los datos para que intérpretes y corrijas tu modelo de negocio. [9]

SOFTWARE	CARACTERÍSTICAS	PRECIO EN EL MERCADO
Sistema optimizador de campañas publicitarias online SOCPO	Presenta la implementación de redes neuronales artificiales a las campañas del sistema de publicidad Google AdWords.	Trabajo Terminal
Simulador para seleccionar a qué clientes se les puede otorgar créditos mediante técnicas de minería de datos	Desarrolla un simulador web que sirva de apoyo a las micro y pequeñas empresas para tomar mejores decisiones con técnicas de	Trabajo Terminal

	minería de datos.	
Sistema de inteligencia de negocios en la nube aplicado al comercio electrónico	Muestra la propuesta de desarrollo de un sistema de información gerencial (de inteligencia del negocio) para apoyo en la toma de decisiones de una PyME.	Pagos por presupuestos personalizados
Syte	Combina el comportamiento en tiempo real con una IA visual para predecir qué productos es más probable que adquiera cada cliente.	Pagos por presupuestos personalizados

*Tabla 1. Herramientas similares.*

## 2. Objetivo

### Objetivo General

Implementar una herramienta web para realizar el análisis de una base de datos que contiene los detalles de compra de los clientes de una tienda de conveniencia, de tal forma que mediante técnicas de minería de datos, se obtenga una predicción sobre el producto que se deba promocionar.

### Objetivo Específicos

1. Determinar qué tipos atributos de las bases de datos son útiles para el análisis.
2. Analizar patrones de compra de las transacciones de los clientes.
3. Diseñar la arquitectura de la red neuronal para entrenarla con los datos de prueba.
4. Desarrollar una herramienta web intuitiva para los usuarios.

## 3. Justificación

Como consumidores hemos observado que en muchas ocasiones existen nuevos productos que buscan competir con grandes marcas, ya sea ofreciendo nuevos beneficios, o incluso a veces reinventando el producto, para brindar un mejor beneficio y marcar una diferencia con respecto a los demás competidores en el mercado, sin embargo, muchos de estos productos tienen menores ventas, no prosperan y pasan al olvido, todo esto nos llevó a la pregunta ¿cuáles productos que están siendo poco vendidos pueden generar más ventas?

Para resolver la problemática anterior, existen herramientas y servicios que se mencionan en la *Tabla 1* de este documento, de las cuáles requieren del apoyo de analistas y científicos de datos que se encargan de darle un significado a los datos. Dentro de este análisis, los resultados a obtener son: identificación de clientes, planificación de inventario y mitigación de riesgos, atención al cliente personalizada e información precisa en tiempo real, logrando así que las tiendas de conveniencia identifiquen que deben de reestructurar en su organización y qué aspectos modificar en su logística, pero para ello, tiene que realizar un gasto para este estudio y que en ocasiones puede o no impulsar el negocio.[10] Esto debido a que las tiendas de conveniencia no hacen un correcto uso de estos conocimientos. [11]

Con base en lo anterior, buscamos dotar a las tiendas un sistema que les permita saber qué productos serán conveniente vender durante cierto periodo tiempo según sean los patrones de compra de los clientes, la época del año y la zona geográfica en la que se encuentran, y ellos tomarán la decisión de cómo establecer sus promociones, sin necesidad de requerir atención personalizada para impulsar o reestructurar el modelo de negocio. En resumen, ofrecemos una herramienta de predicción, y las tiendas se encargan de hacer uso de la información recabada.

#### 4. Productos o Resultados esperados

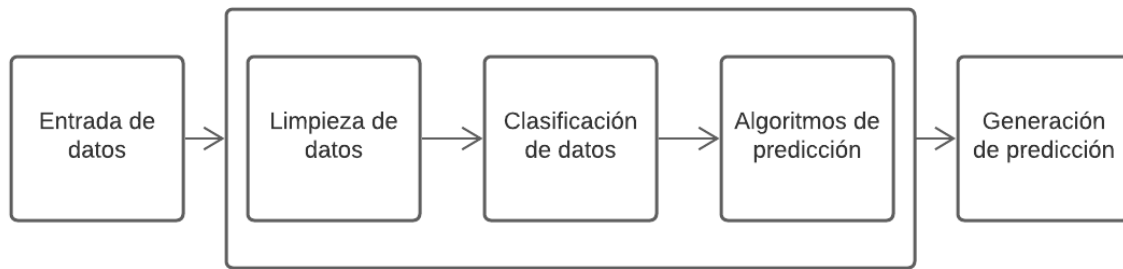


Figura 1. Diagrama de bloques del sistema.

- Una base de datos con detalles de compras realizadas por los clientes de la tienda.
- Desarrollo de una red neuronal que predice posibles resultados a campañas publicitarias de ciertos productos
- Desarrollo de un procedimiento de minería de datos sobre la base de datos.
- Aplicación web desarrollada en lenguaje Java con el framework Spring para el acceso a la herramienta.
- Documentación Técnica.
- Manual de Usuario.

#### 5. Metodología

Para el desarrollo de este trabajo terminal, los integrantes del equipo en conjunto con los directores, optamos por utilizar la metodología Scrum. Scrum es una metodología de desarrollo ágil de software concebido por Jeff Sutherland y su equipo de desarrollo a principios de la década de 1990.

Los principios Scrum son congruentes con el manifiesto ágil y se utilizan para guiar actividades de desarrollo dentro de un proceso de análisis que incorpora las siguientes actividades estructurales: requerimientos, análisis, diseño, evolución y entrega. Dentro de cada actividad estructural, las tareas del trabajo ocurren con un patrón del proceso llamado sprint. El trabajo realizado dentro de un sprint se adapta al problema en cuestión y se define y/o modifica en tiempo real por parte del equipo Scrum. [12]

Scrum contiene los siguientes eventos:

- Planificación del Sprint (*Sprint Planning*)
- Scrum Diario (*Daily Scrum*)
- Revisión del Sprint (*Sprint Review*)
- Retrospectiva del Sprint (*Sprint Retrospective*)

Estos eventos existen con el fin de establecer una regularidad y minimizar la necesidad de reuniones no definidas en Scrum. Estos eventos son bloques de tiempo (time boxes), de tal forma que todos cuentan con una duración máxima.

También se definen los siguientes artefactos:

- Lista de Producto (*Product Backlog*)
- Lista de Pendientes del Sprint (*Sprint Backlog*)
- Incremento (*Increment*)

Los artefactos en Scrum se definen para así fomentar la transparencia de la información de tal manera que todos los involucrados tengan el mismo entendimiento de que es lo que se está llevando a cabo, además de que nos crean oportunidades para realizar inspecciones y adaptaciones.

Se pretenden realizar 4 Sprints en cada Trabajo Terminal:

#### **Trabajo Terminal 1.**

- **Sprint 1:** Realización del análisis y diseño de la herramienta web
- **Sprint 2:** Investigación sobre las técnicas de minería de datos y la implementación del análisis predictivo, así como también la recolección de datos.
- **Sprint 3:** Sección de técnicas para el tratamiento de datos y limpieza de los mismos.
- **Sprint 4:** Clasificación de datos e implementación de algoritmos de predicción.

#### **Trabajo Terminal 2.**

- **Sprint 5:** Elaboración de la interfaz web de la herramienta.
- **Sprint 6:** Elaboración de pruebas de la herramienta.
- **Sprint 7:** Generación del manual de usuario.
- **Sprint 8:** Generación del reporte técnico.

## 6. Cronograma

CRONOGRAMA Nombre del alumno(a): Aguilar Reyes Abraham Zaid

TT No.:

Título del TT: Herramienta web de selección de productos en tiendas de conveniencia para su promoción utilizando analítica de datos.

[illegible]

CRONOGRAMA Nombre del alumno(a): Franco Ake Alan

TT No.:

Título del TT: Herramienta web de selección de productos en tiendas de conveniencia para su promoción utilizando analítica de datos.

[illegible]

CRONOGRAMA Nombre del alumno(a): Medero Luján Alejandro

TT No.:

Título del TT: Herramienta web de selección de productos en tiendas de conveniencia para su promoción utilizando analítica de datos.

[illegible]



## 7. Referencias

- [1] I. Witten, E. Frank and M. Hall, *Data mining*, 3rd ed. Burlington, Mass.: Morgan Kaufmann Publishers, 2011, pp. 26-27.
- [2] C. Aggarwal, *Data mining*. New York: Springer, pp. 3-5.
- [3] D. Larose and C. Larose, *Data mining and predictive analytics*, 2nd ed. New Jersey: Wiley, 2015, pp. 3-5.
- [4] L.J Aguilar. Inteligencia de Negocios y Analítica de Datos. México: Alfaomega, 2019, pp. 86.
- [5] C. Kozyrkov, "What on earth is data science? | Hacker Noon", *Hackernoon.com*, 2018. [Online]. Available: <https://hackernoon.com/what-on-earth-is-data-science-eb1237d8cb37>. [Último acceso: 04 Marzo 2021].
- [6] J. Edwards, "Predictive analytics: Transforming data into future insights", *CIO*, 2019. [Online]. Available: <https://www.cio.com/article/3273114/what-is-predictive-analytics-transforming-data-into-future-insights.html>. [Último acceso: 04 Marzo 2021].
- [7] S. Pawar, "Diseño de páginas web profesionales, creativas y únicas", *istratega - Agencia Diseño Web*, 2021. [Online]. Available: <https://www.istratega.com/>. [Último acceso: 02 Marzo 2021].
- [8] S. Pawar, "Diseño de páginas web profesionales, creativas y únicas", *istratega - Agencia Diseño Web*, 2021. [Online]. Available: <https://www.istratega.com/>. [Último acceso: 02 Marzo 2021].
- [9] L. Pinto, "Syte | The World's First Product Discovery Platform", *Syte.ai*, 2021. [Online]. Available: <https://www.syte.ai/>. [Último acceso: 02 Marzo 2021].
- [10] G. Jaim "Predictive Analytics in the Retail Industry", 2019 [Online] Available: <https://blogs.mastechinfotrellis.com/predictive-analytics-in-the-retail-industry>
- [11] A. Corral- Garza, I. Echevarria Garza, L. Flores Bunes "BIG DATA ANALYTICS COMO RECURSO INTANGIBLE GENERADOR DE VALOR EN LA INDUSTRIA DEL RETAIL: CASO DE ESTUDIO OXXO", 2020 [Online] Available: [http://www.web.facpya.uanl.mx/Vinculategica/Vinculategica6\\_1/11%20CORRA\\_ECHEVERRIA\\_FLORES.pdf](http://www.web.facpya.uanl.mx/Vinculategica/Vinculategica6_1/11%20CORRA_ECHEVERRIA_FLORES.pdf)
- [12] R. Pressman, *Ingeniería del software*, 7th ed. México: McGraw-Hill, pp. 69-70.

## 8. Alumnos y Directores

*Aguilar Reyes Abraham Zaid.*- Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Boleta: 2015080012, Tel. 5519708940, email [aaguilar1409@alumno.ipn.mx](mailto:aaguilar1409@alumno.ipn.mx)

Firma: ACUSE DE RECIBIDO

*Franco Ake Alan.*- Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Boleta: 2015020402, Tel. 5520657474, email [afranco1400@alumno.ipn.mx](mailto:afranco1400@alumno.ipn.mx)

Firma: ACUSE DE RECIBIDO

*Medero Luján Alejandro.*- Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Boleta: 2018630901, Tel. 5560665787, email: [AMEDEROL1700@alumno.ipn.mx](mailto:AMEDEROL1700@alumno.ipn.mx)

Firma: ACUSE DE RECIBIDO

CARÁCTER: Confidencial  
FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos 108, 113 y 117 de la Ley Federal de Transparencia y Acceso a la Información Pública.  
PARTES CONFIDENCIALES: Número de boleta y teléfono.

## Directores

*Araujo Díaz David.*- Profesor de la Escuela Superior de Cómputo (Depto. de Posgrado), Ing. en Comunicaciones y Electrónica (ESIME-Zacatenco), M. en C. en Ingeniería Eléctrica en la Opción de Computación (CINVESTAV-IPN). Áreas de Interés: Realidad Virtual, Diseño y Simulación de Circuitos Electrónicos, Modelación Matemática, Cómputo Paralelo, Robótica, Procesamiento de Imágenes, Reconocimiento de Patrones, Computabilidad, Complejidad Algorítmica y Seguridad Informática. Tel: 57-29-60-00 Ext. 52038. Email: [daraujo@ipn.mx](mailto:daraujo@ipn.mx)

  
Firma: ACUSE DE RECIBIDO

*Méndez Segundo Laura.*- M. en C. con especialidad en computación CINVESTAV 1998, Licenciatura en Informática, Universidad Veracruzana 1991. Certificado de SCRUM Master en el 2017, Áreas de Especialidad: Bases de Datos, Ingeniería de Software, UML. Área de Interés cómputo educativo y procesamiento de imágenes. Tel: 57-29-60-00 Ext. 52032, Email: [lmendez@ipn.mx](mailto:lmendez@ipn.mx)



Firma: ACUSE DE RECIBIDO

CARÁCTER: Confidencial  
FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos 108, 113 y 117 de la Ley Federal de Transparencia y Acceso a la Información Pública.  
PARTES CONFIDENCIALES: Número de boleta y teléfono.