

Predicción de la Intención de Compra en el Comercio Electrónico: Caso de éxito

Prediction of Purchase Intention in E-commerce: Exit Case

AUTOR

Zamora Pérez Angie Leonor
Universidad Internacional La Rioja
México-Ciudad de México
zamoraperez@hotmail.com
<https://orcid.org/0009-0008-5585-1036>

Como citar: Zamora Pérez , A. L. (2025). Predicción de la Intención de Compra en el Comercio Electrónico: Caso de éxito. <i>Revista Internacional De Investigación Y Desarrollo Global</i> , 4(3), 1–14.	Fecha de recepción: 2025-06-15 Fecha de aceptación: 2025-07-15 Fecha de publicación: 2025-08-15
---	---

Resumen

El comercio electrónico ha experimentado un crecimiento acelerado en los últimos años, potenciado por la transformación digital y, especialmente, por la pandemia de COVID-19, que modificó radicalmente los hábitos de consumo a nivel mundial. Este estudio analiza la intención de compra en línea mediante la aplicación de técnicas de minería de datos y algoritmos de clasificación. Para ello, se trabajó con un conjunto de datos de 12.546 sesiones de compra, complementado con información local en Ecuador. El análisis incluyó variables relacionadas con la duración de las visitas, tipos de páginas consultadas, tasas de rebote y valores de página, aplicando un enfoque cuantitativo y correlacional. Los modelos de aprendizaje automático evaluados —Extra Trees Classifier, Random Forest y Gaussian Naive Bayes— fueron comparados en función de su precisión y métrica AUC. Los resultados muestran que el modelo **Extra Trees Classifier** alcanzó un rendimiento superior (93,99% de exactitud y AUC de 0,92), lo que confirma su idoneidad para predecir la intención de compra en plataformas minoristas. Este hallazgo permite proponer un modelo de predicción que contribuya a personalizar la experiencia del usuario, optimizar estrategias de marketing y aumentar las tasas de conversión en el comercio electrónico ecuatoriano.

Palabras clave: Comercio Electrónico, Intención De Compra, Minería De Datos, Aprendizaje Automático, Predicción.

Abstract

Electronic commerce has experienced accelerated growth in recent years, driven by digital transformation and, in particular, by the COVID-19 pandemic, which radically changed consumer habits worldwide. This study analyzes online purchase intention through the application of data mining techniques and classification algorithms. The dataset included 12,546 purchase sessions, complemented with local information from Ecuador. The analysis considered variables such as visit duration, page types, bounce rates, and page values, applying a quantitative and correlational approach. Machine learning models—Extra Trees Classifier, Random Forest, and Gaussian Naive Bayes—were evaluated and compared using accuracy and AUC metrics. Results show that the **Extra Trees Classifier** achieved the best performance (93.99% accuracy and 0.92 AUC), confirming its suitability for predicting purchase intention in retail platforms. These findings support the proposal of a predictive model that contributes to personalizing user experience, optimizing marketing strategies, and increasing conversion rates in Ecuadorian e-commerce.

Keywords: E-Commerce, Purchase Intention, Data Mining, Machine Learning, Prediction.

Introducción

El comercio electrónico es hoy en día una parte indispensable del comercio minorista global. A semejanza de otras áreas la industria ha sufrido una transformación digital, hoy en día los consumidores se benefician de las transacciones en línea. A su vez la cantidad de compradores digitales aumenta cada año, por lo que se establece una comparación del incremento de compras en línea por categoría. Según una proyección realizada por el proveedor de datos Statista, el crecimiento de las compras en línea se establece con un mayor porcentaje en la electrónica de consumo (Statista, 2021a). Este incremento en ventas se proyecta sobre todo en los productos electrónicos, sin embargo, se espera que el crecimiento en la compra de víveres de por internet aumente con los diversos periodos de confinamiento que se han en el mundo por consecuencia de la pandemia.

A nivel mundial en el año 2020, durante la crisis de COVID-19, el comportamiento de los consumidores y los hábitos de compra cambiaron fundamentalmente. Aproximadamente 2.000 millones de personas adquirieron bienes o servicios en línea y, durante el mismo año, los comercios reportaron haber superado los 4,2 billones de dólares estadounidenses, en ventas minoristas a nivel mundial (Jílková & Králová, 2021). Por lo que se concluye que la crisis de COVID-19 generó un cambio en la conducta de los consumidores, representado por un incremento del manejo de medios digitales para la adquisición de bienes y servicios.

Algunas cifras relevantes reportadas al CEPAL (2020) por uno de líderes de América Latina en lo correspondiente a comercio electrónico indican que durante la pandemia tuvieron más de 1.7 millones de nuevos compradores en su plataforma, de los cuáles el 80% de compradores habituales, concretaron por primera vez una compra de alimentos y el 20% realizó su primera compra en esta época.

Las entregas diarias crecieron en más de 1,1 millones. Los consumidores tienden a realizar sus compras en línea desde su computadora o desde su smartphone destacando que la comodidad de las compras en línea las convierte en una tendencia emergente entre los consumidores, especialmente la Generación Y (Lim et al., 2016).

El Ecuador no es una excepción a este tipo de tendencia, por lo que se presenta este estudio, con la finalidad de demostrar que, recabando la información de las sesiones de compra de los consumidores, se puede detectar el patrón de comportamiento de los clientes para orientar la publicidad, el marketing y las ofertas a clientes potenciales, según una encuesta de 2021, alrededor del 70% de los compradores utilizan un teléfono móvil, mientras que el 30 % indica que usa una computadora (Statista, 2021b). El conjunto de datos utilizado en este análisis proviene de un sitio Business to Consumer y son compartidos por Sakar C. & Kastro (2018), y el mismo se ha complementado con información recabada a nivel local, los datos pertenecen a un año de monitoreo, por lo

que se presenta un análisis de minería de datos de la intención de compra de los visitantes al sitio, lo que permite reconocer un patrón de comportamiento del consumidor según la tendencia de las sesiones.

Una de las técnicas de análisis de minería de datos es la clasificación de una cantidad tan voluminosa de datos en dos o más clases distantes. Aplicando el aprendizaje supervisado, se utilizan datos etiquetados, para generar un modelo que permite predecir la influencia de compra en un sitio de comercio minorista.

El trabajo propuesto compara la precisión de varias técnicas de clasificación. Los algoritmos de clasificación consisten en examinar las características de un elemento presente en el conjunto de datos y asignarlo a uno de los conjuntos predefinidos de clases. Los elementos que van a ser clasificados están generalmente representados por los registros que se contienen de ese elemento en una tabla de base de datos o un archivo. La tarea de clasificación se caracteriza por contar con una correcta definición de las clases, y de una formación de entrenamiento que consiste en ejemplos preclasificados. La tarea es construir o aplicar un modelo de algún tipo que pueda ser empleado en los datos que aún no hayan sido clasificados con el fin de clasificarlos.

Desde esta problemática se generan la pregunta de investigación: ¿Cómo ayudaría un algoritmo de predicción de la intención de compras en línea al crecimiento exponencial de un comercio minorista?

Identificando como variable dependiente la eficiencia de los algoritmos ejecutados y como independiente la cantidad de sesiones evaluadas y clasificadas en el dataset.

Material y métodos

Material

En la metodología de análisis, se realizó un análisis exploratorio con enfoque cuantitativo en las unidades muestreadas, las cuales pertenecen a 12546 sesiones de compras captadas en un año con lo cual se genera un análisis numérico de correlación.

Para el efecto se ha hecho uso de diversos tipos de investigación, teniendo en primera instancia la investigación documental a través de la cual se ha logrado describir y explicar los conocimientos previamente establecidos en las fuentes confiables de información para la consolidación del proyecto.

El diseño de la investigación es observacional y longitudinal basado en las siguientes variables:

Efecto del número de diferentes tipos de páginas visitadas en esa sesión y el tiempo total dedicado a cada una de estas categorías de páginas.

Efecto del porcentaje de visitantes que ingresan al sitio desde esa página y luego se van ("rebote") sin activar ninguna otra solicitud al servidor de análisis durante esa sesión.

Efecto del valor promedio de una página web que un usuario visitó antes de completar una transacción de comercio electrónico "Valor de la página".

Como técnica de análisis de la información se aplica un análisis de correlación de Pearson, el cual permite reconocer la dependencia entre los atributos identificados, utilizando un alcance predictivo se generará la recomendación del algoritmo del que se obtenga una mayor eficiencia, con fines prácticos de aplicación. En la tabla 2 se describe los atributos del conjunto de datos utilizado en este análisis los que provienen de un sitio Bussiness to Consumer y son compartidos por Sakar C. & Kastro (2018), y el mismo se ha complementado con información recabada a nivel local en un sitio web de comercio electrónico, los datos pertenecen a un año de monitoreo.

Tabla 1. Attributes and description

Atributos	Descripcion
Administrative	Número de páginas visitadas
Administrative_Duration	Tiempo de cada visita
Informational	Páginas del mismo tipo visitadas por el usuario



Informational_Duration	Tiempo de visita de páginas de este tipo
ProductRelated	Páginas relacionadas con el producto
ProductRelated_Duration	Tiempo en las páginas relacionadas con el producto
BounceRates	Porcentaje de visitantes del sitio
ExitRates	Porcentaje de páginas vistas que desencadenan una visita a nuestra página
PageValues	Promedio entre valor de la página y transacción finalizada
SpecialDay	Aproximación a días especiales (Día de la madre, San Valentín, entre otros)
Month	Mes de la visita
OperatingSystems	Sistema operativo desde el cual se accedió
Browser	Navegador
Region	Valor numérico que presenta el cantón de acceso
TrafficType	Valor número del tipo de tráfico web
VisitorType	Visitante nuevo, recurrente u otro
Weekend	Booleano que representa si es o no fin de semana
Revenue	Booleano que representa si se concretó o no la compra

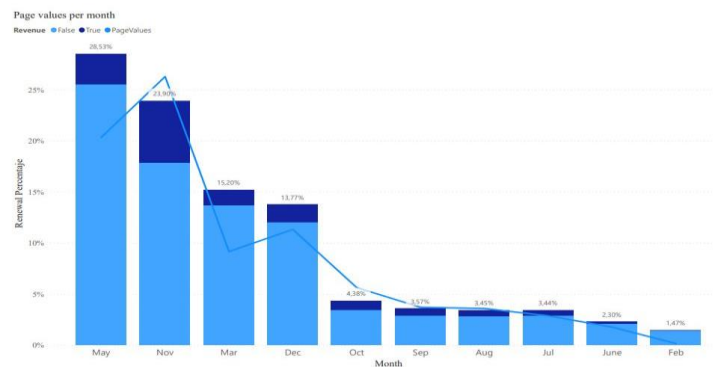
Análisis descriptivo de los datos recopilados

En esta fase se analizó la información correspondiente a las sesiones de compra, en primer lugar, se verificó que no existan datos nulos en el dataset, adicionalmente se evidencio que de los sistemas operativos que se usaron para ingresar al sitio de comercio electrónico el más utilizado es Android Mobile y el menos utilizado es Linux, de los Navegadores se puede resaltar que la mayor cantidad de sesiones es desde Google Chrome.

Dado que se cuenta con datos cuantitativos se generaron gráficos para presentar las características representativas de las variables de estudio(Charry et al., 2020).

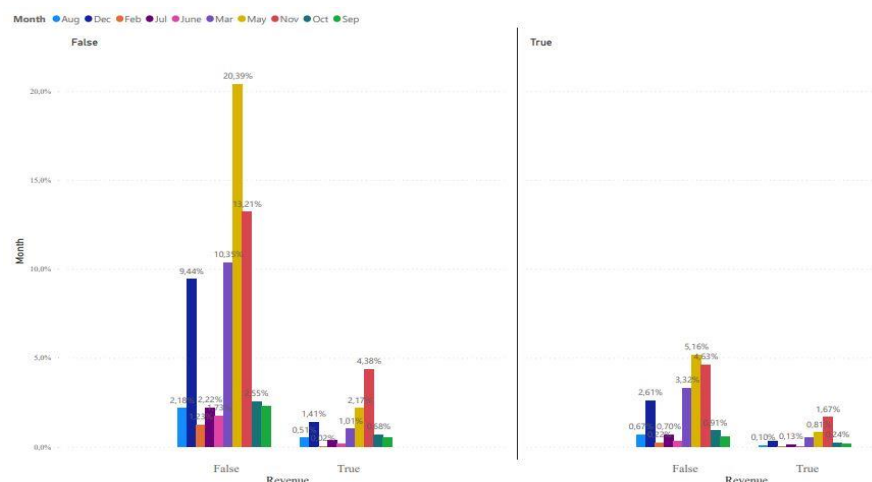
En la figura 3 se describe el porcentaje de compradores que concretaron la transacción con el valor de verdadero y los que no efectuaron compras con Falso, para el efecto se visualiza que el mes de noviembre el 6.06% de visitantes si efectuaron compras, a diferencia del mes de mayo que a pesar de haber tenido mayor cantidad de visitas solo el 2.98% concreto la compra. Se observa a través de la línea de tendencia que el atributo Pages Values tiende a incrementar en noviembre y diciembre, lo que representa que estos meses la mayor cantidad de usuarios ha finalizado una transacción de comercio electrónico.

Figura 1. Porcentaje de sesiones en las cuales se concretó una transacción por mes



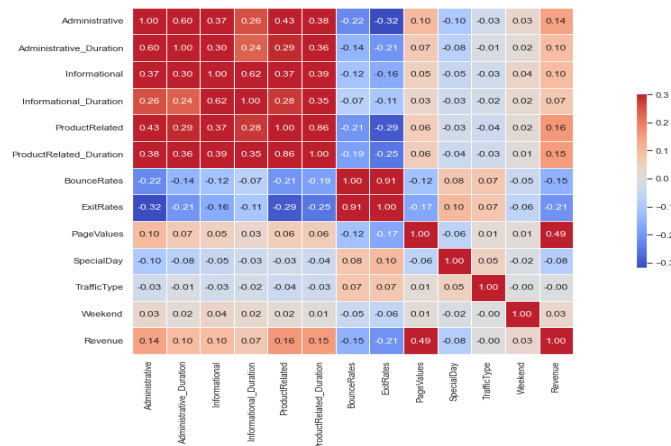
La figura 4 permite representar en el gráfico izquierdo la tendencia de visitantes que concretaron o no una compra en fin de semana y a la derecha los que han generado transacciones de compra los fines de semana, con lo que se puede destacar que en los días laborables del mes de mayo se efectuaron la mayor cantidad de transacciones, seguido de noviembre, marzo y diciembre.

Figura. 2. Porcentaje de clientes que concretaron una compra según el mes, divididos en fines de semana y días laborables



Se evaluó la clase Revenue utilizando el método de correlación de Pearson's, con el mismo que se identificó el factor de correlación entre los atributos del dataset y la clase. En base a este análisis se procede a eliminar los atributos month, informational_duration, weekend, browser, operatingsystems, traffictype.

Figura 3. Matriz de correlación



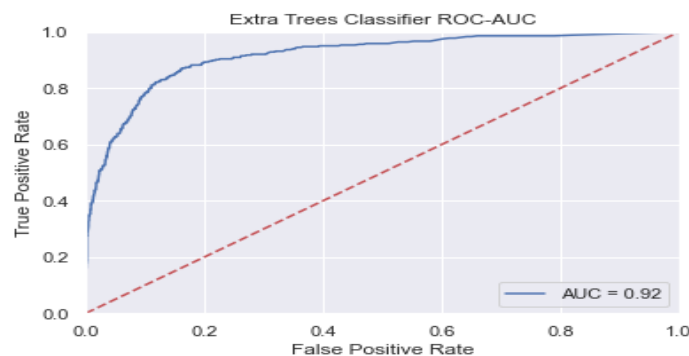
Resultados.

Tabla 2. Modelos y exactitud con datos de prueba

Model	Accuracy(in %)
Extra Trees Classifier model accuracy	93.99%
Gaussian Naive Bayes model accuracy	78.35%
Random Forest Classifier model accuracy	87.24%

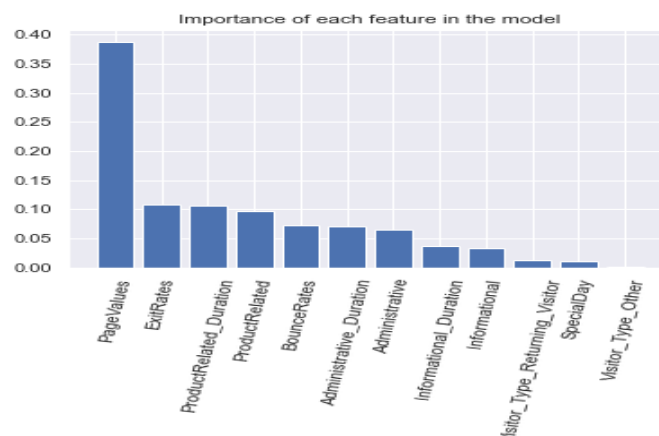
El AUC (Area Under Curve) fue la métrica seleccionada para cuantificar el comportamiento de los clasificadores(Muschelli, 2020). Se deriva del cálculo del área bajo la curva ROC (Receiver Operating Characteristics), la cual es resultado de trazar la tasa de verdaderos positivos contra la tasa de falsos negativos. Diferentes estudios han usado el AUC como métrica para conocer el desempeño de los clasificadores y demostrado sus ventajas, aunque otros han señalado sus deficiencias. Esta medida puede ser engañosa principalmente para conjuntos de datos desbalanceados donde la clase minoritaria representa menos del 10% del conjunto de datos. Es por ese motivo que en este dataset primero se aplicó la técnica de sobremuestreo. Para el entrenamiento del modelo se utilizaron el 70% de los datos y para las pruebas se usaron los datos locales, los cuales corresponden a un 30 % del total de las sesiones del dataset. En la figura 4 se genera una curva ROC, para el clasificador de mejor rendimiento, verificando el área under the ROC curve, para el efecto se usó la función `sklearn.metrics.roc_curve`.

Figura 4. Representa la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos (“1 – especificidad”)



El bosque aleatorio y los árboles adicionales también permiten estimar la importancia de cada característica en el modelo, por lo que se muestra en la figura 6 la importancia según los atributos identificados como más significativos.

Figura 5. Importancia de atributos en relación con el modelo ejecutado

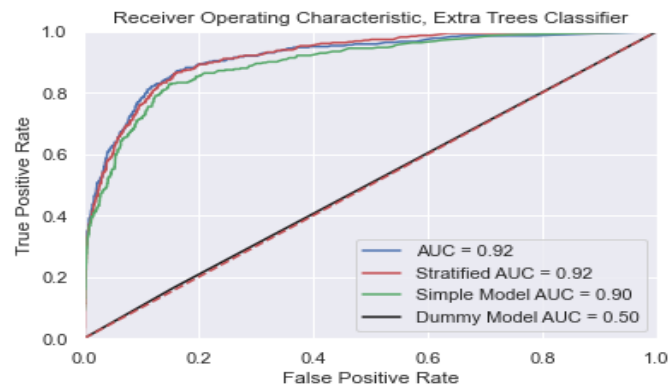


Los resultados de la ejecución del modelo se presentan a través de un análisis de características operativas del receptor (ROC), el cual es un enfoque gráfico para analizar el rendimiento de un clasificador (Tan, 2009). En este se muestra el desempeño relativo del modelo Extra Trees Classifier, generando un AUC de 0.92.

Para finalizar se utiliza DummyClassifier el cual permitió comparar el comportamiento del clasificador Extra Trees Classifier, con un clasificador ficticio que maneja las conjeturas basadas en un conjunto de datos estratificado, con lo que se obtiene que nuestro modelo genera un rendimiento bastante alto y que permite deducir que puede ser utilizado

para como modelo de predicción en conjunto con los atributos identificados de mayor importancia.

Figura 6. Comparación del modelo incluyendo DummyClassifier



Conclusiones

El comercio electrónico, es la acción de efectuar transacciones de compra-venta de bienes y servicios en línea, es uno de los muchos campos revolucionados por la ciencia de datos (Chaubey et al., 2022). Es importante para las empresas de comercio electrónico incrementar el porcentaje de visitantes del sitio web que completan la compra en las tiendas en línea. Para lograr este objetivo, las empresas de comercio electrónico, así como los investigadores de la academia, han dedicado esfuerzos en analizar y modelar los comportamientos de los usuarios de la página web. Especialmente en los últimos años, ha habido una tendencia en la investigación a utilizar métodos de aprendizaje automático para predecir el comportamiento de los usuarios.

En esta investigación se plantea el análisis de algoritmos de clasificación y proponemos un modelo de predicción de compra para predecir la intención de finalizar una compra en una sesión. El modelo propuesto se prueba con datos de un sitio web de Ecuador que mantiene sucursales en los cantones principales de la provincia del Guayas. El motivo de este artículo es ayudar e ilustrar cómo se puede trabajar con datos etiquetados y modelos de clasificación.

Los experimentos realizados demuestran que los modelos de aprendizaje automático generan un excelente rendimiento en el proceso de identificación de la intención de compra, por lo que se concluye que el modelo Extra Trees Classifier puede ser usado para ayudar a las tiendas a generar una experiencia de compra personalizada, a través de recomendaciones de productos y generación de ofertas en relación con los datos de cada sesión.

Referencias bibliográficas

- CEPAL. (2020). *Recuperación económica tras la pandemia COVID-19: empoderar a América Latina y el Caribe para un mejor aprovechamiento del comercio electrónico y digital*. <https://www.kas.de/de/web/regionalprogramm-adela>
- Charry, K., Coussement, K., Demoulin, N., & Heuvinck, N. (2020). Descriptive Analysis. In *Marketing Research with IBM® SPSS Statistics*. <https://doi.org/10.4324/9781315525532-2>
- Chaubey, G., Gavhane, P. R., Bisen, D., & Arjaria, S. K. (2022). Customer purchasing behavior prediction using machine learning classification techniques. *Journal of Ambient Intelligence and Humanized Computing*. <https://doi.org/10.1007/s12652-022-03837-6>
- Jílková, P., & Králová, P. (2021). Digital Consumer Behaviour and eCommerce Trends during the COVID-19 Crisis. *International Advances in Economic Research*, 27(1), 83–85. <https://doi.org/10.1007/s11294-021-09817-4>
- Lim, Y. J., Osman, A., Salahuddin, S. N., Romle, A. R., Abdullah, S., Muda, M., Mohd, R., & Hassan, S. (2016). Online Purchase Behavior of Generation Y in Malaysia. *Procedia Economics and Finance*, 37(16), 292–298. [https://doi.org/10.1016/s2212-5671\(16\)00050-2](https://doi.org/10.1016/s2212-5671(16)00050-2)
- Muschelli, J. (2020). ROC and AUC with a Binary Predictor: a Potentially Misleading Metric. *Journal of Classification*, 37(3), 696–708. <https://doi.org/10.1007/s00357-019-09345-1>
- Sakar C. & Kastro, Y. (2018). *Online Shoppers Purchasing Intention Dataset*.
- Statista. (2021a). *eCommerce report 2021* | . <https://www.statista.com/study/42335/e-commerce-report/>
- Statista. (2021b). *Ecuador: online shopping devices 2021* | Statista. <https://www.statista.com/statistics/921189/ecuador-online-purchases-device/>
- Tan, P.-N. (2009). *Receiver Operating Characteristic BT - Encyclopedia of Database Systems* (L. LIU & M. T. ÖZSU (eds.); pp. 2349–2352). Springer US.



https://doi.org/10.1007/978-0-387-39940-9_569

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés



CC BY-NC-ND 4.0

<https://creativecommons.org/licenses/by-nc-nd/4.0/>