



ANÁLITICA PREDICTIVA EN PREFERENCIAS AUTOMOTRICES: LA REGRESIÓN MULTINOMIAL EN LA INCLINACIÓN DE ESTUDIANTES

Elba Mendoza Macias

Instituto Politécnico Nacional, Escuela Superior de Cómputo
emendozam@ipn.mx

Edgardo Adrián Franco Martínez

Instituto Politécnico Nacional, Escuela Superior de Cómputo
edfrancom@ipn.mx

Gerardo Suarez Santos

Instituto Politécnico Nacional, Escuela Superior de Cómputo
gsuarezs1800@alumno.ipn.mx

Resumen

El crecimiento de la industria automotriz ha impulsado la adopción de herramientas digitales y técnicas de ciencia de datos, especialmente tras la pandemia de 2020 que llevó a una migración hacia entornos digitales. Aunque las estrategias basadas en datos han mejorado la eficiencia y seguridad de los automóviles, su aplicación en México aún es limitada, adaptando modelos desarrollados en países avanzados. Este estudio se enfoca en identificar las preferencias automovilísticas de estudiantes universitarios, considerando factores socioeconómicos y características del vehículo. Se usó un modelo de regresión logística multinomial con un 80% de confianza, revelando que los estudiantes prefieren autos híbridos, sedanes, y marcas como Nissan, con énfasis en confiabilidad mecánica y eficiencia de combustible.

Palabras clave: analítica predictiva, preferencias automotrices, regresión multinomial.

La industria automotriz en el mundo crece día con día, expandiendo cada vez más los diversos sectores y mercados que surgen conforme nuevas necesidades o para satisfacer diversos requerimientos que no eran considerados con anterioridad.

Sin embargo, el mercado también se encuentra en constante renovación, como en la pandemia del año 2020, donde las actividades económicas del mundo debieron ser migradas a un entorno digital,

principalmente la adquisición de bienes, entre ellos los automóviles.

Aunado a lo anterior, diversas empresas decidieron adoptar diversas herramientas de ciencias de datos o predictivas para acelerar sus ventas o mejorar sus estrategias, las cuales se basan en diversos factores del usuario final, como la actividad en línea para generar publicidad personalizada. A pesar de que estas llevan un tiempo implementándose, en mercados emergentes como México, se encuentran en desarrollo o simplemente no se



adaptan al mercado, utilizándose las versiones antes implementadas en países desarrollados.

El caso de México es muy particular, ya que el sector automotriz es uno de los mercados más grandes y de mayor expansión, tal es el caso que, en el año 2021, durante la emergencia sanitaria del Covid-19, este mismo creció un 38.54% (Estudio Contar, 2021), lo que sirvió de manifiesto para mostrar las nuevas necesidades de los consumidores para un transporte seguro y asequible.

Lo anterior resaltó la necesidad de usar un vehículo propio para evitar el contacto con el exterior y con otras personas, donde tendencias como las SUV (Vehículos Utilitarios Deportivos) estaban en auge antes de la pandemia, fue en decrecimiento en los últimos años, aumentando la demanda por subcompactos.

La implementación de la ciencia de datos en la industria automotriz ha permitido un crecimiento enorme en cuanto a innovaciones y avances en la misma, donde la producción se ha incrementado, la seguridad, eficiencia de los motores y las estrategias de venta se adaptan y mejoran con el pasar del tiempo, sin embargo, casos como el de México, hace que se requiera una adaptación al análisis del consumidor nacional.

Situación problemática

Dado esta necesidad de adaptar un modelo de ciencia de datos para el mercado mexicano, se tomará en cuenta las características del consumidor mexicano, en este caso el de los estudiantes universitarios del área metropolitana de la Ciudad de México, con objetivo de entender la preferencia en cuanto a características de los

automóviles que prefieren adquirir a forma de variables dependientes, entre ellos la marca, tipo de motor, cilindrada, tipo de chasis, capacidad de carga, entre otros.

Por el otro lado, se consideraron como variables independientes el perfil socioeconómico del individuo, los integrantes de su familia, edad, ocupación, distancia recorrida, presupuesto, etcétera. El uso de estas características permite comprender las preferencias en vehículos automotores en el determinado sector de la población antes mencionado.

Metodología

Dado que el objeto de estudio de la presente investigación está asociado a la identificación de características deseables que posee un vehículo en relación al perfil socioeconómico del cliente, se plantea un enfoque cuantitativo, de acuerdo con Sampieri et.al. (2014). Para realizar la recolección de los datos, se construyó un instrumento digital mediante la plataforma de Google Forms, enfocándose en una muestra representativa de las personas mayores a 18 años del área metropolitana de la Ciudad de México.

Se considero una muestra 40 individuos para dar validez a la recolección, todo esto tomando en cuenta un universo 200 estudiantes de ciencia de datos, por fines de tiempo y recursos se estableció un marco de confianza al 80%, además de un error de 15% de tolerancia, mismo que es perfecto para evitar el sobreajuste, por lo que la muestra es representativa de la población.

Para la limpieza y extracción de los datos se utilizó Microsoft Excel, gracias a la confiabilidad respaldada por los años de uso en diversas áreas, además de ser



estandarizado por su manejo de información en formatos tabulares.

El uso del lenguaje Python fue necesario para el procesamiento y construcción del modelo, gracias a la estabilidad que ofrece y la facilidad que presenta al uso de funciones estadísticas y de análisis de datos, aunado a una vasta comunidad y selección de librerías con diversas funciones (Gonzalez, 2023).

Resultados

Una vez realizada la aplicación del instrumento de recopilación de datos, se limpió y procesó la información con la finalidad de darle un tratamiento adecuado para su análisis, de esta forma, evaluar cuáles son los factores más determinantes en el perfil socioeconómico de un cliente para la compra de un automóvil.

La presencia de diversas marcas en el país es notoria, obteniendo una predominancia la marca japonesa Nissan, frente a la estadounidense General Motors, seguidos por automóviles de procedencia europea, japonesa, coreana, y china.

Lo cual es comprobable ya que los automóviles de los dueños participantes en la aplicación del instrumento son:

- Nissan Altima SE 2002
- Chevrolet Captiva 2012
- Kia Rio (Año desconocido)
- Chevrolet Aveo 2017
- Nissan Versa 2019
- Volkswagen Vento (Año desconocido)
- Chevrolet Meriva 2006
- Toyota Yaris Sedan 2022
- Chevrolet Spark (Año desconocido)

Todos ellos pertenecientes a estudiantes pertenecientes al Área metropolitana de la Ciudad de México, cuya ocupación principal es ser estudiante, con casos aislados de ocupaciones secundarias como técnicos y auxiliares administrativos.

El 50% de la muestra tienen familias de 4 personas, que además de su uso personal, también hacen uso del vehículo para recorridos familiares, seguido por un 19% con familias de 5 personas, el 14% de tres integrantes, 6 miembros o más con 9.5% y finalmente las de 3, con 4.8%. Mismos grupos que hacen traslados de más de una hora y en 15 minutos o menos.

La población encuestada, al momento de adquirir un automóvil, el 66.7% contempla un presupuesto de \$10,000 a \$50,000, mientras que el 14.3% de \$100,000 a \$200,000, seguido por rangos de \$200,000 a \$500,000 y \$50,000 a \$100,000 con el 9.5% cada uno respectivamente. Cabe mencionar que el 71.4% no sabe si tenía una disponibilidad de crédito, mientras que el 28.6% restante sabe que tiene una línea de crédito de \$10,000 a \$500,000.

En cuanto a preferencias de automóviles, los motores híbridos llevaron predominancia con el 47.8%, después los de gasolina con el 42.9% y el 9.5% restantes prefirieron eléctricos. En este caso, la cilindrada predilecta fue de 4, con el 55.6%, seguida de 5 con el 22.2% y finalmente de 6 y de 10 o más, con el 11.1% cada una.

Sin dudas, la carrocería preferida por el estudiante es el tipo sedán, con 28.6%, seguido por el coupé con el 19%, después hatchback y crossover con el 14.3% cada una, luego convertible con el 9.5% y finalmente miniván, SUV y coupé SUV con el 4.8% por tipo.



El aspecto más buscado en un automóvil fue la confiabilidad mecánica con el 61.9%, seguido por la eficiencia de combustible con un 28.6%, finalmente la calidad de interiores y sistemas de entretenimiento-sistema de audio, cada uno con el 4.8%.

Finalmente, la preferencia de marcas de automóviles es variada, como el mercado de automóviles mexicano, empatando en primer lugar Honda, Mazda, Volkswagen, KIA, Mercedes-Benz y Nissan siendo representado en la figura 1:

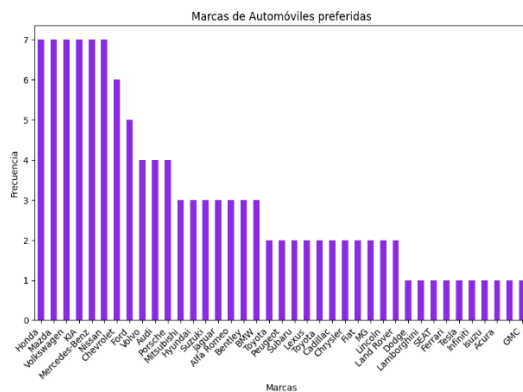


Figura 1. Marcas preferidas por la población.

El modelo predictivo seleccionado fue una regresión logística multinomial, dado el principio de la población con una confianza del 80%, fue una de las cosas acertadas, en donde el modelo final la obtuvo de igual forma, validando su construcción.

Informe de Clasificación:				
	precisión	recall	f1-score	support
No	0.75	1.00	0.86	3
Sí	1.00	0.50	0.67	2
accuracy			0.80	5
macro avg	0.88	0.75	0.76	5
weighted avg	0.85	0.80	0.78	5

Figura 2. Informe de clasificación del modelo de regresión logística multinomial.

En cuanto a predicciones, se trabajó con una matriz de confusión realizada por el mismo modelo con la finalidad de validación,

cuyos resultados demostró 3 predicciones positivas que son verdaderas, es decir, que se cumplen bajo las condiciones en las que fueron planteadas, tan solo arrojando dos predicciones incorrectas, por lo que se valida la confianza del 80%.

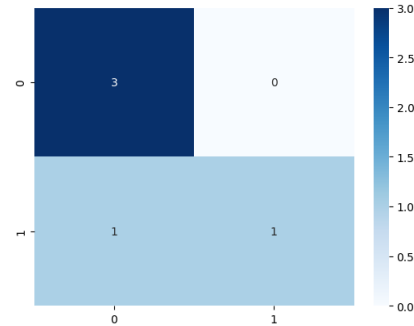


Figura 4. Matriz de confusión.

Conclusión

Un estudio del consumidor mexicano basado en ciencia de datos permite que la construcción de modelos de predicción funcione adecuadamente identificando las necesidades específicas de la población, sobre todo en un país donde la industria automotriz es de las más importantes.

Sin embargo, a pesar de los resultados positivos, esto abre las puertas a investigaciones y estudios futuros donde se utilice una cantidad de datos más grande y variada, sometiendo el modelo a diversos campos de la sociedad, permitiendo la adaptabilidad de estas herramientas a más grupos de personas cada vez más grandes.

La adopción de diversas técnicas como la aquí presentada permitiría a la industria automotriz expandirse de forma más rápida en el país, creando estrategias de ventas más enfocadas en el cliente, personalizadas y hechas a medida del público, permitiendo tener un acercamiento mayor con los posibles compradores y mejorando su



experiencia, además de fomentar la competitividad del mercado y crecimiento de la industria en general.

Referencias

Estudio Contar (12 de julio de 2021). Customer Journey Automotriz. Estudio Contar. <https://tinyurl.com/yckwz2s9>

Sampieri, R. H., Collado, C. F., Lucio, P. B., Valencia, S. M., & Torres, C. P. M. (2014). *Metodología de la investigación*.

Gonzalez, R. (2023). El papel de Python en el análisis de datos y la inteligencia artificial. *es.linkedin.com*. <https://tinyurl.com/5cfm9fxb>