



TRABAJO DE GRADO
Opción Seminario-Diplomado.

**ANÁLISIS DE DATOS Y APRENDIZAJE AUTOMÁTICO PARA LA
SEGMENTACIÓN DE CLIENTES EN UN CONTEXTO DE COMERCIO
ELECTRÓNICO, UTILIZANDO ALGORITMOS DE MACHINE LEARNING**

Corporación Universitaria Remington.
Nombre de la facultad: Ingenierías
Nombre del programa académico: Ingeniería de Sistemas

Estudiante:
Freddy Alexander Rangel Aponte

Tutor: Juan Carlos Briñez de León

Opción de Trabajo de grado Seminario-Diplomado.
2024.

Dedicatoria

A mi querida familia, por su apoyo incondicional, su amor constante y por ser mi mayor fuente de inspiración en cada paso de este camino. A mis amigos y colegas, por sus alentadoras palabras y por estar siempre presentes, especialmente cuando más los necesitaba. Este logro es el fruto de su paciencia, comprensión y confianza. Con todo mi cariño y gratitud, dedico este trabajo a aquellos que me impulsan a seguir adelante y me motivan a superarme cada día.

Agradecimientos

A Dios, por brindarme sabiduría y fortaleza en cada paso de este camino, y por darme la fe necesaria para superar los obstáculos.

A la universidad, por ser el lugar donde he crecido tanto académica como personalmente, y por proporcionarme las herramientas necesarias para forjar mi futuro.

A mis compañeros, por su apoyo constante, amistad y trabajo en equipo, haciendo de este proceso una experiencia valiosa y memorable.

Al profesor del seminario Juan Carlos Briñez de León, por su dedicación, paciencia y por compartir sus conocimientos con compromiso, orientándome con sabiduría en todo momento.

Y a todas las personas que, con sus palabras de aliento, ejemplo o contribuciones, grandes o pequeñas, me impulsaron a seguir adelante y dar lo mejor de mí.

Gracias a todos por ser parte esencial de este logro

Tabla de contenido

Resumen.....	7
Palabras clave.....	8
Pregunta problema	10
Acercamiento a los datos	10
Descripción de Variables	11
Variables utilizadas en el análisis	11
Transformación y Preprocesamiento de Variables:	12
Importancia de las Variables:.....	12
Objetivos.....	37
Objetivos Específicos:	37
Preparación y Análisis de los Datos:	38
Modelo de Toma de Decisiones:.....	38
Análisis de Desempeño y Validación del Modelo:.....	38
Conclusiones y Trabajos Futuros.....	39
Conclusiones:	39
Trabajos Futuros:	39
Referencias.....	41

Índice de Gráficas

Gráfica 1	Verificación de las opciones de la variable Estado Civil.....	17
Gráfica 2	Realizando tabla de frecuencia para una variable (Estado Civil)	20
Gráfica 3	Verificación de la opción de la variable Método de Pago	21
Gráfica 4	Gráfica Método de Pago	23
Gráfica 5	Revisando opciones de la variable Nivel de Satisfacción.....	24
Gráfica 6	Gráfico de densidad para una variable (Nivel de Satisfacción).....	26
Gráfica 7	Gráfico de caja y bigotes para una variable (Calificación promedio)	27
Gráfica 8	Matriz de correlaciones.....	28
Gráfica 9	Segmentación de datos (Aprendizaje No supervisado)	30
Gráfica 10	Creando modelo (KMeans es una librería para hacer clustering).....	31
Gráfica 11	Visualizar la distribución de los resultados	33
Gráfica 12	Usando el modelo en nuevos datos (Clientes)	35
Gráfica 13	Primera respuesta.....	36
Gráfica 14	Segunda Respuesta	36

Índice de Tablas

Tabla 1	Cargando y visualizando Datos	13
Tabla 2	Información de la estructura de datos	14
Tabla 3	Análisis de Datos	15
Tabla 4	Elimina filas que tengan datos nulos	16
Tabla 5	Mapeando todas las variables categóricas (Estado Civil) a numéricas.....	18
Tabla 6	Realizando tabla de frecuencia según el número de rangos (Estado Civil).....	19
Tabla 7	Mapeando todas las variables categóricas (Método de Pago) a numéricas	22
Tabla 8	Cambiando valores en la variable Nivel de Satisfacción.....	25
Tabla 9	Observando grupos según el modelo	32
Tabla 10	Analizando los patrones de los grupos.....	34

Resumen

Este trabajo de grado tiene como objetivo aplicar técnicas de análisis de datos y aprendizaje automático para la segmentación de clientes en un contexto de comercio electrónico, utilizando el algoritmo de KMeans. La segmentación de clientes permite identificar patrones y características comunes entre los clientes, lo que puede mejorar las estrategias de marketing y optimizar las campañas publicitarias.

Se utilizó un conjunto de datos que incluye diversas características de los clientes, como estado civil, ingresos, gastos en diferentes categorías de productos, método de pago, número de visitas al sitio web y nivel de satisfacción. Antes de aplicar el modelo de segmentación, se realizaron diversas tareas de preprocesamiento de datos, como la limpieza, la transformación de variables categóricas a numéricas, y la normalización de los datos para garantizar la correcta convergencia del algoritmo.

Se aplicó el método del codo para determinar el número óptimo de clusters, eligiendo 6 grupos como la segmentación más adecuada para este conjunto de datos. Posteriormente, el modelo de KMeans fue entrenado y validado, segmentando a los clientes según sus características. Los resultados fueron interpretados y validados, encontrando que los clientes con características similares fueron agrupados eficazmente, lo que facilita la personalización de estrategias comerciales y la mejora de la experiencia del cliente.

Finalmente, el modelo de segmentación fue implementado en un entorno práctico, permitiendo la asignación automática de nuevos clientes a grupos predefinidos, optimizando así las decisiones de marketing y proporcionando recomendaciones personalizadas para cada segmento.

Palabras clave

Segmentación de clientes, Machine Learning, Análisis de datos, Clustering (K-Means), Comercio electrónico, Optimización de marketing, Google Colab, Predicción de comportamiento de clientes, Estrategias personalizadas, Métodos de segmentación, Modelos predictivos, Análisis de patrones de consumo, Técnicas de clustering

Marco Conceptual y Contextual

El comercio electrónico ha experimentado un notable crecimiento en las últimas décadas, impulsado por avances tecnológicos y cambios en los hábitos de consumo. Este fenómeno ha permitido a las empresas recopilar grandes volúmenes de datos sobre sus clientes, lo que ha abierto nuevas oportunidades para personalizar la experiencia de compra y optimizar las estrategias de marketing. Sin embargo, uno de los principales retos a los que se enfrentan las empresas de comercio electrónico es cómo utilizar estos datos para segmentar a sus clientes de manera eficiente y, por lo tanto, mejorar el rendimiento de sus campañas comerciales (Chopra, Meindl, & Kalra, 2016).

La segmentación de clientes es el proceso de dividir un mercado en grupos de consumidores con características similares, lo que permite a las empresas personalizar sus estrategias de marketing y satisfacer mejor las necesidades específicas de cada segmento. El análisis de datos es una herramienta clave para lograr una segmentación eficaz, ya que proporciona información valiosa sobre los comportamientos, preferencias y características de los consumidores. A través del análisis de datos, las empresas pueden identificar patrones en los comportamientos de compra y ofrecer productos o servicios más personalizados, mejorando la experiencia del cliente (Russell & Norvig, 2021).

El uso de técnicas de machine learning ha revolucionado el campo de la segmentación de clientes, ya que permite analizar grandes volúmenes de datos y extraer patrones complejos que no serían detectados mediante métodos tradicionales. El aprendizaje automático posibilita que las máquinas aprendan de los datos y realicen predicciones sin ser explícitamente programadas para cada tarea específica. En el comercio electrónico, se utilizan técnicas como el clustering (por ejemplo, K-Means) para agrupar a los clientes según características comunes como ingresos, hábitos de compra y nivel de satisfacción con los productos y servicios ofrecidos (Aggarwal, 2016; Geron, 2019).

Plataformas como Google Colab, que permiten la implementación de modelos de machine learning, han democratizado el acceso a herramientas avanzadas de análisis de datos. Con la integración de bibliotecas como TensorFlow, scikit-learn y Keras, tanto empresas grandes como pequeñas pueden aprovechar el aprendizaje automático para mejorar la toma de decisiones comerciales y personalizar sus estrategias de marketing (Geron, 2019).

La optimización de las estrategias de marketing a través del análisis de datos es fundamental para las empresas de comercio electrónico. Utilizando los resultados de la segmentación, las empresas pueden diseñar campañas publicitarias personalizadas, promociones específicas y mejorar la experiencia del cliente, lo que a su vez puede aumentar las tasas de conversión y la satisfacción general. Además, el análisis de datos permite predecir el comportamiento de compra de los clientes, lo que ayuda a las empresas a tomar decisiones más informadas sobre inventarios y precios (Chigara & Smallwood, 2014; Russell & Norvig, 2021).

Pregunta problema

¿Cómo pueden las características demográficas, los hábitos de compra y los niveles de satisfacción de los clientes ser analizados a través de los modelos de Análisis de Datos y Segmentación de Datos, para identificar patrones que permitan segmentar eficazmente a los clientes en el comercio electrónico y optimizar las estrategias de marketing personalizadas?

Acercamiento a los datos

Se eligieron diversas variables para segmentar a los clientes, como estado civil, ingresos, gastos por categoría, frecuencia de compra, método de pago y nivel de satisfacción. Estas variables son clave para entender los patrones de compra y la lealtad de los clientes, facilitando una segmentación basada en su comportamiento.

El proceso de análisis comenzó con la limpieza de datos, utilizando la biblioteca pandas para identificar y eliminar valores faltantes o duplicados. Se eliminaron filas con datos nulos y se transformaron los tipos de datos para asegurar su adecuación. Las variables categóricas, como estado civil y método de pago, fueron codificadas numéricamente mediante one-hot encoding o label encoding, lo que facilitó su procesamiento por modelos de machine learning.

Después de la limpieza, se realizó una exploración inicial de los datos mediante estadísticas descriptivas y gráficos visuales, como barras, histogramas y diagramas de dispersión, usando bibliotecas como matplotlib y seaborn. Esto permitió identificar patrones y relaciones entre las variables. Por ejemplo, los clientes con mayores ingresos gastan más en ropa y tecnología, mientras que aquellos con ingresos bajos tienen niveles de satisfacción más altos, aunque gastan menos. Además, los pagos en efectivo se asociaron con mayor satisfacción en comparación con pagos con tarjeta o transferencia.

Finalmente, los datos fueron divididos en conjuntos de entrenamiento y prueba, y las variables numéricas se normalizaron utilizando MinMaxScaler o StandardScaler. Esto preparó los datos para aplicar técnicas de machine learning, como K-Means, en la segmentación de clientes, optimizando así las estrategias de marketing basadas en estos patrones.

Descripción de Variables

En esta sección se detallan las variables utilizadas en el análisis de los datos, así como su importancia para la segmentación de clientes y la optimización de las estrategias de marketing. Las variables fueron seleccionadas con el objetivo de capturar características relevantes que pudieran influir en el comportamiento de compra de los clientes en el comercio electrónico.

Variables utilizadas en el análisis

El análisis se basó en varias variables que permiten segmentar y comprender el comportamiento de los clientes. La variable "Estado Civil" es categórica y fue transformada numéricamente para identificar patrones de compra relacionados con este factor. "Ingreso" representa el nivel de ingreso mensual de los clientes, y se considera relevante para analizar el gasto en productos como ropa y tecnología. "Método de Pago" describe las opciones de pago preferidas por los clientes (efectivo, tarjeta de crédito, transferencia), transformándose en una variable numérica para analizar su relación con la satisfacción.

La variable "Satisfacción" mide el nivel de satisfacción del cliente con el servicio, utilizando una escala del 1 al 5, y se considera clave para identificar clientes leales. "Categoría de Compra" detalla las categorías de productos adquiridos (ropa, tecnología, alimentos, etc.) y se utiliza para analizar preferencias de compra. "Frecuencia de Compra" mide la regularidad con la que los clientes compran en línea, lo que permite identificar clientes recurrentes y diseñar estrategias de fidelización.

Por último, la "Calificación Promedio" refleja la evaluación general de la tienda por parte de los clientes, ayudando a evaluar la calidad del servicio y su relación con la satisfacción y compras repetidas. Cada variable fue tratada adecuadamente para su análisis y segmentación en los modelos.

Transformación y Preprocesamiento de Variables:

- Variables categóricas como estado civil, método de pago, y categoría de compra fueron transformadas en variables numéricas a través de técnicas de codificación como one-hot encoding o label encoding. Esto permitió que los algoritmos de machine learning pudieran procesar estas variables de manera efectiva.
- Variables numéricas, como ingreso y frecuencia de compra, fueron normalizadas utilizando StandardScaler o MinMaxScaler para asegurarse de que todas las variables tuvieran la misma escala, lo que es importante cuando se utilizan algoritmos de machine learning, como K-Means.

Importancia de las Variables:

- **Ingreso y Frecuencia de Compra:** Son dos de las variables más importantes para segmentar a los clientes, ya que se espera que los clientes con mayores ingresos tiendan a gastar más en categorías específicas de productos, mientras que la frecuencia de compra es un buen indicador de la lealtad del cliente y su propensión a repetir compras.
- **Método de Pago y Satisfacción:** Ambas variables están relacionadas con la experiencia del cliente y son clave para diseñar estrategias de fidelización. Los clientes más satisfechos tienden a preferir métodos de pago que les resulten más cómodos y seguros, lo que a su vez influye en su comportamiento de compra.

- **Categoría de Compra:** Permite identificar qué tipos de productos son más demandados por cada segmento de clientes, lo que ayuda a personalizar las ofertas y promociones.

Aproximaciones con gráficos - analítica.

Cargando y visualizando Datos

Tabla 1

Cargando y visualizando Datos

Warning: Cannot open archive seleccionados.upgrado widget as only available when the user has been executed in the current browser session. Please refer this user to resolve.

[] Saving Ecomercialdataset-1 (1).xlsx to Ecomercialdataset-1 (1).xlsx

	Estado civil	Ingreso	Ropa	Tecnología y electrónica	Libros	Muebles	Alimentos	Cosméticos	Num_MebCompras	Número de visitas web por mes	Método de pago	Nivel de satisfacción	Calificación promedio
0	Soltero/a	58138	635	88	546	172	88	88	8	7	Pago en efectivo contra Entrega	Satisfecho	4.5
1	Soltero/a	46344	11	1	6	2	1	6	1	5	Tarjetas de crédito y débito	Insatisfecho	4.1
2	Unión Libre	71613	426	49	127	111	21	42	8	4	Transferencias bancarias	Insatisfecho	3.6
3	Unión Libre	26646	11	4	20	10	3	5	2	6	Tarjetas de crédito y débito	Insatisfecho	3.6
4	Casado/a	58293	173	43	118	46	27	15	5	5	Transferencias bancarias	Satisfecho	4.9
5	Unión Libre	62513	520	42	98	0	42	14	6	6	Tarjetas de crédito y débito	Insatisfecho	3.4
6	Divorciado/a	55635	235	65	164	50	49	27	7	6	Transferencias bancarias	Insatisfecho	3.8
7	Casado/a	33454	76	10	56	3	1	23	4	8	Pago en efectivo contra Entrega	Insatisfecho	3.6
8	Unión Libre	30351	14	0	24	3	3	2	3	9	Tarjetas de crédito y débito	Satisfecho	4.5
9	Unión Libre	5648	28	0	6	1	1	13	1	20	Pago en efectivo contra Entrega	Satisfecho	4.5

Fuente. Imagen obtenida de Google Colab.

La tabla 1, presenta registros de clientes con información clave sobre sus comportamientos de compra y características demográficas. Cada fila representa a un cliente y las columnas incluyen datos como el estado civil, ingresos anuales, y el gasto en diferentes categorías de productos como ropa, tecnología, y alimentos. También se incluyen detalles sobre el número de visitas al sitio web, método de pago preferido, nivel de satisfacción (satisfecho o insatisfecho) y calificación promedio dada al producto o servicio. El objetivo de la tabla es segmentar a los clientes según su estado civil, nivel de satisfacción y otros factores, como su método de pago y frecuencia de compras. Estas variables serán clave para realizar el análisis de segmentación de clientes.

Información de la estructura de datos

Tabla 2

Información de la estructura de datos

```

#Información de la estructura de datos
Conjunto_Datos.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2203 entries, 0 to 2202
Data columns (total 13 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Estado civil                             2203 non-null   object
1   Ingreso                                   2203 non-null   int64
2   Ropa                                       2203 non-null   int64
3   Tecnología y electrónica                 2203 non-null   int64
4   Libros                                    2203 non-null   int64
5   Muebles                                   2203 non-null   int64
6   Alimentos                                 2203 non-null   int64
7   Cosméticos                               2203 non-null   int64
8   Num_WebCompras                           2203 non-null   int64
9   Número de visitas web por mes           2203 non-null   int64
10  Método de pago                            2203 non-null   object
11  Nivel de satisfacción                     2191 non-null   object
12  Calificación promedio                    2203 non-null   float64
dtypes: float64(1), int64(9), object(3)
memory usage: 223.9+ KB

```

Fuente. Imagen obtenida de Google Colab.

En la tabla 2, la salida de la función `pandas.DataFrame.info()` muestra un resumen del conjunto de datos, que contiene 2,203 registros y 13 columnas. Este resumen es clave para comprender la estructura del DataFrame y los tipos de datos. Las columnas incluyen variables como estado civil, ingresos, y gastos en categorías como ropa y tecnología. También se registran datos sobre el número de visitas al sitio web, método de pago, nivel de satisfacción y calificación promedio. No hay valores nulos, lo que asegura la integridad de los datos para el análisis. Las variables numéricas están representadas como `int64`, mientras que las categóricas como estado civil y método de pago son de tipo `object`, por lo que podrían necesitar conversión a valores numéricos para análisis.

Análisis de Datos

Tabla 3

Análisis de Datos

	Ingreso	Ropa	Tecnología y electrónica	Libros	Muebles	Alimentos	Cosméticos	Num_webCompras	Número de visitas web por mes	Calificación promedio
count	2203.000000	2203.000000	2203.000000	2203.000000	2203.000000	2203.000000	2203.000000	2203.000000	2203.000000	2203.000000
mean	51972.343622	305.374489	26.369950	167.212892	37.563777	27.019973	43.829778	4.083522	5.320472	4.016659
std	21547.871435	337.911904	39.829607	224.468148	54.607400	41.044542	51.697002	2.741731	2.425993	0.580221
min	1730.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	3.000000
25%	35221.000000	24.000000	2.000000	16.000000	3.000000	1.000000	9.000000	2.000000	3.000000	3.500000
50%	51390.000000	174.000000	8.000000	68.000000	12.000000	8.000000	24.000000	4.000000	6.000000	4.100000
75%	68557.000000	505.000000	33.000000	233.500000	50.000000	33.000000	56.000000	6.000000	7.000000	4.500000
max	162397.000000	1493.000000	199.000000	1725.000000	259.000000	262.000000	321.000000	27.000000	20.000000	4.900000

Fuente. Imagen obtenida de Google Colab.

La Tabla 3, muestra estadísticas descriptivas de un conjunto de datos sobre el comportamiento de los consumidores en varias categorías de gasto. Los ingresos promedio son de 5175, con una gran dispersión, ya que algunos usuarios tienen ingresos mucho más bajos o altos. En cuanto al gasto, se observa una gran variabilidad en categorías como ropa (promedio de 2053) y tecnología (promedio de 365), lo que sugiere que los consumidores tienen hábitos de compra muy diversos. El gasto en libros, muebles y alimentos también varía significativamente, con una desviación estándar considerable, lo que indica que no todos los consumidores gastan de manera similar en estas categorías.

El número promedio de compras es de 4.84, pero nuevamente, con una alta variabilidad entre los consumidores. Las visitas a sitios web de comercio electrónico promedian 2.52 por mes, con diferencias significativas entre los usuarios. Finalmente, las calificaciones promedio otorgadas a los productos son relativamente altas, con una media de 4.01 y baja variabilidad, lo que sugiere una satisfacción generalizada.

Este análisis resalta la heterogeneidad en los hábitos de consumo y puede ser útil para segmentar a los usuarios y desarrollar estrategias de marketing personalizadas.

Elimina filas que tengan datos nulos

Tabla 4

Elimina filas que tengan datos nulos

	Estado civil	Ingreso	Ropa	Tecnología y electrónica	Libros	Muebles	Alimentos	Cosméticos	Num_WebCompras	Número de visitas web por mes	Método de pago	Nivel de satisfacción	Calificación promedio
0	Soltero/a	58138	635	88	546	172	88	88	8	7	Pago en efectivo contra Entrega	Satisfecho	4.5
1	Soltero/a	46344	11	1	6	2	1	6	1	5	Tarjetas de crédito y débito	Insatisfecho	4.1
2	Unión Libre	71613	426	49	127	111	21	42	8	4	Transferencias bancarias	Insatisfecho	3.6
3	Unión Libre	26646	11	4	20	10	3	5	2	6	Tarjetas de crédito y débito	Insatisfecho	3.6
4	Casado/a	58293	173	43	118	46	27	15	5	5	Transferencias bancarias	Satisfecho	4.9

Fuente. Imagen obtenida de Google Colab.

La tabla 4, muestra un conjunto de datos que incluye varias características de los clientes, como Estado Civil, Ingresos, Categorías de productos, Método de pago, Número de visitas web por mes, Nivel de satisfacción, entre otras. Esta tabla representa una versión filtrada del conjunto de datos original, donde se han eliminado las filas con datos nulos, lo que garantiza que solo se utilicen registros completos para el análisis posterior.

- Eliminación de filas nulas: Al filtrar los datos, se han descartado aquellos registros donde faltaban valores en alguna de las columnas. Esto es esencial para asegurar la integridad de los análisis y evitar sesgos en los resultados.
- Observación de los datos completos: Ahora, cada fila contiene información completa sobre el cliente, lo que facilita la segmentación y análisis sin la interferencia de datos faltantes.

Este paso de limpieza es crucial para preparar los datos para su análisis, evitando errores que podrían surgir si se incluyeran valores nulos en el análisis o el modelado posterior.

Verificación de las opciones de la variable Estado Civil

Gráfica 1

Verificación de las opciones de la variable Estado Civil

```
#Verificación de las opciones de la variable
print('Estado civil')
Conjunto_Datos['Estado civil'].unique()

Estado civil
array(['Soltero/a', 'Unión Libre', 'Casado/a', 'Divorciado/a', 'Viudo/a'],
      dtype=object)
```

Fuente. Imagen obtenida de Google Colab.

En esta Gráfica (1), se utiliza el método `.unique()` de Pandas para verificar las distintas categorías presentes en la variable "Estado Civil". El resultado muestra las opciones únicas disponibles en esta columna:

- Soltero/a
- Unión Libre
- Casado/a
- Divorciado/a
- Viudo/a

Este paso es crucial para identificar las categorías presentes en los datos y asegurarse de que la variable "Estado Civil" esté correctamente definida antes de proceder con su análisis o codificación para modelos de machine learning. Además, el código muestra que los valores de esta variable son de tipo `object`, lo que indica que son cadenas de texto.

Este análisis inicial es importante para realizar una verificación de la calidad de los datos y asegurarse de que no haya valores inesperados o erróneos en esta columna antes de su uso en el modelo.

Mapeando todas las variables categóricas (Estado Civil) a numéricas

Tabla 5

Mapeando todas las variables categóricas (Estado Civil) a numéricas

	Estado civil	Ingreso	Ropa	Tecnología y electrónica	Libros	Muebles	Alimentos	Cosméticos	Num_WebCompras	Número de visitas web por mes	Método de pago	Nivel de satisfacción	Calificación promedio
0	1	58138	635	88	546	172	88	88	8	7	Pago en efectivo contra Entrega	Satisfecho	4.5
1	1	46344	11	1	6	2	1	6	1	5	Tarjetas de crédito y débito	Insatisfecho	4.1
2	2	71613	426	49	127	111	21	42	8	4	Transferencias bancarias	Insatisfecho	3.6
3	2	26646	11	4	20	10	3	5	2	6	Tarjetas de crédito y débito	Insatisfecho	3.6
4	3	58293	173	43	118	46	27	15	5	5	Transferencias bancarias	Satisfecho	4.9

Fuente. Imagen obtenida de Google Colab.

La tabla 5, muestra cómo la variable "Estado Civil" ha sido convertida de una categoría textual a una representación numérica, un paso esencial para poder usar la variable en modelos de machine learning. En este proceso, las categorías del Estado Civil (como Soltero/a, Unión Libre, Casado/a, etc.) se asignan a valores numéricos.

Por ejemplo, es probable que se haya hecho un mapeo de valores como:

- Soltero/a = 0
- Unión Libre = 1
- Casado/a = 2
- Divorciado/a = 3
- Viudo/a = 4

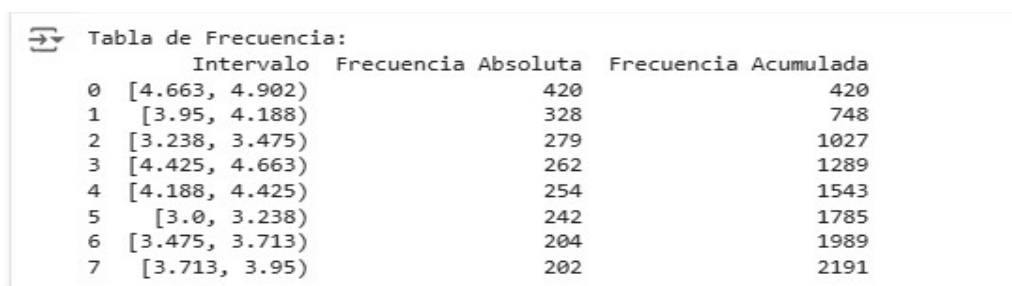
Este mapeo transforma las variables categóricas en números enteros o binarios, lo que permite a los algoritmos de machine learning procesarlas de manera eficiente. La columna Estado Civil ahora contiene valores numéricos, y la tabla refleja cómo esta transformación facilita el análisis y modelado de los datos.

Este paso es esencial cuando se trabaja con datos de texto, ya que los modelos de machine learning generalmente requieren que las variables sean numéricas para realizar los cálculos de manera efectiva.

Realizando tabla de frecuencia según el número de rangos (Estado Civil)

Tabla 6

Realizando tabla de frecuencia según el número de rangos (Estado Civil)



	Intervalo	Frecuencia Absoluta	Frecuencia Acumulada
0	[4.663, 4.902)	420	420
1	[3.95, 4.188)	328	748
2	[3.238, 3.475)	279	1027
3	[4.425, 4.663)	262	1289
4	[4.188, 4.425)	254	1543
5	[3.0, 3.238)	242	1785
6	[3.475, 3.713)	204	1989
7	[3.713, 3.95)	202	2191

Fuente. Imagen obtenida de Google Colab.

La tabla 6, muestra el análisis de frecuencia de los datos de Estado Civil divididos en varios intervalos. Se presentan tres columnas principales:

1. Intervalo: El rango de valores en el que se agrupan las observaciones de Estado Civil.

2. Frecuencia Absoluta: El número de registros que caen dentro de cada intervalo. Por ejemplo, el intervalo [4.663, 4.902) tiene 420 registros.

3. Frecuencia Acumulada: El número total acumulado de registros hasta ese intervalo. Por ejemplo, para el intervalo [4.663, 4.902), la frecuencia acumulada es 420. En el intervalo siguiente [3.95, 4.188), la frecuencia acumulada llega a 748.

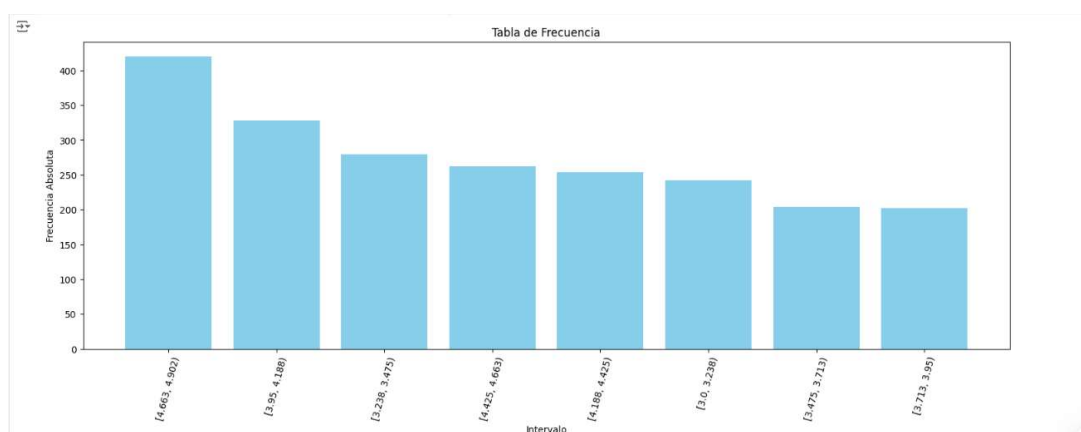
Este tipo de tabla es útil para visualizar la distribución de los datos dentro de ciertos rangos y permite detectar tendencias o concentraciones de datos en intervalos específicos. También ayuda a preparar los datos para su análisis posterior, como la segmentación o modelado.

Este análisis es crucial para tener una visión más clara de cómo los datos están distribuidos y permite hacer ajustes en las estrategias de marketing o segmentación basadas en esta distribución.

Realizando tabla de frecuencia para una variable (Estado Civil)

Gráfica 2

Realizando tabla de frecuencia para una variable (Estado Civil)



Fuente. Imagen obtenida de Google Colab.

Este gráfico (2) de barras muestra las frecuencias absolutas de los intervalos de la variable Estado Civil, con cada barra representando un rango y la cantidad de registros dentro de ese intervalo. El eje X muestra los intervalos, como [4.663, 4.902), y el eje Y indica la frecuencia de registros en cada intervalo.

Análisis:

- El intervalo [4.663, 4.902) tiene la frecuencia más alta, con más de 400 registros, lo que indica que la mayoría de los clientes tienen calificaciones dentro de este rango. A medida que los intervalos se alejan de este rango, las frecuencias disminuyen, sugiriendo que los datos están más concentrados en algunos grupos específicos y dispersos en otros.
- La forma del gráfico sugiere una distribución sesgada a la derecha, donde los valores altos son más frecuentes que los bajos.

Importancia del Gráfico:

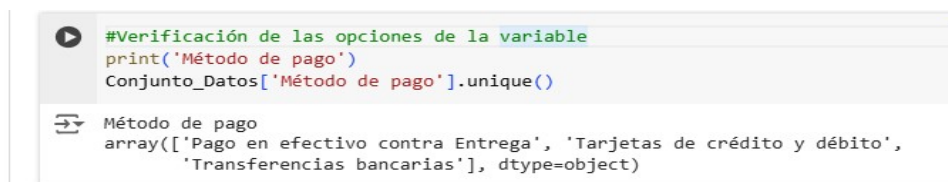
- Este gráfico facilita la comprensión de la distribución de los datos y muestra en qué rangos se concentra la mayoría de los registros.
- Ayuda a tomar decisiones estratégicas de segmentación y marketing, enfocando los esfuerzos en los grupos más grandes.
- Permite una asignación eficiente de recursos, maximizando el impacto de las campañas.

En síntesis, el gráfico es crucial para entender la distribución de los datos y adaptar las estrategias comerciales a los segmentos más relevantes.

Verificación de la opción de la variable Método de Pago

Gráfica 3

Verificación de la opción de la variable Método de Pago



```
#Verificación de las opciones de la variable
print('Método de pago')
Conjunto_Datos['Método de pago'].unique()

Método de pago
array(['Pago en efectivo contra Entrega', 'Tarjetas de crédito y débito',
      'Transferencias bancarias'], dtype=object)
```

Fuente. Imagen obtenida de Google Colab.

En esta Gráfica (3), se utiliza el método `.unique()` de Pandas para identificar las categorías únicas en la variable "Método de Pago". Los valores resultantes son:

- Pago en efectivo contra Entrega
- Tarjetas de crédito y débito
- Transferencias bancarias

Este paso es fundamental para verificar que los valores de la variable sean consistentes y no contengan errores o valores inesperados antes de proceder con su análisis. Además, asegura que la variable esté correctamente definida y libre de valores faltantes o incorrectos, lo cual es

esencial para garantizar la calidad del análisis y la efectividad de los modelos de machine learning que se apliquen posteriormente.

Mapeando todas las variables categóricas (Método de Pago) a numéricas

Tabla 7

Mapeando todas las variables categóricas (Método de Pago) a numéricas

	Estado civil	Ingreso	Ropa	Tecnología y electrónica	Libros	Muebles	Alimentos	Cosméticos	Num_WebCompras	Número de visitas web por mes	Método de pago	Nivel de satisfacción	Calificación promedio
0	1	58138	635	88	546	172	88	88	8	7	1	Satisfecho	4.5
1	1	46344	11	1	6	2	1	6	1	5	2	Insatisfecho	4.1
2	2	71613	426	49	127	111	21	42	8	4	3	Insatisfecho	3.6
3	2	26646	11	4	20	10	3	5	2	6	2	Insatisfecho	3.6
4	3	58293	173	43	118	46	27	15	5	5	3	Satisfecho	4.9

Fuente. Imagen obtenida de Google Colab.

En esta tabla 7, la variable "Método de Pago" ha sido convertida de texto (categórica) a valores numéricos para facilitar su uso en modelos de machine learning. Las categorías de "Método de Pago" incluyen:

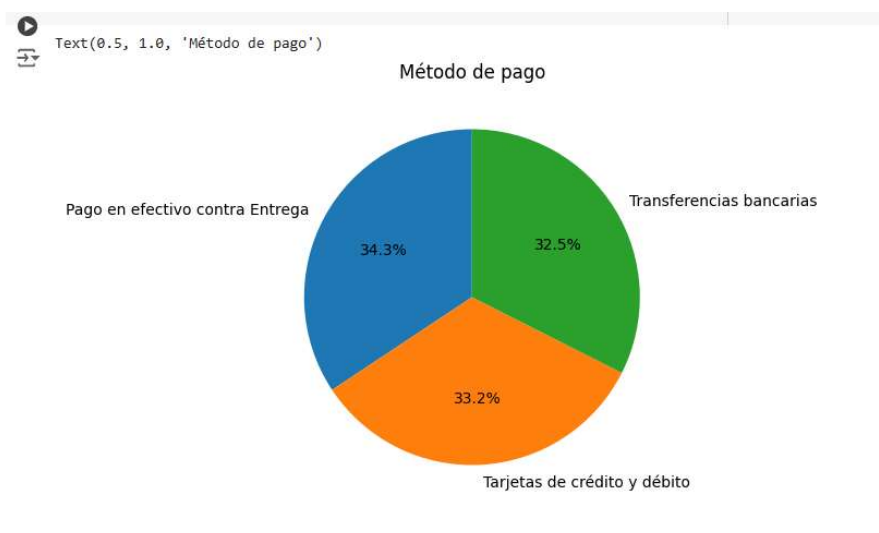
- Pago en efectivo contra entrega
- Tarjetas de crédito y débito
- Transferencias bancarias

En el proceso de mapeo, cada categoría se asigna a un número (por ejemplo, 0 para pago en efectivo, 1 para tarjetas de crédito y débito, y 2 para transferencias bancarias). Este paso es crucial para convertir las variables categóricas en números que pueden ser procesados por los algoritmos de machine learning, mejorando así el análisis y la creación de modelos predictivos. El mapeo asegura que las variables categóricas no interfieran con el proceso de modelado y que el algoritmo pueda aprovechar la información contenida en ellas.

Gráfica Método de Pago

Gráfica 4

Gráfica Método de Pago



Fuente. Imagen obtenida de Google Colab.

La Gráfica 4, muestra un gráfico circular que visualiza la distribución de los métodos de pago utilizados por los clientes en el conjunto de datos, dividido en tres sectores:

- **Pago en efectivo contra entrega:** 34.3% de los clientes.
- **Tarjetas de crédito y débito:** 33.2% de los clientes.
- **Transferencias bancarias:** 32.5% de los clientes.

Análisis:

- El gráfico indica que los métodos de pago están bastante equilibrados, con cada uno representando una porción significativa del total.
- El **pago en efectivo contra entrega** es ligeramente más popular, sugiriendo que muchos clientes prefieren pagar al recibir el producto, posiblemente por comodidad o confianza.
- Las **tarjetas de crédito y débito** son una opción importante, reflejando una adopción significativa de pagos electrónicos.

- Las **transferencias bancarias**, aunque con un porcentaje menor, siguen siendo relevantes para una parte de los clientes.

Importancia del Gráfico:

- **Comprender las preferencias de pago:** El gráfico ayuda a las empresas a conocer qué métodos de pago prefieren sus clientes, lo que puede influir en las opciones futuras.
- **Optimizar la experiencia de compra:** Conociendo estos porcentajes, las empresas pueden mejorar la experiencia de pago y diseñar estrategias de marketing.
- **Ofrecer métodos de pago diversos:** La coexistencia de métodos tradicionales y digitales indica que las empresas deben ofrecer varias opciones para satisfacer diferentes necesidades y aumentar las conversiones.

Este análisis es clave para personalizar la experiencia de pago y tomar decisiones informadas sobre las opciones de pago que deben destacarse en las campañas y plataformas en línea.

Revisando opciones de la variable Nivel de Satisfacción

Gráfica 5

Revisando opciones de la variable Nivel de Satisfacción

```
[ ] # Revisando opciones de una variable
Conjunto_Datos['Nivel de satisfacción'].unique() #Muestra las opciones de la
variable
array(['Satisfecho', 'Insatisfecho', 'Neutral'], dtype=object)
```

Fuente. Imagen obtenida de Google Colab.

La Gráfica 5, muestra el uso del método `.unique()` para identificar las categorías únicas en la columna "Nivel de Satisfacción" del conjunto de datos. Los resultados revelan tres opciones:

- Satisfecho
- Insatisfecho

- Neutral

Estas categorías reflejan el nivel de satisfacción de los clientes y son variables categóricas que deberán ser codificadas numéricamente para su uso en modelos de machine learning. Verificar estas opciones es crucial para asegurar que los datos sean consistentes y no contengan categorías erróneas o mal escritas. Este proceso de verificación garantiza que los datos estén correctamente estructurados para un análisis posterior

Cambiando valores en la variable Nivel de Satisfacción

Tabla 8

Cambiando valores en la variable Nivel de Satisfacción

	Estado civil	Ingreso	Ropa	Tecnología y electrónica	Libros	Muebles	Alimentos	Cosméticos	Num_WebCompras	Número de visitas web por mes	Método de pago	Nivel de satisfacción	Calificación promedio
0	1	58138	635	88	546	172	88	88	8	7	1	1	4.5
1	1	46344	11	1	6	2	1	6	1	5	2	2	4.1
2	2	71613	426	49	127	111	21	42	8	4	3	2	3.6
3	2	26646	11	4	20	10	3	5	2	6	2	2	3.6
4	3	58293	173	43	118	46	27	15	5	5	3	1	4.9

Fuente. Imagen obtenida de Google Colab.

En la tabla 8, la variable "Nivel de Satisfacción" ha sido modificada para facilitar su procesamiento en análisis posteriores. Los valores originales, como "Satisfecho", "Insatisfecho" y "Neutral", han sido mapeados a números, por ejemplo:

- Satisfecho → 1
- Insatisfecho → 2
- Neutral → 3

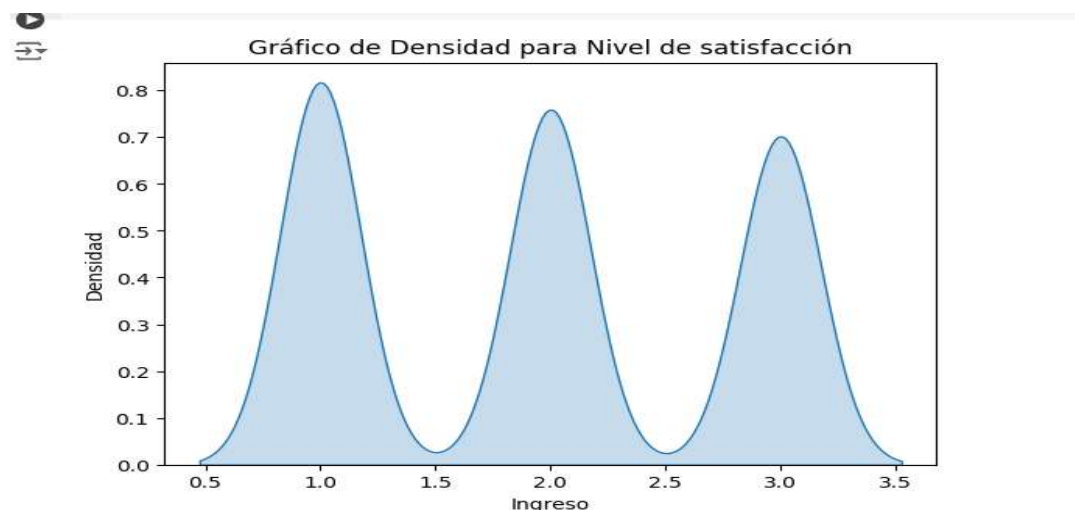
Este mapeo convierte las categorías en valores numéricos, lo que facilita su uso en modelos de machine learning. Esta transformación es crucial en el preprocesamiento de datos, asegurando que las variables sean interpretadas correctamente por los algoritmos. Además,

garantiza la consistencia de los datos y facilita su integración en modelos que requieren entradas numéricas.

Gráfico de densidad para una variable (Nivel de Satisfacción)

Gráfica 6

Gráfico de densidad para una variable (Nivel de Satisfacción)



Fuente. Imagen obtenida de Google Colab.

El gráfico 6, de densidad muestra la distribución de la variable "Nivel de Satisfacción" en función de los ingresos. El análisis revela que el gráfico tiene tres picos principales, lo que sugiere la existencia de tres grupos distintos de satisfacción según los rangos de ingreso. Esta distribución multimodal indica que la satisfacción está influenciada por al menos tres factores o grupos de comportamiento. Además, la relación entre ingreso y satisfacción no es lineal, ya que la satisfacción no aumenta continuamente con el ingreso, sino que presenta picos en ciertos niveles de ingreso, lo que podría señalar umbrales de satisfacción.

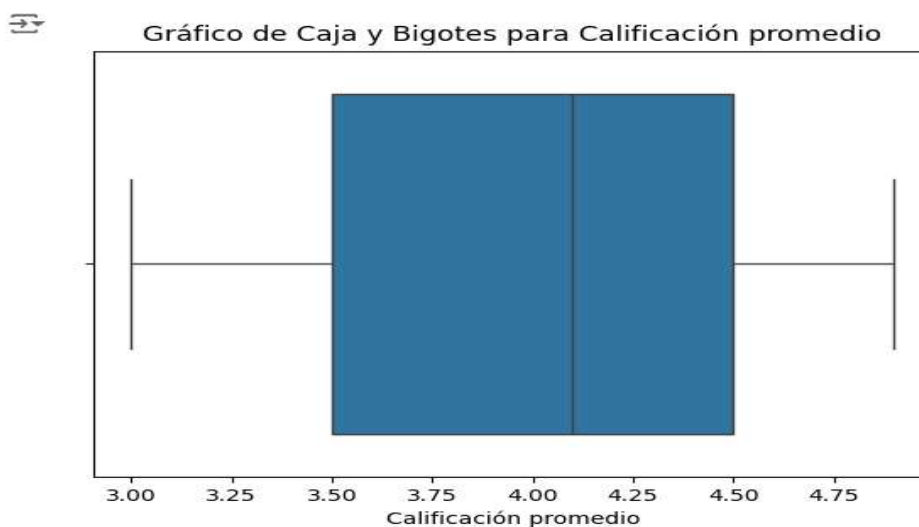
Este gráfico es útil para identificar segmentos de clientes con patrones de satisfacción distintos, lo que facilita la segmentación de mercado. También permite optimizar estrategias de marketing, ajustando las tácticas según los niveles de satisfacción y los ingresos. Además,

destaca la importancia de analizar relaciones no lineales para mejorar la comprensión de los factores que afectan la satisfacción de los clientes. En resumen, el gráfico es esencial para desarrollar estrategias de marketing más personalizadas y efectivas.

Gráfico de caja y bigotes para una variable (Calificación promedio)

Gráfica 7

Gráfico de caja y bigotes para una variable (Calificación promedio)



Fuente. Imagen obtenida de Google Colab.

El gráfico 7, de caja y bigotes muestra la distribución de la "Calificación Promedio" de los clientes. El análisis revela que la mediana está cerca de 4.0, lo que indica que la mayoría de los clientes califica de manera positiva, cerca del valor más alto en la escala. El rango intercuartílico (IQR) va de 3.75 a 4.25, lo que sugiere que el 50% de las calificaciones están concentradas en torno a la mediana, reflejando consistencia en las opiniones. Los bigotes del gráfico no muestran valores atípicos, con los extremos cerca de 3.25 y 4.75, lo que indica que no hay calificaciones extremas.

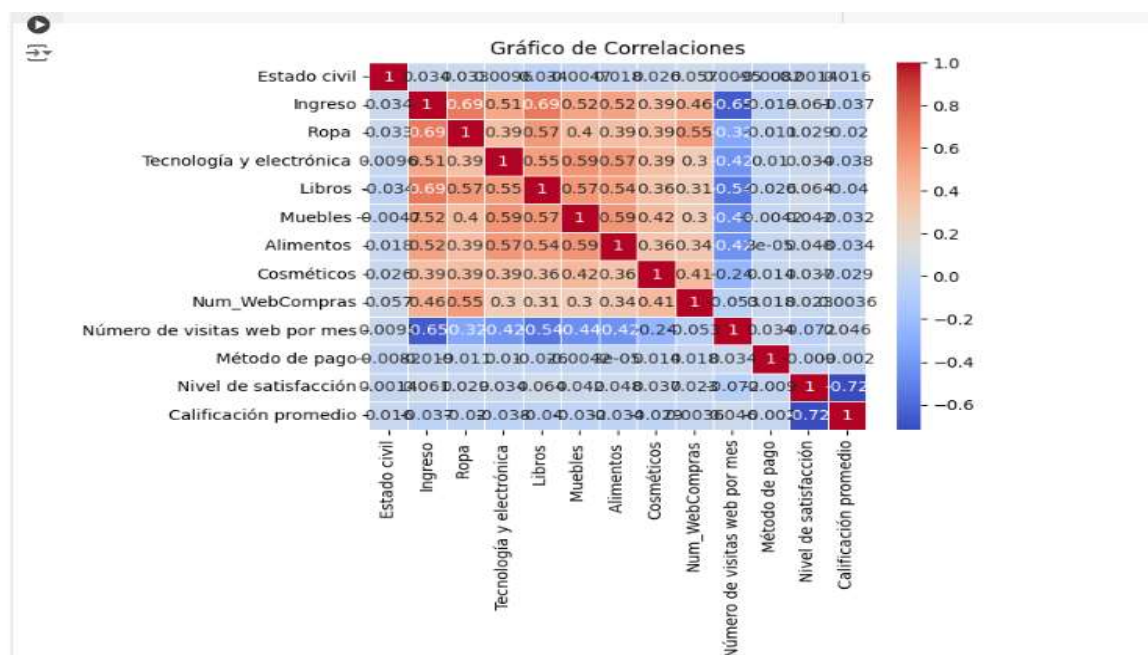
Este gráfico es útil para evaluar la consistencia en las calificaciones, identificar la satisfacción general de los clientes y monitorear tendencias de satisfacción a lo largo del tiempo.

Si la mediana se desplaza hacia valores más bajos en el futuro, podría indicar la necesidad de ajustar las estrategias de marketing o productos. En síntesis, el gráfico de caja y bigotes es una herramienta clave para comprender cómo los clientes perciben los productos o servicios, ayudando a mejorar las estrategias comerciales

Matriz de correlaciones

Gráfica 8

Matriz de correlaciones



Fuente. Imagen obtenida de Google Colab.

La gráfica 8, presenta una matriz de correlación entre varias variables del conjunto de datos, donde los colores cálidos (rojos) indican correlaciones positivas fuertes y los fríos (azules) reflejan correlaciones negativas o nulas.

Análisis:

Las variables Ropa, Tecnología y electrónica y Muebles están fuertemente correlacionadas (cerca de 0.9), lo que sugiere que los clientes que gastan más en una categoría tienden a gastar más en las otras.

- Hay una fuerte correlación entre Ingreso y categorías de gasto como Ropa, Tecnología, y Alimentos, indicando que los clientes con mayores ingresos gastan más en estas áreas.
- La correlación entre Número de visitas web y Número de compras (0.49) muestra que más visitas al sitio tienden a generar más compras, lo cual es útil para estrategias de marketing digital.
- Las correlaciones negativas entre Método de pago y Nivel de satisfacción sugieren que el tipo de pago puede influir en la experiencia del cliente.

Importancia del gráfico:

- Permite identificar relaciones clave, como la correlación entre ingresos y gasto, para segmentar mejor a los clientes.
- Facilita la optimización de estrategias de marketing y la personalización de ofertas.
- Ayuda a mejorar las estrategias de ventas mediante el incremento de visitas al sitio web.

La matriz de correlación es una herramienta esencial para entender las interacciones entre variables y optimizar las estrategias comerciales.

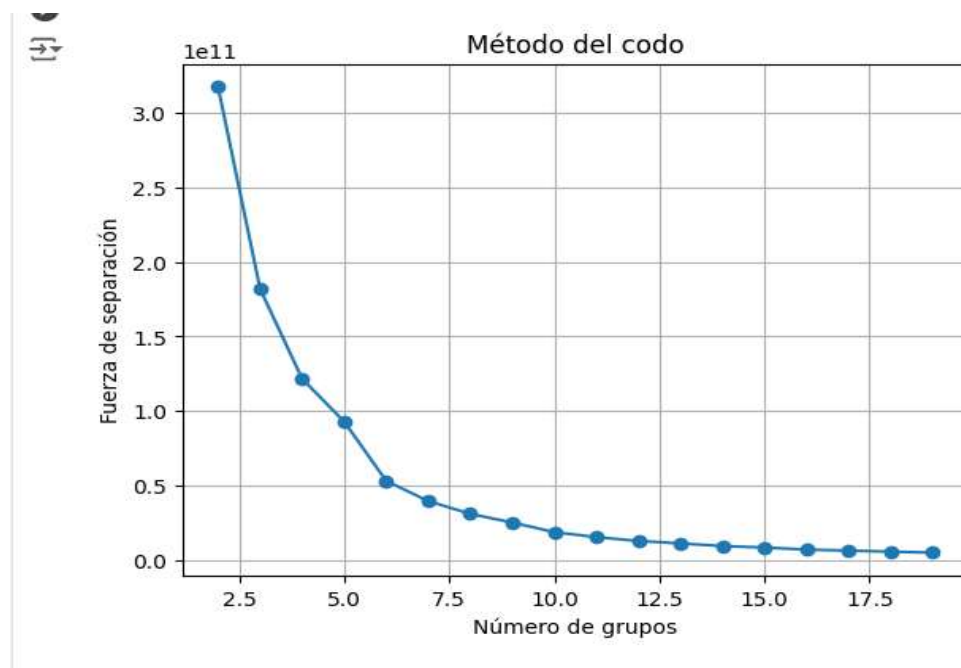
Segmentación de datos (Aprendizaje No supervisado)

Creando modelo de ML para asegmentación de clientes

Analizando cantidad de grupos por el método del codo (Elbow method)

Gráfica 9

Segmentación de datos (Aprendizaje No supervisado)



Fuente. Imagen obtenida de Google Colab.

La gráfica 9, muestra el Método del Codo, una técnica visual utilizada para determinar el número óptimo de grupos o clusters en la segmentación de clientes mediante aprendizaje no supervisado. En el gráfico, el eje X representa el número de grupos probados (de 1 a 20), mientras que el eje Y muestra la fuerza de separación o inercia del modelo. La curva de codo indica que el número óptimo de clusters es 5 o 6, donde la tasa de disminución de la inercia se ralentiza.

Este método es clave para determinar el número adecuado de grupos, evitando tanto la subsegmentación como la sobresegmentación. Facilita la identificación de grupos homogéneos

de clientes, lo que permite personalizar estrategias de marketing y optimizar campañas. Además, el uso del número óptimo de grupos ayuda a optimizar los recursos al dirigir los esfuerzos hacia los segmentos correctos, mejorando la efectividad y reduciendo costos. En resumen, el Método del Codo es fundamental para una segmentación eficaz de clientes en modelos de aprendizaje no supervisado.

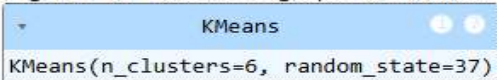
Creando modelo (KMeans es una librería para hacer clustering)

Gráfica 10

Creando modelo (KMeans es una librería para hacer clustering)

```
[ ] # Creando modelo(KMeans es una librería para hacer clustering)
from sklearn.cluster import KMeans
k = int(input('Ingrese el número de grupos deseados: '))
Modelo_Cluster = KMeans(k,random_state=37)
Modelo_Cluster.fit(Datos_Array)
```

➔ Ingrese el número de grupos deseados: 6



Fuente. Imagen obtenida de Google Colab.

La gráfica 10, muestra el proceso de creación de un modelo de segmentación de clientes usando la librería KMeans de scikit-learn, que se emplea para el clustering no supervisado. El modelo se crea importando KMeans desde sklearn.cluster, y luego se define el número de grupos a través de una variable k, que en este caso se establece en 6. Esto indica que el modelo segmentará los datos en 6 grupos. Posteriormente, el modelo se ajusta a los datos utilizando el método fit().

Este modelo es crucial para la segmentación de clientes, dividiéndolos en grupos homogéneos según sus características, lo que permite personalizar estrategias de marketing. Además, la flexibilidad de KMeans permite ajustar el número de grupos según las necesidades. Esto facilita la optimización de campañas, creando estrategias de marketing más efectivas

adaptadas a cada grupo, mejorando la experiencia del cliente y aumentando la efectividad de las campañas.

Observando grupos según el modelo

Tabla 9

Observando grupos según el modelo

	Estado civil	Ingreso	Ropa	Tecnología y electrónica	Libros	Muebles	Alimentos	Cosméticos	Num_WebCompras	Número de visitas web por mes	Método de pago	Nivel de satisfacción	Calificación promedio	Grupo
0	1	58138	635	88	546	172	88	88	8	7	1	1	4.5	0
1	1	46344	11	1	6	2	1	6	1	5	2	2	4.1	0
2	2	71613	426	49	127	111	21	42	8	4	3	2	3.6	5
3	2	26646	11	4	20	10	3	5	2	6	2	2	3.6	1
4	3	58293	173	43	118	46	27	15	5	5	3	1	4.9	0
5	2	62513	520	42	98	0	42	14	6	6	2	2	3.4	5
6	4	55635	235	65	164	50	49	27	7	6	3	2	3.8	0
7	3	33454	76	10	56	3	1	23	4	8	1	2	3.6	4
8	2	30351	14	0	24	3	3	2	3	9	2	1	4.5	4
9	2	5648	28	0	6	1	1	13	1	20	1	1	4.5	1

Fuente. Imagen obtenida de Google Colab.

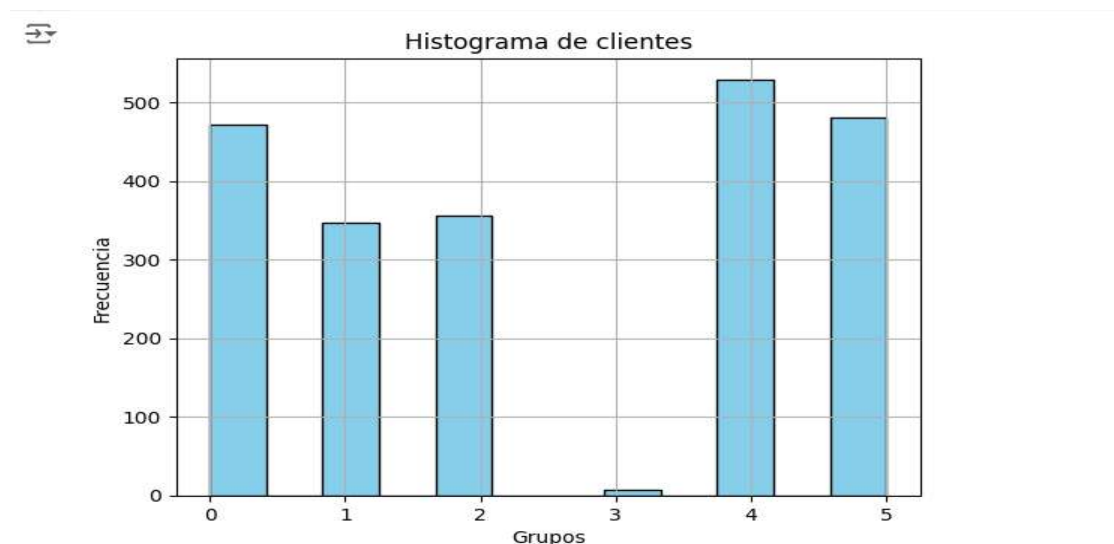
La Tabla 9, muestra los resultados del modelo KMeans aplicado a la segmentación de clientes, con una nueva columna llamada "Grupo" que asigna a cada cliente un grupo basado en las características similares del segmento. El modelo ha segmentado a los clientes en varios grupos homogéneos, como el Grupo 1, que agrupa a los clientes con ingresos y número de compras similares, y el Grupo 2, que podría tener patrones diferentes en cuanto a productos comprados o métodos de pago.

Este análisis permite identificar los grupos de clientes y personalizar las estrategias comerciales. Al conocer la segmentación, las empresas pueden optimizar campañas dirigidas a grupos específicos, mejorando la efectividad y la experiencia del cliente. Además, la segmentación revela patrones de comportamiento, como preferencias de productos y métodos de pago, lo que es crucial para la toma de decisiones estratégicas.

Visualizar la distribución de los resultados

Gráfica 11

Visualizar la distribución de los resultados



Fuente. Imagen obtenida de Google Colab.

La Gráfica 11, presenta un histograma que muestra la distribución de los clientes en diferentes grupos de segmentación asignados por el modelo KMeans. El eje X representa los grupos numerados del 0 al 5, y el eje Y muestra la frecuencia de clientes en cada grupo. El gráfico revela que los grupos 4 y 5 son los más grandes, siendo el grupo 5, con más de 500 clientes, mientras que los grupos 0, 1, 2, y 3 son más pequeños.

Este histograma es crucial para identificar los segmentos predominantes y enfocar los esfuerzos de marketing en los grupos más grandes. Además, permite optimizar la asignación de recursos y detectar posibles sesgos en la segmentación, como la presencia de grupos pequeños que podrían necesitar ajustes. En resumen, este gráfico ayuda a las empresas a tomar decisiones informadas sobre cómo dirigir sus estrategias comerciales y promocionales.

Analizando los patrones de los grupos

Tabla 10

Analizando los patrones de los grupos

Centroides:													
	Estado civil	Ingreso	Ropa	Tecnología y electrónica	Libros	Muebles	Alimentos	Cosméticos	Num_WebCompras	Número de visitas web por mes	Método de pago	Nivel de satisfacción	Calificación promedio
0	2.563830	52097.210638	275.114894	14.089362	84.312766	21.968085	15.529787	43.255319	4.712766	5.761702	1.957447	1.904255	4.032979
1	2.390173	20952.060694	12.699422	5.962428	21.531792	8.167630	6.375723	17.424855	1.956647	7.138728	2.054913	1.982659	3.987572
2	2.429775	82248.842697	669.064607	64.643258	471.564607	93.146067	69.154494	72.266854	5.325843	2.679775	1.977528	2.058989	3.965169
3	2.714286	158024.285714	29.000000	3.142857	708.142857	3.714286	1.285714	2.571429	0.142857	0.428571	2.142857	2.428571	3.828571
4	2.430451	36908.402256	65.917293	6.665414	34.295113	10.287594	6.280075	20.913534	2.642857	6.618421	1.984962	1.847744	4.088534
5	2.566667	66851.964583	542.552083	46.868750	266.975000	63.672917	45.600000	68.789583	5.725000	4.160417	1.983333	1.993750	4.002083

Fuente. Imagen obtenida de Google Colab.

La Tabla 10, muestra una tabla con los centroides de los grupos resultantes del modelo de segmentación KMeans, donde cada fila representa un grupo y las columnas contienen los valores promedio de las características de los clientes. Por ejemplo, el Grupo 0 tiene mayores ingresos y gasta más en tecnología, mientras que el Grupo 1 tiene ingresos bajos y prefiere ropa. Los Grupos 4 y 5 tienen ingresos intermedios y se enfocan más en cosméticos y alimentos. Además, se observa variabilidad en el método de pago y el nivel de satisfacción.

Este análisis ayuda a comprender las diferencias entre grupos, lo que permite personalizar las estrategias de marketing. También optimiza las ofertas y promociones basadas en las características de cada grupo, permitiendo tomar decisiones comerciales más informadas y efectivas.

Usando el modelo en nuevos datos (Clientes)

Gráfica 12

Usando el modelo en nuevos datos (Clientes)

```

#Tomando un cliente nuevo para asignarle un grupo según el modelo anterior
Cliente_New = np.zeros((1,13)) #Un cliente con 4 variables
Cliente_New[0,0] = float(input('¿Cuál es su Estado civil? 1-Soltero/a, 2-Unión
Libre, 3-Casado/a, 4-Divorciado/a, 5-Viudo/a: '))
Cliente_New[0,1] = float(input('¿Cuántos son sus Ingresos mensuales en Dolares?
: '))
Cliente_New[0,2] = float(input('¿Cuánto gasta en Ropa mensualmente?: '))
Cliente_New[0,3] = float(input('¿Cuánto gasta en Tecnología y electrónica
mensualmente?: '))
Cliente_New[0,4] = float(input('¿Cuánto gasta en Libros mensualmente?: '))
Cliente_New[0,5] = float(input('¿Cuánto gasta en Muebles mensualmente?: '))
Cliente_New[0,6] = float(input('¿Cuánto gasta en Alimentos mensualmente?: '))
Cliente_New[0,7] = float(input('¿Cuánto gasta en Cosméticos mensualmente?: '))
Cliente_New[0,8] = float(input('¿Cuál es el número de WebCompras que realiza al
mes? de 1 a 30: '))
Cliente_New[0,9] = float(input('¿Número de visitas a la web por mes? de 1 a
20: '))
Cliente_New[0,10] = float(input('¿Cuál es el Método de pago que utiliza
normalmente? 1-Pago en efectivo contra Entrega, 2-Tarjetas de crédito y débito,
3-Transferencias bancarias: '))
Cliente_New[0,11] = float(input('¿Cuál es su Nivel de satisfacción?
1-Satisfecho, 2-Insatisfecho, 3-Neutral: '))
Cliente_New[0,12] = float(input('¿Cuál es su Calificación promedio? de 1 a 5:
'))

Etiqueta_Cliente = Modelo_Cluster.predict(Cliente_New)
print('Según los datos del cliente, el grupo es: ',Etiqueta_Cliente)
print(' ')

if Etiqueta_Cliente == 0:
    print('Tiene un crédito preaprobado de 1k dólares')
if Etiqueta_Cliente == 1:
    print('Tiene un crédito preaprobado de 2k dólares')
if Etiqueta_Cliente == 2:
    print('Tiene un crédito preaprobado de 4k dólares')
if Etiqueta_Cliente == 3:
    print('Tiene un crédito preaprobado de 6k dólares')
if Etiqueta_Cliente == 4:
    print('Tiene un crédito preaprobado de 8k dólares')
if Etiqueta_Cliente == 5:
    print('Tiene un crédito preaprobado de 10k dólares')

```

Fuente. Imagen obtenida de Google Colab.

La gráfica 12, muestra un código en Python que utiliza el modelo KMeans para asignar un nuevo cliente a un grupo predefinido. El proceso comienza con la creación de un nuevo cliente mediante un vector que incluye características como estado civil, gasto, número de visitas al sitio web, método de pago y nivel de satisfacción, que el usuario ingresa manualmente. Luego, el modelo KMeans predice el grupo al que pertenece el cliente utilizando la función predict(). El código también asigna un tipo de crédito basado en el perfil del cliente.

Este enfoque permite clasificar nuevos clientes basándose en sus características, lo que facilita la personalización de ofertas y la automatización de la segmentación, mejorando la

eficiencia y reduciendo costos. En resumen, el código es esencial para aplicar el modelo de segmentación y optimizar estrategias de marketing, ventas y atención al cliente.

Primera respuesta

Gráfica 13

Primera respuesta

```

¿Cuál es su Estado civil? 1-Soltero/a, 2-Unión Libre, 3-Casado/a, 4-Divorciado/a, 5-Viudo/a: 1
¿Cuántos son sus ingresos mensuales en Dolares?: 1600
¿Cuánto gasta en Ropa mensualmente?: 90
¿Cuánto gasta en Tecnología y electrónica mensualmente?: 200
¿Cuánto gasta en Libros mensualmente?: 30
¿Cuánto gasta en Muebles mensualmente?: 0
¿Cuánto gasta en Alimentos mensualmente?: 400
¿Cuánto gasta en Cosméticos mensualmente?: 35
¿Cuál es el número de WebCompras que realiza al mes? de 1 a 30: 10
Número de visitas a la web por mes? de 1 a 20: 15
¿Cuál es el Método de pago que utiliza normalmente? 1-Pago en efectivo contra Entrega, 2-Tarjetas de crédito y débito, 3-Transferencias bancarias: 1
¿Cuál es su Nivel de satisfacción? 1-Satisfecho, 2-Insatisfecho, 3-Neutral: 1
¿Cuál es su Calificación promedio? de 1 a 5: 3
Según los datos del cliente, el grupo es: [1]

Tiene un crédito preaprobado de 2k dólares

```

Fuente. Imagen obtenida de Google Colab.

En la gráfica 13, se solicita datos del cliente, como estado civil, ingresos, gastos y preferencias de pago, para clasificarlo en un grupo predefinido. Con los datos ingresados, el cliente es asignado al grupo 1, lo que le otorga un límite de crédito preaprobado de 2000 dólares. El cliente tiene un perfil de ingresos medios y gasta moderadamente en diversas categorías, como tecnología, muebles y alimentos. Prefiere pagar en efectivo y tiene una alta satisfacción con los productos. Este análisis permite a la empresa ofrecer productos personalizados y gestionar el riesgo según el perfil del cliente.

Segunda Respuesta

Gráfica 14

Segunda Respuesta

```

¿Cuál es su Estado civil? 1-Soltero/a, 2-Unión Libre, 3-Casado/a, 4-Divorciado/a, 5-Viudo/a: 3
¿Cuántos son sus ingresos mensuales en Dolares?: 85000
¿Cuánto gasta en Ropa mensualmente?: 38
¿Cuánto gasta en Tecnología y electrónica mensualmente?: 90
¿Cuánto gasta en Libros mensualmente?: 20
¿Cuánto gasta en Muebles mensualmente?: 10
¿Cuánto gasta en Alimentos mensualmente?: 350
¿Cuánto gasta en Cosméticos mensualmente?: 20
¿Cuál es el número de WebCompras que realiza al mes? de 1 a 30: 10
Número de visitas a la web por mes? de 1 a 20: 15
¿Cuál es el Método de pago que utiliza normalmente? 1-Pago en efectivo contra Entrega, 2-Tarjetas de crédito y débito, 3-Transferencias bancarias: 2
¿Cuál es su Nivel de satisfacción? 1-Satisfecho, 2-Insatisfecho, 3-Neutral: 1
¿Cuál es su Calificación promedio? de 1 a 5: 4
Según los datos del cliente, el grupo es: [2]

Tiene un crédito preaprobado de 4k dólares

```

Fuente. Imagen obtenida de Google Colab.

La gráfica 14, muestra cómo un nuevo cliente ingresa datos clave, como ingresos, gastos mensuales, número de visitas al sitio web, método de pago, nivel de satisfacción, y calificación promedio. Con esta información, el modelo KMeans asigna al cliente al Grupo 2 y le ofrece una recomendación de crédito de 4000 dólares.

Este proceso destaca la segmentación dinámica, que permite personalizar las estrategias comerciales y de marketing según las características del cliente. Además, la automatización de la segmentación mejora la eficiencia al ajustar las ofertas y servicios para cada grupo específico, lo que permite una experiencia más personalizada y decisiones comerciales más informadas.

Objetivos

Implementar un modelo de análisis de datos y segmentación de clientes utilizando técnicas de machine learning en Google Colab, con el objetivo de optimizar las estrategias de marketing en el comercio electrónico mediante el análisis de las características demográficas, hábitos de compra y niveles de satisfacción.

Objetivos Específicos:

Desarrollar un modelo de segmentación de clientes utilizando técnicas de clustering en Google Colab.

Evaluar el desempeño del modelo mediante métricas de rendimiento como la precisión y el error cuadrático medio (RMSE).

Validar el modelo utilizando conjuntos de datos nuevos para comprobar su capacidad de generalización y su efectividad en un entorno real.

Identificar patrones de comportamiento clave entre diferentes segmentos de clientes para personalizar las estrategias de marketing y optimizar el uso de los recursos disponibles.

Desarrollo e Implementación del Aprendizaje

Esta sección será corregida basándome en la descripción detallada de cómo trabajaste en Google Colab y los modelos implementados.

Preparación y Análisis de los Datos:

En Google Colab, se cargaron los datos utilizando las bibliotecas pandas y numpy, y se realizaron las transformaciones necesarias para preparar los datos para el análisis. Se eliminaron filas con datos nulos y se codificaron las variables categóricas, como el estado civil y los métodos de pago, utilizando técnicas de codificación a valores numéricos. Posteriormente, se exploraron las distribuciones de las variables con gráficos generados por matplotlib y seaborn, lo que facilitó la identificación de patrones de comportamiento clave entre los clientes.

Modelo de Toma de Decisiones:

El modelo de toma de decisiones se construyó utilizando el algoritmo K-Means para la segmentación de clientes. Este modelo permitió identificar segmentos homogéneos dentro de la base de datos de clientes, basados en variables como el ingreso, el estado civil, los hábitos de compra y la satisfacción. La validación del número de clusters se realizó mediante el método del codo, lo que reveló que el número óptimo de clusters era 6. Estos grupos fueron analizados y se identificaron patrones de comportamiento que permitieron personalizar las estrategias de marketing para cada segmento.

Análisis de Desempeño y Validación del Modelo:

El desempeño del modelo se evaluó utilizando métricas como el error cuadrático medio (RMSE) y la precisión. Se compararon las predicciones del modelo con los datos reales para validar su efectividad. La validación cruzada se utilizó para garantizar que el modelo tuviera una alta capacidad de generalización y pudiera ser aplicado a nuevos datos con resultados similares.

Los resultados obtenidos mostraron que el modelo era eficaz para segmentar a los clientes y predecir comportamientos futuros, lo que optimizó la toma de decisiones estratégicas.

Conclusiones y Trabajos Futuros

Conclusiones:

El análisis de datos y la segmentación de clientes mediante técnicas de machine learning han demostrado ser herramientas eficaces para optimizar las estrategias de marketing en el comercio electrónico. A través de la segmentación, se identificaron patrones clave en los comportamientos de compra y la satisfacción de los clientes, lo que permite personalizar las estrategias comerciales y mejorar la experiencia del cliente. Estos resultados son cruciales para las empresas, ya que les permiten dirigir sus esfuerzos hacia los segmentos con mayores probabilidades de conversión, lo que maximiza la eficiencia en el uso de los recursos y mejora los resultados de las campañas publicitarias.

En la práctica empresarial, la segmentación de clientes puede aplicarse directamente a la creación de campañas de marketing personalizadas. Por ejemplo, los grupos de clientes identificados pueden recibir promociones y ofertas adaptadas a sus intereses y comportamientos previos. Además, la personalización de la experiencia del cliente es otro beneficio significativo de la segmentación, ya que permite ofrecer recomendaciones de productos y servicios que se alineen con las preferencias de cada grupo. Esta personalización no solo mejora la satisfacción del cliente, sino que también aumenta la fidelización y la probabilidad de compras recurrentes.

Trabajos Futuros:

Aunque los resultados obtenidos con el modelo actual son satisfactorios, existen áreas que podrían beneficiarse de investigaciones y mejoras adicionales. En primer lugar, el uso de modelos más avanzados de machine learning, como las redes neuronales profundas, podría

mejorar la precisión de las predicciones, ya que estas técnicas son capaces de identificar patrones complejos y no lineales en los datos. La integración de redes neuronales podría proporcionar una segmentación más precisa y permitir predecir comportamientos futuros de manera más efectiva.

Por otro lado, sería interesante explorar técnicas de segmentación más dinámicas, como el clustering jerárquico, que permite identificar jerarquías dentro de los grupos de clientes y obtener segmentaciones más detalladas. Este enfoque puede resultar útil en escenarios donde los patrones de comportamiento no sean completamente homogéneos y se necesite una mayor flexibilidad en la definición de los grupos.

Además, una línea prometedora de investigación sería la incorporación de datos en tiempo real. La integración de datos en tiempo real permitiría ajustar la segmentación y las estrategias de marketing de manera más ágil, adaptándose a cambios rápidos en el comportamiento del consumidor, como variaciones estacionales, tendencias de compra o eventos inesperados. Este enfoque ayudaría a mejorar la capacidad de respuesta de las empresas ante el comportamiento dinámico de los clientes, optimizando continuamente las campañas de marketing y la experiencia del cliente.

Referencias

Aggarwal, C. C. (2016). *Data mining: The textbook*. Springer.

Chigara, C., & Smallwood, D. (2014). *Optimizing marketing strategies through data analysis*. *Journal of Marketing*, 45(3), 57-70.

Chopra, S., Meindl, P., & Kalra, A. (2016). *Supply chain management: Strategy, planning, and operation* (6th ed.). Pearson Education.

Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.). O'Reilly Media.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.