

1. Desarrollar un modelo de clasificación utilizando Máquinas de Soporte Vectorial (SVM), para predecir si un cliente comprará o no un producto en un sitio web de comercio electrónico, basándose en su comportamiento de navegación. Los datos específicos para analizar, incluyen el número de accesos a la página del producto, el tiempo total pasado en la página y si el producto fue agregado al carrito.

Acceso a la pagina	Tiempo en la pagina	Agregacion al carrito	Compra del producto
9	23.438.745.967.916.500	1	1
12	354.935.869.970.497	1	0

Preprocesamiento y Análisis: Normalizar y ajustar los datos según sea necesario para el modelado y realizar un análisis exploratorio para entender las características y la distribución de los datos.

Modelado con SVM: Utilizar el algoritmo SVM para entrenar un modelo de clasificación que pueda predecir la probabilidad de que un cliente realice una compra en función de su comportamiento en el sitio web.

Evaluación del Modelo: Calcular métricas de rendimiento como precisión, recall, F1-score y mostrar la matriz de confusión para evaluar la efectividad del modelo en la clasificación de los clientes que realizarán una compra.

Visualización: Generar una visualización, preferiblemente en 3D, del hiperplano de decisión de la SVM que muestre cómo el modelo clasifica los datos entre las posibles acciones de compra y no compra.

¿Cómo afecta la selección de kernel al rendimiento del modelo SVM?: Investigar qué kernel (lineal, polinomial, RBF, sigmoidal) es más adecuado para los datos del e-commerce y cómo diferentes kernels influyen en la precisión, el recall y el F1-score del modelo.

Realice una presentación en Power Bi para mostrar a los socios de la empresa el comportamiento modelo análisis



ecommerce_data.csv

v

2. Archivo

Una empresa de telecomunicaciones está experimentando una tasa de abandono de clientes (churn) más alta de lo deseado. La empresa quiere utilizar análisis de datos y aprendizaje automático para identificar los factores que contribuyen al churn y predecir qué clientes están en riesgo de abandonar. Esto permitirá a la empresa tomar medidas proactivas para retener a esos clientes. utilice un modelo de red neuronal para identificar a los clientes en riesgo de churn. Las variables para tener en cuenta son:

Duración del contrato: El tiempo que el cliente ha estado con la empresa.

Uso mensual: Cantidad de uso del servicio por el cliente en un mes típico.

Facturación mensual: Cuánto paga el cliente por los servicios cada mes.

Satisfacción del cliente: Puntuación de satisfacción basada en encuestas recientes.

Soporte técnico: Número de veces que el cliente ha contactado al soporte técnico.

Variable Objetivo (Etiqueta): Churn: Si el cliente dejó la empresa (1) o no (0).

Duracion del contrato	Uso mensual	Facturacion mensual	Satisfaccion del cliente	Soporte tecnico	Churn
0.7740740509661974	20.437.368.924.193.100	25.101.477.740.733.800	0.145657054814895	0.8214601551837409	1
-0.143013884719579	0.9117612804142561	27.613.002.196.521.100	0.26830894876243916	0.23743457396898515	0
-16.385.995.267.488.500	18.728.544.531.505.900	-10.404.934.810.462.300	0.54743920361829	-1.521.008.873.615.400	0

Preprocesamiento de Datos: Preparar los datos para el modelado, incluyendo la normalización o estandarización si es necesario.

Modelado de Red Neuronal: Utilizar MLPClassifier de scikit-learn para crear una red neuronal que pueda predecir la probabilidad de churn de los clientes.

Evaluación del Modelo: Determinar la efectividad del modelo utilizando métricas como precisión, recall, F1-score y la matriz de confusión.

Realice una presentación en Power Bi para mostrar a los socios de la empresa el comportamiento modelo análisis

3. Una empresa que vende productos por internet desea mejorar sus estrategias de marketing y ventas personalizando sus campañas según los diferentes tipos de comportamientos de compra de sus clientes. La empresa quiere utilizar la técnica de clustering, específicamente K-Means, para segmentar a sus clientes en grupos con características y comportamientos similares. Las variables para tener en cuenta son:

Total, de compras: Cantidad total de productos comprados por el cliente.

Frecuencia de compras: Número de visitas que resultaron en una compra.

Categoría favorita: Categoría de producto más comprada (se puede representar con un valor numérico).

Valor promedio de compra: Promedio de gasto en cada compra.

Tiempo en el sitio: Tiempo total promedio que el cliente pasa en el sitio por visita.

Descripción de Resultados

Analizar los clusters resultantes para comprender las características distintivas de cada grupo. Por ejemplo, examinar el centroide de cada cluster puede revelar las tendencias comunes en cada segmento.

Determinar las diferencias en Total de compras, Frecuencia de compras, Categoría favorita, Valor promedio de compra, y Tiempo en el sitio entre los clusters.

Identificar patrones como qué cluster representa a clientes de alto valor, clientes frecuentes, clientes nuevos o casuales, etc.

Describe algunas métricas y explique la eficiencia del modelo

Realice una presentación en Power Bi para mostrar a los socios de la empresa el comportamiento modelo análisis

Utilizar esta información para personalizar las estrategias de marketing, como campañas de correo electrónico, ofertas especiales, y recomendaciones de productos basadas en la categoría favorita.

Este enfoque permitirá a la empresa de comercio electrónico comprender mejor a sus clientes y adaptar sus estrategias de marketing y ventas para aumentar la efectividad y mejorar la satisfacción del cliente.