

---

Nombre del proyecto
Clasificación técnica mediante SVM y Árboles de Decisión.

Profesor	Gladys María Villegas Rugel
Materia	Aprendizaje Automático
Grupo	
Autores	Jorge Murillo
Fecha	04/02/2026

Resumen del Proyecto
Este informe presenta el desarrollo y evaluación de modelos de clasificación supervisada aplicados a un conjunto de datos de marketing digital, con el objetivo de predecir la decisión de compra de usuarios en una red social.

Actividad	Fecha
Creación del informe	04/02/2026

# Informe Académico: Predicción de Compra en Redes Sociales

## 1. Introducción

En el contexto actual del marketing digital, las empresas generan y recopilan grandes volúmenes de datos relacionados con el comportamiento de los usuarios en plataformas digitales. Analizar estos datos de forma eficiente se ha convertido en un factor clave para mejorar la efectividad de las campañas publicitarias y optimizar el uso de recursos.

La predicción del comportamiento del consumidor, particularmente la decisión de compra representa un problema relevante que puede abordarse mediante técnicas de Machine Learning. En este sentido, los modelos de clasificación supervisada permiten identificar patrones en los datos históricos y realizar predicciones sobre nuevos usuarios, apoyando así la toma de decisiones basada en datos.

## 2. Dominio del Problema

El problema analizado se sitúa en el dominio del **marketing digital y la analítica de clientes**, específicamente en el estudio de campañas publicitarias en redes sociales. El conjunto de datos utilizado contiene información demográfica básica de usuarios, tales como edad, género y salario estimado, junto con una variable objetivo que indica si el usuario realizó o no una compra.

El objetivo es determinar la probabilidad de compra de un usuario a partir de dichas características, abordando el problema como una tarea de **clasificación binaria**. Este tipo de análisis resulta fundamental para la segmentación de clientes, la personalización de anuncios y la mejora del retorno de inversión en estrategias publicitarias.

## 3. Metodología

La metodología seguida en este estudio se estructuró en varias etapas. En primer lugar, se realizó un análisis exploratorio de los datos, con el fin de comprender la distribución de las variables, identificar posibles relaciones entre ellas y detectar valores irrelevantes para el modelo.

Posteriormente, se llevó a cabo el preprocesamiento de los datos, que incluyó la eliminación de variables no informativas, la codificación de variables categóricas y el escalado de las características numéricas para garantizar un entrenamiento adecuado de los modelos.

Una vez preparados los datos, se entrenaron tres modelos de clasificación supervisada: Regresión Logística, Support Vector Machine (SVM) y Árbol de Decisión. Finalmente, los

modelos fueron evaluados mediante métricas de desempeño como la matriz de confusión, la precisión (accuracy) y la validación cruzada, permitiendo comparar su rendimiento de forma objetiva.

## **4. Resultados**

Los resultados obtenidos muestran diferencias claras en el desempeño de los modelos evaluados. La Regresión Logística presentó un rendimiento aceptable, aunque limitado por su naturaleza lineal. El Árbol de Decisión logró una buena capacidad de clasificación, pero mostró una mayor variabilidad en sus resultados.

El modelo Support Vector Machine (SVM) destacó como el de mejor desempeño general, alcanzando mayores niveles de precisión y una mayor estabilidad en la validación cruzada. Las representaciones gráficas, como las matrices de confusión y las fronteras de decisión, evidencian una mejor separación entre las clases en comparación con los otros modelos.

## **5. Conclusiones**

A partir del análisis realizado, se concluye que las variables edad y salario estimado son factores determinantes en la decisión de compra de los usuarios. Asimismo, se observa que el uso de modelos no lineales, como SVM, mejora significativamente la capacidad predictiva del sistema frente a enfoques más simples.

El estudio demuestra la utilidad de las técnicas de Machine Learning para apoyar la toma de decisiones en marketing digital y resalta la importancia de un adecuado preprocesamiento de los datos y una correcta evaluación de los modelos.

## **6. Anexos**

El anexo incluye las gráficas generadas durante el desarrollo del proyecto en el entorno Jupyter Notebook. Entre ellas se encuentran visualizaciones del análisis exploratorio de datos, matrices de confusión, resultados de la validación cruzada y fronteras de decisión de los modelos entrenados. Estas representaciones gráficas respaldan visualmente los resultados obtenidos y facilitan la interpretación del desempeño de cada modelo.















