

Técnicas de Aprendizaje de Máquina

Proyecto 2

Pontificia Universidad Javeriana
Cristian Javier Diaz Alvarez

16 de marzo de 2025

La segmentación de clientes es una técnica que permite dividir un conjunto de clientes en grupos o segmentos con características similares. Esto facilita la creación de estrategias personalizadas, como campañas de marketing o mejoras en productos, al entender mejor el comportamiento y las necesidades de cada grupo. A través de métodos de análisis de datos como K-Means, es posible identificar patrones en los datos sin etiquetas, permitiendo a las empresas tomar decisiones más informadas y efectivas.

Descripción

El objetivo de este taller es que los estudiantes apliquen técnicas de clustering no supervisado para segmentar clientes, y luego utilicen esos segmentos como etiquetas para entrenar un modelo de árbol de decisión. A través del análisis de un conjunto de datos entregado (en formato CSV), los estudiantes deben identificar patrones en los clientes utilizando K-Means, y posteriormente, entrenar un modelo supervisado para predecir a qué segmento pertenece un cliente nuevo basado en sus características.

Es fundamental que los estudiantes no solo implementen los modelos de manera técnica, sino que también realicen un análisis exhaustivo en cada etapa del proceso. Se espera que, más allá de los algoritmos, puedan describir y justificar las decisiones tomadas durante el preprocesamiento, el análisis exploratorio y la interpretación de los resultados obtenidos.

Específicamente, el estudiante enfocará su esfuerzo en:

- Análisis Exploratorio y Preprocesamiento de Datos
- Aplicación de K-Means
- Entrenamiento y Evaluación del Árbol de Decisión
- Interpretación y Conclusiones
- Interpretación de Resultados
- Presentación y Estructura del Notebook

1. Análisis Exploratorio y Preprocesamiento de Datos (25 %)

Claridad y profundidad del análisis inicial, manejo de las variables y transformación adecuada de los datos.

- Breve descripción del problema a resolver y el objetivo del taller.
- Presentación del dataset.
- Visualización y análisis de las características del dataset.
- Identificación de posibles valores atípicos o inconsistencias

- Transformación de variables numéricas (normalización, estandarización).
- Codificación de variables categóricas.
- Justificación de cada paso en el preprocesamiento.

2. Aplicación de K-Means (25 %)

Uso correcto del método del codo para determinar el número óptimo de clusters y análisis exhaustivo de los grupos encontrados.

- Determinación del número óptimo de clusters utilizando el método del codo.
- Aplicación de K-Means al dataset.
- Descripción detallada de cada uno de los grupos encontrados (segmentación de clientes).

3. Entrenamiento y Evaluación del Árbol de Decisión (20 %)

Calidad del modelo, interpretación de las variables importantes y justificación de los resultados.

- Entrenamiento del modelo de árbol de decisión utilizando los clusters como etiquetas..
- Análisis de las variables más importantes en la toma de decisiones del árbol.
- Evaluación del modelo en conjunto de validación.
- Análisis de las métricas de desempeño.
- Reflexión sobre el uso del modelo y sus posibles aplicaciones en un contexto empresarial.

4. Interpretación y Conclusiones (20 %)

Descripción de los hallazgos, su relevancia y las conclusiones extraídas.

- Resumen del análisis y hallazgos.
- Reflexión sobre el proceso de segmentación y predicción.

5. Presentación y Estructura del Notebook (10 %)

Claridad, estructura y organización del notebook entregado.