

Análisis de Resultados

Proyecto: Clasificación del Índice de Entorno de Calidad de Vida

Objetivos del Proyecto

El proyecto tiene como finalidad predecir el "Índice de Entorno" utilizando las siguientes variables:

- Espacios verdes
- Transporte público
- Industrias
- Existencia de cavas
- Nivel de riesgo de inundación

Los interrogantes a responder son:

1. ¿Con qué precisión puede un modelo de aprendizaje automático clasificar el Índice de Entorno de Calidad de Vida utilizando las variables seleccionadas?
2. ¿Cuáles variables tienen el impacto más significativo en la clasificación?
3. ¿Puede el modelo identificar áreas urbanas específicas que necesitan mejoras para aumentar la calidad de vida?

Análisis de las Variables

Del análisis realizado, se observó que la variable con mayor peso en todos los modelos es el índice de "Industrias", mientras que la variable con menor incidencia es la de "Cavas". Esto responde al segundo interrogante del trabajo.

Desarrollo del Análisis

Se realizaron tres modelos de machine learning para abordar el problema:

- Random Forest
- Regresión Logística
- Árbol de Decisión

Cada modelo generó sus resultados, los cuales fueron comparados al final. Se detectó una interrogante adicional debido a que los tres modelos predecían de manera idéntica la etiqueta "Muy Bajo", que tenía pocos registros en el dataset original pero fue notablemente incrementada tras el balanceo de datos.

Metodología

Se evaluaron los modelos con y sin balanceo de datos para verificar si el balanceo alteraba en alguna medida el resultado final. Los resultados mostraron diferencias marcadas en la precisión del modelo, a pesar de que no fue producto de balanceo de los

datos sino más bien de la poca cantidad de datos que quedaron disponibles para entrenar el modelo en el segundo caso:

Comparación de Performance de los Modelos

Precisión de los Modelos

Modelo	Precisión (Datos Originales)	Precisión (Datos Sintéticos)
Random Forest	79.07%	97.97%
Árbol de Decisión	74.42%	97.22%
Regresión Logística	79.07%	87.55%

Las matrices de confusión ratifican estos datos.

Conclusión

Cualquiera de los tres modelos elegidos puede clasificar el “Índice de Entorno” en función de las variables propuestas con niveles de exactitud aceptables. El modelo Random Forest demostró ser el más preciso independientemente del conjunto de datos utilizado.

Identificación de Áreas Urbanas para Mejoras

Los modelos de Random Forest y Árbol de Decisión, debido a su alta precisión con datos balanceados, pueden ser utilizados para identificar áreas urbanas específicas que necesitan mejoras para aumentar la calidad de vida.

Conclusión Final

El modelo de Random Forest resultó ser el más preciso y robusto para clasificar el Índice de Entorno de Calidad de Vida. Este modelo puede ser utilizado para identificar áreas urbanas específicas que necesitan mejoras, proporcionando una herramienta valiosa para la planificación urbana y la toma de decisiones. La variable más influyente en todos los modelos fue el índice de “Industrias”, mientras que la menos influyente fue la de “Cavas”.

Notas:

1. Los resultados obtenidos deben ser interpretados considerando el impacto del balanceo de datos, especialmente en etiquetas con menor cantidad de registros originales.
2. La visualización de la importancia de las características y las matrices de confusión proporcionan información adicional para entender mejor el comportamiento de los modelos.

