

ACTIVIDAD 3. MÉTODOS DE APRENDIZAJE SUPERVISADO

INTELIGENCIA ARTIFICIAL

ISABEL DELGADO CORREAL

ISMAEL CARVAJAL GONZÁLEZ

LIZETH PAOLA BUITRAGO QUINTERO

UNIVERSIDAD IBEROAMERICANA

FACULTAD INGENIERÍA EN CIENCIA DE DATOS

DIANA CAROLINA CANDIA

11 DE JUNIO DE 2023

INTRODUCCIÓN:

Medellín, una metrópolis en constante evolución, ha implementado exitosamente un sistema de transporte masivo eficiente y sostenible por medio del Metro de Medellín. Esta infraestructura de transporte ha revolucionado la vida cotidiana de sus habitantes al mejorar la conectividad entre distintas zonas urbanas.

En este contexto, los métodos de aprendizaje supervisado tienen un gran potencial para optimizar aún más el sistema de transporte masivo, mediante el análisis de datos y la generación de soluciones innovadoras basadas en inteligencia artificial. Estos métodos permiten entrenar modelos predictivos que puedan identificar patrones y relaciones entre las características de los datos y los resultados esperados, lo que permite predecir con precisión el comportamiento del sistema y desarrollar estrategias efectivas para mejorar la movilidad urbana.

Este informe examina cómo aplicar los métodos de aprendizaje supervisado en el contexto del Metro de Medellín.

El presente trabajo demuestra el valor de incorporar la Ciencia de Datos y la Inteligencia Artificial en la optimización de la movilidad urbana, impulsando la creación de soluciones innovadoras que mejoren la vida cotidiana de los ciudadanos y fomenten un transporte más eficiente y sostenible tanto en Medellín como en otras áreas urbanas a nivel mundial.



EXPLICACIÓN DEL PROYECTO:

El algoritmo de aprendizaje supervisado en este ejemplo es la regresión lineal, mientras que el objetivo es predecir un valor numérico continuo basado en una serie de características. La regresión lineal es una técnica que se utiliza cuando se sospecha que

existe una relación lineal entre las variables independientes (características) y la variable dependiente (valor objetivo).

En este caso, se le proporciona un conjunto de datos con coordenadas X (longitudes) e Y (latitudes). El objetivo es construir un modelo que pueda predecir la latitud (variable dependiente) a partir de la longitud (variable independiente) utilizando la regresión lineal.

Después de cargar los datos y dividirlos en conjuntos de entrenamiento y prueba, se crea y entrena el modelo de regresión lineal utilizando los datos de entrenamiento. Luego, se realizan las predicciones en el conjunto de prueba y se evalúa el desempeño del modelo utilizando distintas métricas (MSE, MAE, R^2 Score).

También se implementa el algoritmo k-Vecinos más Cercanos (k-NN) es un método de aprendizaje supervisado, que se puede aplicar tanto a problemas de clasificación como de regresión. La idea principal de k-NN es predecir etiquetas de clase (en clasificación) o valores numéricos (en regresión) para nuevas instancias basándose en los ejemplos más cercanos en el conjunto de datos de entrenamiento.

En este caso, se aplica el algoritmo k-NN para la regresión utilizando `KNeighborsRegressor` de `scikit-learn`. El objetivo es predecir la latitud (variable dependiente) a partir de la longitud (variable independiente) utilizando k-NN.

Finalmente, para visualizar los resultados del modelo y su ajuste a los datos, se crean gráficos de dispersión utilizando `matplotlib`. Se muestran dos gráficos: uno para el modelo de regresión lineal y otro para el algoritmo de k-Vecinos más Cercanos (k-NN) que también se tomó en cuenta en este ejemplo. Estos gráficos permiten comparar fácilmente el desempeño de los dos modelos y elegir el que mejor se ajuste a sus necesidades.

CONCLUSIÓN:

La aplicación de algoritmos de aprendizaje supervisado, como la regresión lineal y el algoritmo k-NN, en el contexto del Metro de Medellín demuestra un gran potencial en la optimización del sistema de transporte masivo. Estos métodos permiten predecir con precisión valores numéricos continuos, como las latitudes a partir de las longitudes, ayudando a identificar patrones y relaciones que puedan mejorar la movilidad urbana.

A través del análisis de los gráficos de dispersión generados, se puede observar y comparar el desempeño de los dos modelos implementados en este estudio. Esto permite evaluar el ajuste de cada modelo a los datos y, dependiendo de los resultados, tomar

decisiones informadas sobre qué algoritmo se ajusta mejor a las necesidades específicas del proyecto.

En general, la incorporación de la Ciencia de Datos y la Inteligencia Artificial en la optimización de la movilidad urbana tiene el potencial de impulsar la creación de soluciones innovadoras que mejoren la vida cotidiana de los ciudadanos y fomenten un transporte más eficiente y sostenible, no solo en Medellín, sino también en áreas urbanas a nivel mundial.

LINK DEL REPOSITORIO: <https://github.com/isabeldc13008/IA/tree/Actividad-3>

LINK DEL VIDEO: https://laiberocol-my.sharepoint.comhttps://laiberocol-my.sharepoint.com/:v:/g/personal/idelgad6_iberco/ERm8OMORJsJItvYzLITZMfYBTy2ckT65SO1PBdSSBkGYKw/v:/g/personal/idelgad6_iberco/ERm8OMORJsJItvYzLITZMfYBTy2ckT65SO1PBdSSBkGYKw

BIBLIOGRAFÍA:

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Los Elementos del Aprendizaje Estadístico: Minería de Datos, Inferencia y Predicción. Springer Science & Business Media.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Introducción al Aprendizaje Estadístico: con Aplicaciones en R. Springer Science & Business Media.

McKinney, W. (2017). Python para Análisis de Datos: Manipulación de Datos con Pandas, NumPy e IPython. O'Reilly Media.