**Movilidad sostenible**

**Taller 1**

Nos enfrentamos a una serie de fenómenos que están afectando al medio ambiente, manifestándose en el calentamiento global, la contaminación del aire y la pérdida de biodiversidad, entre otros. Establecer políticas que aborden sus causas, junto con el desarrollo de planes de gestión de riesgos que mitiguen sus efectos y permitan una respuesta oportuna, representa uno de los mayores desafíos que tenemos como sociedad.

Uno de estos problemas está relacionado con el cambio climático, el cual abarca tanto el calentamiento global, generado por la emisión de gases de efecto invernadero, como el incremento de desastres ambientales provocados por este calentamiento. En particular, el dióxido de carbono (CO2), aunque esencial para la vida, cuando se encuentra en concentraciones excesivas contribuye de manera significativa al efecto invernadero, elevando la temperatura del planeta y alterando el ciclo natural. Una forma de reducir las emisiones de CO2 es promover planes de movilidad sostenible, que buscan garantizar traslados de personas y mercancías de manera segura, sin poner en riesgo la salud ni los ecosistemas. Apoyar e incentivar el uso de vehículos eléctricos y bicicletas, entre otras medidas, sería clave para mitigar este problema.

En particular, los sistemas de alquiler de bicicletas en entornos urbanos, ofrecen una solución práctica y accesible para el transporte diario, al mismo tiempo que representa una estrategia clave para mitigar los efectos del cambio climático. En muchas ciudades, la implementación de estos sistemas se ha convertido en una herramienta popular para reducir la congestión vehicular, disminuir las emisiones de CO2 y mejorar la calidad del aire. No obstante, uno de los problemas recurrentes es la distribución desequilibrada de las bicicletas a lo largo de las estaciones de alquiler, lo cual reduce la eficiencia del servicio y afecta la percepción que tienen los usuarios sobre la fiabilidad del sistema. Así, uno de los retos fundamentales de estos sistemas es la capacidad para anticipar y gestionar la demanda de bicicletas de manera eficiente.

Se tiene entonces la necesidad de desarrollar soluciones, basadas en datos, para predecir los patrones de demanda en tiempo real, tomando en cuenta variables como el clima, la hora del día, la ubicación, los eventos locales o incluso los patrones históricos de uso. Esto permitiría a los operadores del sistema gestionar la distribución de las bicicletas de manera más eficiente, asegurando que las estaciones estén siempre bien abastecidas y que las bicicletas estén disponibles en los lugares y momentos adecuados.

**A. Objetivos.**

* Aplicar técnicas de regresión para construir un modelo predictivo que permita estimar la demanda sobre el uso de un sistema de alquiler de bicicletas siguiendo el ciclo de machine learning.
* Determinar cuáles son los factores que más inciden en la demanda con base en los datos.

**B. Conjunto de datos**.

El conjunto de datos recoge información sobre la cantidad de bicicletas rentadas en un período de tiempo, junto con información meteorológica y de temporalidad, entre otros. Es importante que revises el diccionario como primer paso para comprender estos datos. Los datos originales han sido tomados a partir de este enlace y han sido modificados para propósitos de este proyecto.

**C. Actividades para realizar.**

1. Exploración y perfilamiento de los datos, utilizando las funcionalidades de la librería pandas. Recuerda que este paso es muy importante para determinar problemas de calidad y tomar decisiones relacionadas con la preparación de los datos para el algoritmo de aprendizaje.
2. Limpieza y preparación de los datos, justificando las decisiones tomadas con base en los resultados obtenidos en el paso anterior.
3. Construcción de un modelo de regresión polinomial. Para determinar el grado de la transformación polinomial emplea las técnicas de selección de modelos sobre los siguientes valores de grado de polinomio: [2, 3]. Utiliza para la selección la métrica RMSE.
4. Construcción de un modelo de regresión regularizada Lasso. Para determinar el valor del hiperparámetro de regularización utiliza las técnicas de selección de modelos sobre los siguientes valores para α: [1, 2, 3, 4, 5]. Utiliza para la selección la métrica RMSE.
5. Elaboración de una tabla comparativa mostrando el rendimiento sobre test de los dos modelos seleccionados (con mejores rendimientos) de las actividades 3 y 4, con las métricas R2, RMSE y MAE.
6. Con base en el modelo Lasso determinar las variables más importantes para la predicción.

**D. Consideraciones.**

* Al hacer la división entrenamiento – test utiliza un valor de semilla de 77 (ramdon\_state).

**E. Análisis de resultados.**

Una vez construido los modelos, deberías estar en capacidad de responder estas preguntas:

* ¿Cuál es el grado de la transformación polinomial que fue seleccionado utilizando la técnica de validación?
* ¿Cuál fue el valor de α que fue seleccionado utilizando la técnica de validación para la regresión Lasso?
* A partir de la tabla comparativa, ¿cuál modelo ofrece el mejor rendimiento sobre el conjunto test? ¿Qué interpretación puedes darles a los valores obtenidos sobre las métricas de rendimiento?
* ¿Cuáles variables fueron seleccionadas con el modelo Lasso? A partir de estas, ¿qué interpretación de cara al problema puedes dar? Reflexiona sobre cómo este nuevo conocimiento podría ayudar a tomar decisiones en el contexto del problema.

**F. Entregable**.

Notebook (\*.ipynb y \*.html) por la plataforma. El Notebook debe estar documentado con las justificaciones de las decisiones tomadas en cada paso del ciclo de ML y las respuestas a las preguntas planteadas en el apartado D. Además, deben ser visibles las ejecuciones de cada celda. Esta entrega debe realizarse por el enlace que será habilitado para este fin, en “Actividades”.

**G. Criterios de evaluación.**

|  |  |
| --- | --- |
| Actividad | Porcentaje |
| Se realiza la exploración y perfilamiento de los datos. | 10% |
| Se realiza la limpieza y preparación de los datos, justificando las decisiones tomadas. | 15% |
| Se construye el modelo de regresión polinomial utilizando una técnica de selección de modelos sobre los valores especificados en la sección C, punto 3. | 15% |
| Se construye el modelo de regresión Lasso utilizando una técnica de selección de modelos sobre los valores especificados en la sección C, punto 4. | 15% |
| Se construye la tabla comparativa mostrando el rendimiento sobre test de los dos mejores modelos, sobre las métricas R2, RMSE y MAE. | 10% |
| Se deriva una tabla que muestra la importancia de cada variable con base en el modelo Lasso. | 5% |
| Se presenta el análisis de resultados con base en las preguntas de la sección E. | 30% |