

Gestión de riesgos ambientales

Actualmente, estamos ante la presencia de una variedad de fenómenos que están afectando el medio ambiente y que se reflejan en el calentamiento global, la contaminación del aire y la pérdida de biodiversidad, entre otros. Establecer políticas que ataquen sus causas y el desarrollo de planes de gestión de riesgos que mitiguen sus efectos y permitan accionar oportunamente es uno de los grandes retos a los que nos enfrentamos como sociedad.

Uno de estos problemas está relacionado con el cambio climático, el cual se refiere tanto al calentamiento global provocado por la emisión de gases de efecto invernadero, como al aumento de los desastres ambientales inducidos por dicho calentamiento. En particular, en el caso del dióxido de carbono (CO_2), aunque es necesario para la vida, su exceso de concentración contribuye significativamente al efecto invernadero, que eleva la temperatura del planeta y desequilibra el ciclo natural. Una manera de reducir las emisiones de CO_2 es fomentar planes de movilidad sostenible, la cual se refiere a la capacidad de hacer traslados, de personas o mercancías, con seguridad y sin afectar la salud ni los ecosistemas. Apoyar e incentivar medidas de movilidad sostenible, como el uso de vehículos eléctricos y bicicletas, contribuiría en gran medida a paliar este problema.

Otro reto lo representa el deterioro de la calidad de los cuerpos de agua, como consecuencia del desarrollo urbano no planificado y sin criterios de preservación del ambiente. Es preocupante observar cómo los ecosistemas acuáticos están siendo afectados por grandes obras de ingeniería, cambios en el uso de la tierra y la sobreexplotación de recursos, entre otros. Resulta necesario entonces implementar mecanismos que permitan monitorear y evaluar las variaciones en las características fisicoquímicas del agua, con el fin de preservar la calidad de vida de estos ecosistemas y garantizarla disponibilidad de agua dulce para la población humana.

En estas situaciones, las técnicas de machine learning pueden ayudar en dos direcciones: en la construcción de planes de prevención y en la estimación de esfuerzos y recursos para mitigar los efectos de estos problemas.

Las dos etapas del proyecto se describen a continuación.

Tarea de regresión.

Objetivo. Aplicar técnicas de regresión para construir un modelo predictivo que permita estimar la demanda sobre el uso de un sistema de alquiler de bicicletas. Este conocimiento puede dar soporte para mejorar el servicio y conocer los factores que inciden en su eficiencia. 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.

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 (por ejemplo, valores ausentes y registros duplicados) 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 (random_state).
- En las técnicas de selección de modelos uno de los parámetros que debes indicar es scoring = 'neg_root_mean_squared_error', con el fin de seleccionar RMSE como métrica de evaluación en la búsqueda.

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 Git Hub. 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 E. Además,

deben ser visibles las ejecuciones de cada celda. Esta entrega debe realizarse en la cuarta semana, en donde encontrarás un espacio para adjuntar los dos archivos.