

Oportunidad Científico de Datos

Descripción:

En el marco de rastreo de proyectos de inversión relacionados con gestión del riesgo, se necesita un algoritmo para predecir si un proyecto está relacionado (clase positiva) o no relacionado (clase negativa) con gestión del riesgo.

Sus tareas implican el análisis de datos, el diseño, el procesamiento y el modelado de una solución de aprendizaje automático para predecir si un proyecto esta relacionado o no con gestión del riesgo y luego evaluar el modelo sobre los datos de prueba (20% del conjunto de datos).

Para la tarea se proporciona dos archivos csv, `proyectos_clase_positiva.csv` y `proyectos_clase_negativa.csv`; que representan ejemplos de las dos clases en cuestión. Cada base de datos tiene 2 campos, el identificador del proyecto y el nombre del proyecto.

Para la evaluación del modelo, usted deberá usar la métrica de accuracy. Adicionalmente debe escoger una métrica secundaria y también elaborar un argumento de por qué eligió dicha métrica.

Los entregables son:

- Un archivo `.py` incluyendo todo el código necesario para definir y evaluar el modelo.
- Un documento con la explicación de la metodología desarrollada, la segunda métrica propuesta, y el desempeño alcanzado con las métricas.