

Prueba de selección DS Machine Learning

Instrucciones:

- ◊ La prueba está en punto decimal y coma milésima.
- ◊ Responda las preguntas de lectura en un documento aparte, ya sea extensión .pdf o .docx.
- ◊ Las preguntas de modelamiento y código deben estar subidas a github, señalando la rama en donde se ha trabajado.
- ◊ Puede insertar ambos modelos en un framework de trabajo, como por ejemplo kedro, si así lo desea.
- ◊ Los códigos deben estar implementados **únicamente en Python**, pero se puede utilizar cualquier librería de preferencia.

1 Preguntas de lectura

Pregunta 1: ¿Cuál es el objetivo del trabajo?

Pregunta 2: Explique la metodología utilizada en el paper.

Pregunta 3: Menciones posibles mejoras al trabajo.

2 Preguntas de modelamiento y programación

Pregunta 1: La flotación es un proceso muy utilizado en la recuperación de los minerales sulfurados de cobre debido a la hidrofobicidad natural que tienen. Esta es utilizada ampliamente para recuperar diversos tipos de minerales, principalmente sulfurados conteniendo especies de cobre, hierro, oro, plata, plomo, zinc, níquel, entre otros.

El objetivo principal es utilizar estos datos para predecir cuánta impureza hay en el concentrado de mineral. Como esta impureza se mide cada hora, si podemos predecir cuánta sílice (impureza) hay en el concentrado de mineral, podemos ayudar a los ingenieros, brindándoles información temprana para tomar medidas. Por lo tanto, podrán tomar acciones correctivas con anticipación (reducir la impureza, si es el caso) y también ayudar al medio ambiente (reducir la cantidad de mineral que va a los relaves a medida que reduce la sílice en el concentrado de mineral). Contenido

La primera columna muestra el intervalo de tiempo y fecha (desde marzo de 2017 hasta septiembre de 2017). Algunas columnas fueron muestreadas cada 20 segundos. Otros fueron muestreados cada hora.

Las columnas segunda y tercera son medidas de calidad de la pulpa de mineral de hierro justo antes de que se introduzca en la planta de flotación. La columna 4 hasta la columna 8 son las variables más importantes que impactan en la calidad del mineral al final del proceso. Desde la columna 9 hasta la columna 22, podemos ver los datos del proceso (nivel y flujo de aire dentro de las columnas

de flotación, que también afectan la calidad del mineral. Las dos últimas columnas son la medición final de la calidad de la pulpa del mineral de hierro del laboratorio. El objetivo es predecir la última columna, que es el % de sílice en el concentrado de mineral de hierro. En base a esto se pide:

- ◇ Predecir el % de concentrado de sílice.
- ◇ ¿Con cuántos pasos (horas) de anticipación podemos predecir el % de sílice en el concentrado?
- ◇ ¿Es posible predecir el % de sílice en concentrado sin usar la columna de % de concentrado de hierro (ya que están altamente correlacionados)?

Puede utilizar la cantidad de modelos que quiera, siempre indicando cuál es el que indicó mejores resultados.