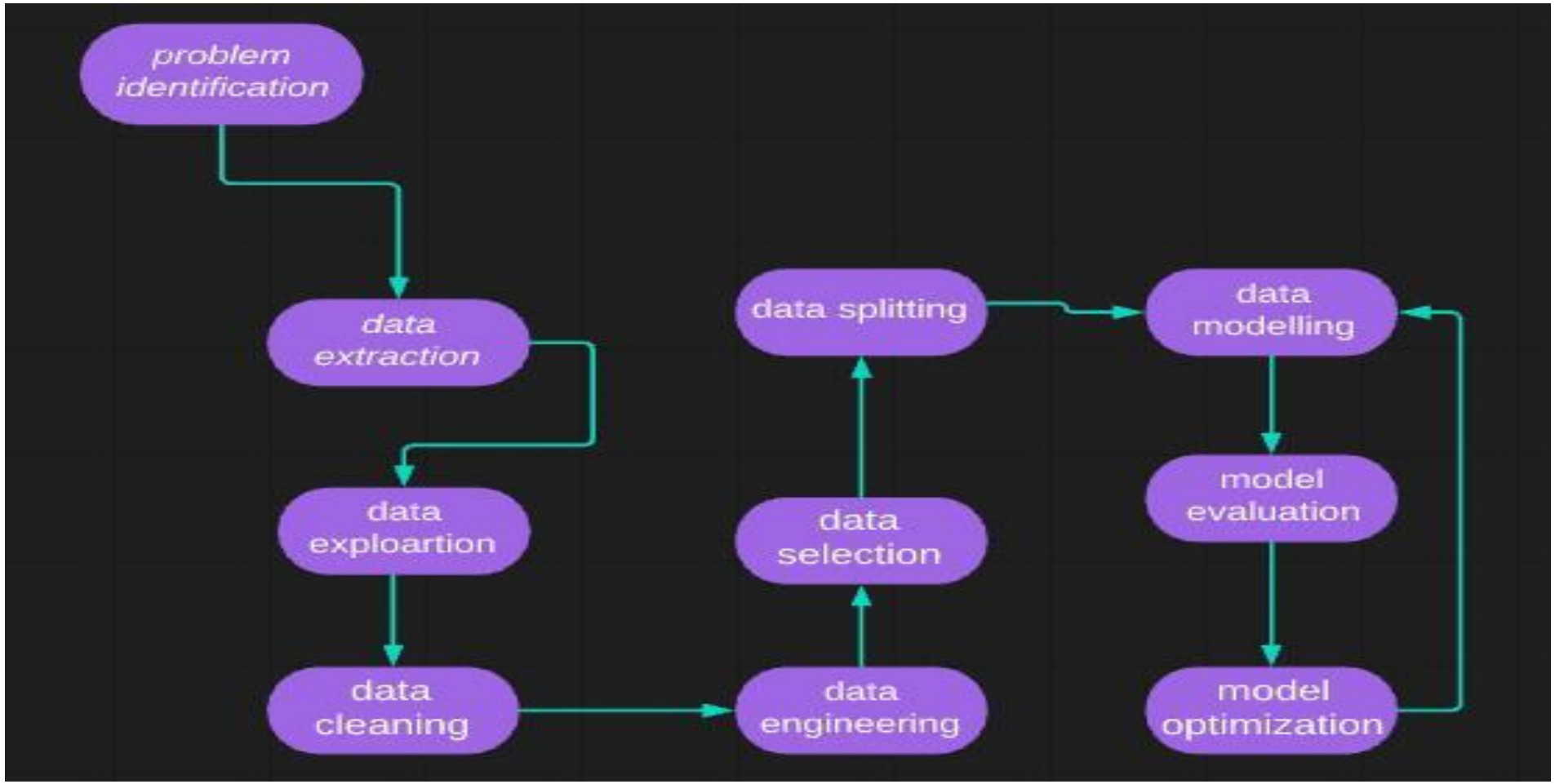
Create a cognos analytics story

Problem statement:

Safe and readily available water is important for public health, whether it is used for drinking, domestic use, food production or recreational purposes. Better water supplies and sanitation, as well as better management of water resources, can contribute greatly to poverty reduction and economic growth. It is known that contaminated water and inadequate sanitation facilitate the transmission of diseases such as cholera, diarrhoea, dysentery, hepatitis A, typhoid, and polio.

Process flow:



Dataset:

The dataset contains water quality metrics for 3276 different water bodies:

- Ph: pH of water (0 to 14).
- Hardness: Capacity of water to precipitate soap in mg/L.
- Solids: Total dissolved solids in ppm.
- Chloramines: Amount of Chloramines in ppm.
- Sulfate: Amount of Sulfates dissolved in mg/L.
- Conductivity: Electrical conductivity of water in µS/cm.
- Potability: Indicates if water is safe for human consumption. Potable – 1 and Not potable - 0

Importing dataset:

Df=pd.read_csv('water_potability.csv')

Df=pd.read_csv('../input/water-potability/water_potability.csv')

Models used for training:

- Logistic Regression
- Support Vector Classifier
- Random Forest Classifier
- XGBoost.