Milestone 2

NAME:BALAJI KOLUSU

**Dataset - Billionaires Statistics Dataset (2023)**

Ideas that were iterated

**Clustering Analysis:**

Initially, we attempted clustering analysis on the billionaire dataset using KMeans. This helped identify natural groupings or clusters based on certain features.

Key Features: 'age', 'finalWorth', 'selfMade', 'gender', 'population\_country', 'category'.

Insights: Clustering aimed to group billionaires with similar characteristics or business profiles.

**Data Preprocessing:**

Handled missing values in the dataset and converted categorical variables to numerical using LabelEncoder.

Notable features encoded: 'selfMade', 'gender', 'category', 'city', 'source', 'title', 'status'.

**Linear Regression Model:**

Applied a simple linear regression model using features like 'age', 'finalWorth', 'selfMade', 'gender', 'population\_country', 'category'.

Linear regression was used to predict a continuous variable, 'finalWorth'.

**SVM Model**

We also tried using SVM model to predict the status of the billionaire which performs poorly than the RandomForestClassifier. It is also presented in the code for reference

**Idea that worked**

**RandomForestClassifier:**

Utilized RandomForestClassifier for classification tasks.

Key Features: 'age', 'finalWorth', 'selfMade', 'gender', 'population\_country', 'category'.

Target Variable: 'status'.

Achieved a Test Accuracy of 60%, indicating the model's effectiveness in predicting the 'status' of billionaires based on given features.

**Analysis on why the selected features gave good results**

**Age**:

Age is often a significant factor in wealth accumulation. Younger billionaires might be more likely to be self-made entrepreneurs, while older ones might inherit wealth.

FinalWorth:

The final net worth is a direct indicator of a billionaire's economic success. It's a crucial feature for predicting their status, whether self-made or inherited wealth.

**SelfMade**:

This binary feature indicates whether the billionaire is self-made. Founders and entrepreneurs are more likely to have 'D' (self-made) status.

**Gender**:

Gender might play a role in wealth accumulation patterns. The model can learn whether certain genders are more likely to be self-made billionaires.

**Population\_Country**:

The population of the billionaire's country could be relevant. In countries with larger populations, there might be more self-made entrepreneurs due to increased economic opportunities.

**Category**:

The category or industry in which the billionaire operates is crucial. Certain industries might be more conducive to self-made wealth.

**Reasons for Success:**

Relevance to Wealth Accumulation: Features like 'Age' and 'FinalWorth' directly relate to a billionaire's financial success.

Informativeness of SelfMade and Gender: Identifying whether a billionaire is self-made and their gender provides insights into wealth origins and potential societal factors.

Economic Context (Population\_Country): Considering the population of the billionaire's country adds an economic context, potentially capturing regional influences on wealth creation.

Industry Influence (Category): The category or industry factor helps the model understand if certain sectors are more likely to produce self-made billionaires.

Diversity in Features: The selected features cover demographic, financial, and contextual aspects, providing a diverse set for prediction.

In summary, the chosen features align well with expectations about how wealth is accumulated and provide a comprehensive view of a billionaire's profile. The success of these features reflects their relevance to the target variable 'status.'

**Results -**

A screenshot of a computer

Description automatically generated

*Our work differentiates with other work on kaggle as no one has developed a classification model to predict the status of the billionaire.*