**Mark Raouf Aziz 211001635**

**Machine learning Assignment**

**Aim of the model:** To predict the winner in the WWE matches by analyzing the data and insights of each fighter.

**Dataset’s brief description:** This dataset comprises data regarding numerous wrestling champions and their respective periods of holding titles. Each row corresponds to a distinct wrestler's championship tenure, encompassing details like the wrestler's identity, the length of their reign (measured in days), the date and venue of the event where they attained the championship, their date of birth, and, if relevant, the date and reason for their demise.

**Task 1.1:** The code imports libraries needed for data analysis and machine learning, adjusts display settings, and reads an uncleaned dataset from a CSV file named 'data.csv'. This dataset will be used for further analysis and machine learning tasks.

**Task 1.2**: The code reads a dataset from a CSV file, keeps the first 140 rows, and filters out rows where wrestlers had more than one championship reign, leaving only records of their first reigns.

**Task 1.3 & 1.4:** Correlation Heatmap: This plot illustrates the relationships between numerical variables in the dataset using colors. It assists in identifying how variables influence each other.

Histogram of Days as Champion: This histogram showcases the distribution of the duration wrestlers held their championship titles, offering insights into typical title reign lengths.

Cause of Death: This countplot displays the frequency of different causes of death among wrestlers, aiding in understanding health and safety patterns within the wrestling community.

**Task 2:** Parsing Dates: It provides a custom function called parse\_date that handles conversion problems and transforms date strings into datetime objects.

Date Conversion: The 'date' and 'date\_of\_birth' columns are converted into datetime format using the parse\_date function.

Engineering Features:

Championship Density: Determines the density of championship reigns by taking into account the entire time period that the dataset spans.

Age at Winning: Determines the wrestlers' age at the time of their championship title victories.

Title Reigns per Year: This metric establishes how many times a wrestler typically wins the belt in a given year.

Location frequency: determines the frequency of each site where events have occurred.

Reign Duration: This new feature merely determines how long each wrestler's title reign will last.

Experience Level: Determines how long a wrestler has been involved in the wrestling business.

Winner: Produces a binary characteristic that indicates whether or not a wrestler held the title for a full year (365 days).

**Task 3:** Based on the amount of unique values in the target variable, this code defines if the task is a regression/clustering or classification activity. It is determined that the task is a classification because the goal variable "winner" only has two possible values.

The code uses the attributes "reign" and "days" to train and assess three classification models (Logistic Regression, Random Forest, and Gradient Boosting) for the classification job. The precision of every model is printed following training and testing.

And this is the results of the accuracies:

**Logistic Regression Accuracy: 0.91**

**Random Forest Accuracy: 0.91**

**Gradient Boosting Accuracy: 0.91**

**Task 4**: This code trains multiple classification models, generates predictions, and evaluates their performance. It prints the accuracy of each model and provides a detailed classification report including precision, recall, and F1-score for each class.

Then, After printing the accuracy of each model, the code stores these accuracies in a dictionary and generates a horizontal bar plot to visually compare the accuracies of different models.

**That is the bar plot comparing the accuracies of different models:** A blue and white bar graph

Description automatically generated

**Limitation:** The dataset should have more important features than the mentioned ones to guarantee more accurate predictions for the WWE upcoming champions.

**Deployment:** This Streamlit application simplifies the process of predicting the WWE Championship winner by offering just one button for users to click. Upon clicking the "Predict Winner" button, the latest championship data, such as the wrestler's reign and the duration of their championship, is fetched from the dataset. This information is then used as input for a pre-trained machine learning model, which makes a prediction about the winner. Subsequently, the app presents the predicted winner's name to the user. This straightforward approach streamlines the prediction task, offering users swift and precise championship winner forecasts.

**Conclusion:** In summary, the goal is to create a model predicting WWE match winners using wrestlers' historical championship data. The dataset includes details like reign duration, date of birth, and cause of death. Tasks involved data cleaning, feature creation, and visualization. Classification models were trained and tested, achieving high accuracy (91%). Through thorough data processing and model evaluation, the project successfully predicts match winners based on wrestlers' championship histories.