PUBG Win Prediction Project Report

# 1. Project Title

Predicting PUBG Match Win Probability Using Machine Learning

# 2. Objective

This project aims to:  
- Analyze in-game player statistics from PUBG matches.  
- Build a regression model to predict the player's win probability (winPlacePerc).  
- Identify the key factors that influence match outcomes.

# 3. Problem Type

- Supervised Machine Learning – Regression  
- Target Variable: winPlacePerc (float between 0 and 1)  
- Models Used: LinearRegression, XGBRegressor

# 4. Steps Taken in the Project

Step 1: Data Loading and Inspection  
- Loaded the PUBG dataset using pandas.  
- Inspected the dataset for shape, data types, and null values.

Step 2: Data Cleaning and Preprocessing  
- Dropped irrelevant columns like Id, matchId, and groupId.  
- Handled missing values using fillna().  
- Converted data types: float to int, string to float, numerical to string where required.  
- Created new features: totalDistance, boostHeals.  
- Encoded categorical columns like matchType using one-hot encoding or label encoding.

Step 3: Exploratory Data Analysis (EDA)  
- Analyzed distributions of key features using histograms and boxplots.  
- Investigated correlations using heatmaps and pairplots.  
- Identified which features were most associated with higher winPlacePerc.

Step 4: Feature Selection  
- Removed low-importance or highly correlated features.  
- Retained top features like walkDistance, boosts, damageDealt, kills, etc.

Step 5: Model Building  
- Split the dataset into train and test sets.  
- Trained a Linear Regression model as a baseline.  
- Trained an XGBRegressor model for improved accuracy.  
- Evaluated using metrics like RMSE and R².

Step 6: Prediction and Evaluation  
- Used trained models to predict winPlacePerc for test data.  
- Checked performance on both single player samples and full test sets.  
- Interpreted prediction scores as win probabilities.

Step 7: Model Interpretation  
- Extracted feature importances from the XGBoost model.  
- Visualized top contributing features.

Step 8: Final Report  
- Summarized the methodology, findings, and recommendations.  
- Provided example predictions and suggestions for future improvements.