Sports Analytics Project Report

Report

1. Introduction

This project focuses on sports analytics to analyze match data, evaluate team performance, and predict match outcomes using machine learning techniques. The objective is to apply data analysis and predictive modeling to extract actionable insights from sports data.

2. Dataset Description

The dataset contains match-level information including participating teams, match scores, venue details, and match winners. Categorical attributes were encoded and numerical features were selected to prepare the data for machine learning models.

3. Data Preprocessing

Data preprocessing steps included handling missing values, encoding categorical variables such as team names and venues, and creating a target variable indicating whether Team 1 won the match. These steps ensured the dataset was suitable for analysis and modeling.

4. Model Development & Evaluation

A Logistic Regression model was implemented to predict match outcomes based on team scores and venue information. Due to the small dataset size, a simple train-test split was used. Model performance was evaluated using accuracy and classification metrics.

5. Conclusion

The project demonstrates the practical use of data analytics and machine learning in sports analysis. With a larger dataset, advanced models such as Random Forest or Gradient Boosting could further improve prediction performance.

Summary

This Sports Analytics project analyzes match data to understand team performance and predict match outcomes using machine learning. The dataset was cleaned and prepared through handling missing values and encoding categorical variables. Exploratory analysis helped identify performance patterns across teams and venues. A Logistic Regression model was used to predict whether Team 1 would win a match based on match-related features. The project highlights the importance of data preprocessing, feature selection, and model choice when working with limited datasets. Overall, the project demonstrates a practical application of sports data analytics and provides a foundation for further enhancements using larger datasets and advanced predictive models.