PROJECT WORK

UNLOCKING OLYMPIC INSIGHTS: USING MACHINE LEARNING TECHNIQUES FOR DATA ANALYSIS

Analyzing Olympic Data with Machine Learning

RADHIKA
322101027

Introduction to Olympic Data Analysis

LOVELY PROFESSIONAL UNIVERSITY

01

Overview of Olympic Games Datasets

Discussing various datasets available from the Olympic Games.

02

Machine Learning Techniques in Sports

Examining the different machine learning methods applied in sports.

03

Applications of Machine Learning

How machine learning is applied to improve sports analytics.

04

Future Implications for Sports Analytics

Analyzing the potential future impacts of data analysis in sports.

Machine Learning Techniques Overview

LOVELY PROFESSIONAL UNIVERSITY

Enhancing Olympic Data Analysis with Machine Learning

01

Supervised Learning for Predictive Models

Utilize historical performance data to predict athlete outcomes, such as medal wins or performance improvements.

02

Unsupervised Learning for Pattern Recognition

Identify hidden patterns and trends within athlete data, enabling insights into training effectiveness or talent identification.

03

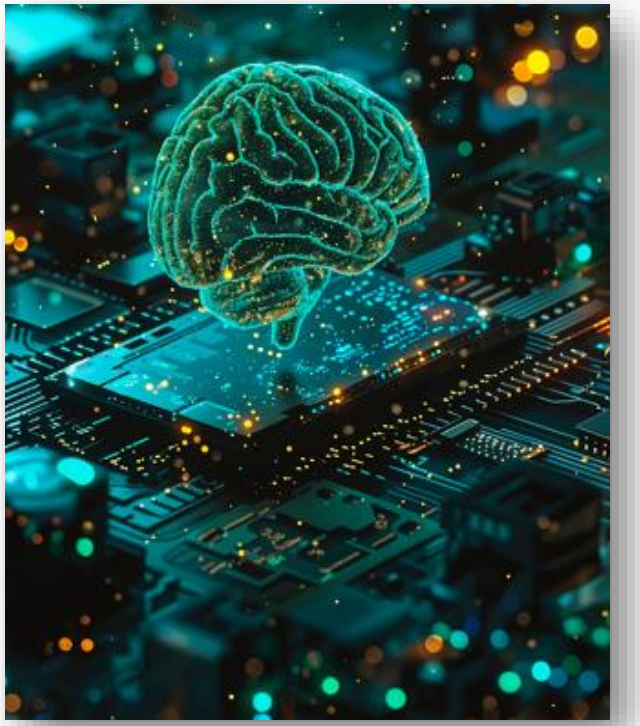
Decision Trees for Classification

Organize data into a tree-like structure to classify athletes based on specific characteristics or performance metrics.

04

Random Forests for Enhanced Predictions

Aggregate multiple decision trees to improve prediction accuracy and handle complex relationships within athlete data.



Key Applications of Machine Learning

Exploring Machine Learning in Olympic Sports Analytics

❑ Performance Enhancement

Identifies training needs to optimize athlete performance.

❑ Talent Identification

Metrics help discover emerging athletes with potential.

❑ Event Planning

Analyzes historical data to improve event organization.

❑ Injury Prevention

Predicting and preventing injuries based on athlete biometrics and training data.

Challenges in Data Analysis

LOVELY PROFESSIONAL UNIVERSITY

01

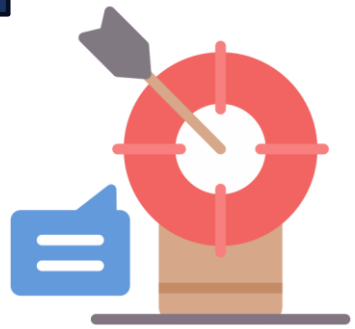
Data Quality, Accuracy, Consistency

Ensuring the accuracy of historical datasets is crucial for reliable analysis.

02

Complexity, Volume, Integration

Managing large volumes of data from various sources adds to the complexity of analysis.



03

Interpretability, Stakeholder Understanding, Machine Learning

Making machine learning results understandable is vital for stakeholder engagement.

04

Resource Allocation, Training, Tools

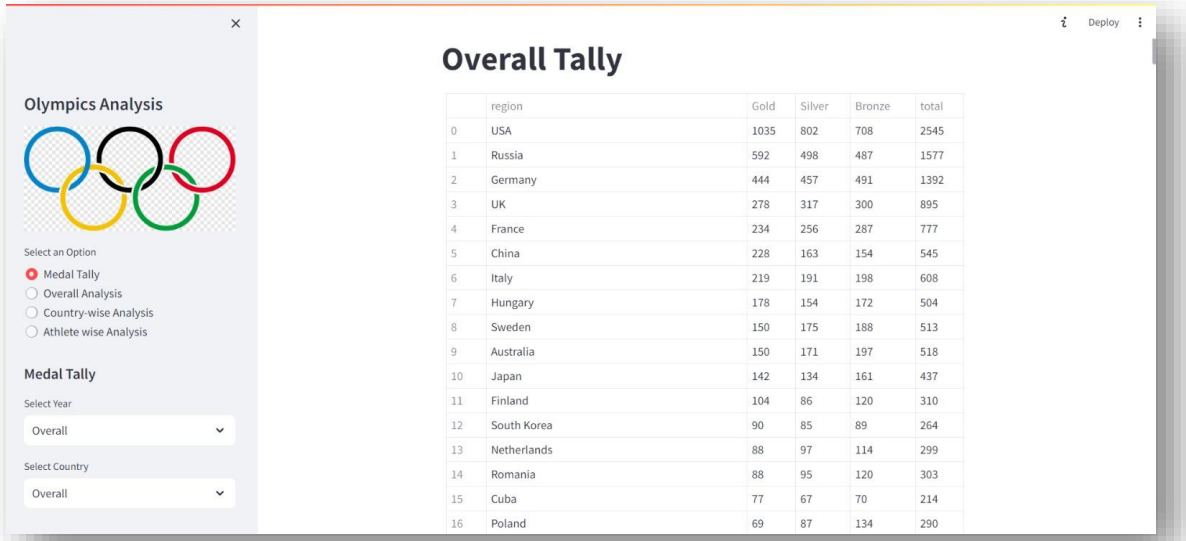
Insufficient resources and training can hinder effective data analysis.

Code Samples

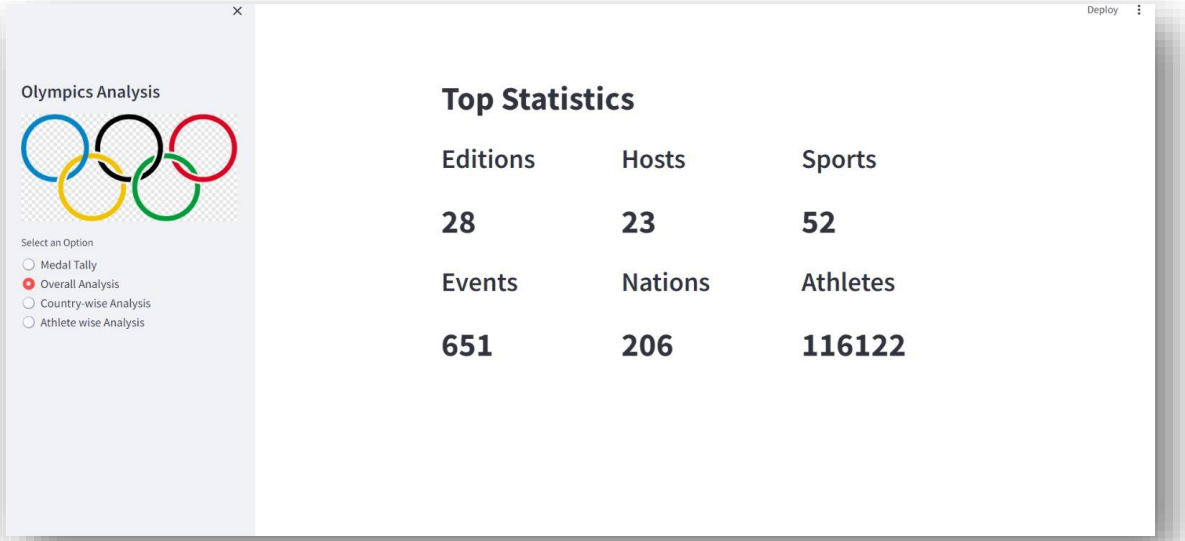
```python
import numpy as np
import pandas as pd
import streamlit as st
import preprocessor,helper
import plotly.express as px
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.figure_factory as ff

df = pd.read_csv('athlete_events.csv')
region_df = pd.read_csv('noc_regions.csv')

df = preprocessor.preprocess(df,region_df)


st.sidebar.title("Olympics Analysis")
st.sidebar.image('https://e7.pngegg.com/pngimages/1020/402/png-clipart-2024-summer-olympics-brand-circle-
user_menu = st.sidebar.radio(
    'Select an Option',
    ('Medal Tally','Overall Analysis','Country-wise Analysis','Athlete wise Analysis')
)

if user_menu == 'Medal Tally':
    st.sidebar.header("Medal Tally")
    years,country = helper.country_year_list(df)
```

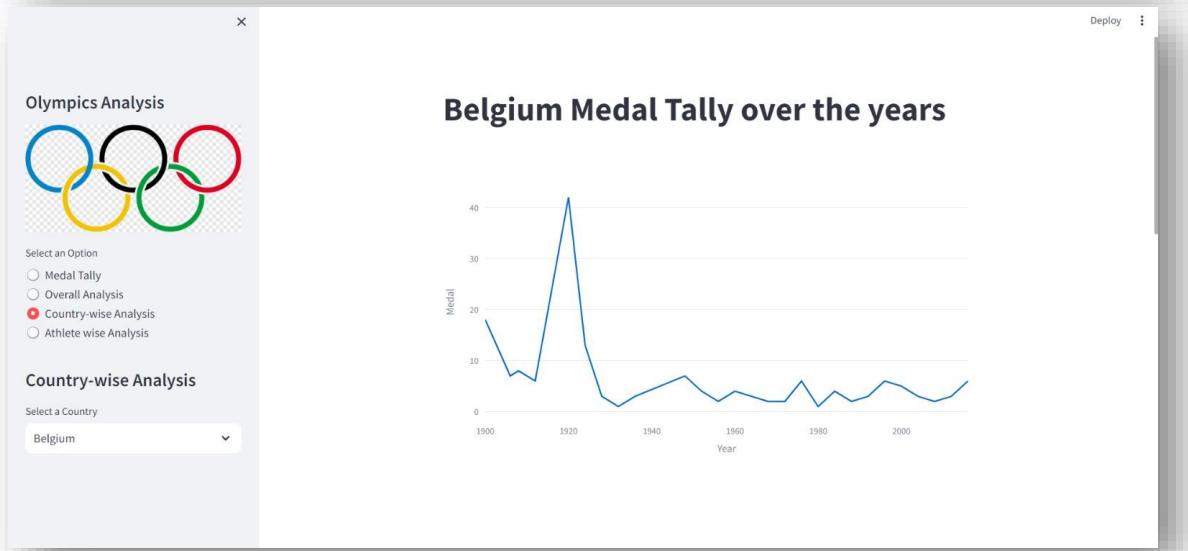
```python
import numpy as np


def fetch_medal_tally(df, year, country):  1 usage
    medal_df = df.drop_duplicates(subset=['Team', 'NOC', 'Games', 'Year', 'City', 'Sport', 'Event', 'Medal'])
    flag = 0
    if year == 'Overall' and country == 'Overall':
        temp_df = medal_df
    if year == 'Overall' and country != 'Overall':
        flag = 1
        temp_df = medal_df[medal_df['region'] == country]
    if year != 'Overall' and country == 'Overall':
        temp_df = medal_df[medal_df['Year'] == int(year)]
    if year != 'Overall' and country != 'Overall':
        temp_df = medal_df[(medal_df['Year'] == year) & (medal_df['region'] == country)]

    if flag == 1:
        x = temp_df.groupby('Year').sum()[['Gold', 'Silver', 'Bronze']].sort_values('Year').reset_index()
    else:
        x = temp_df.groupby('region').sum()[['Gold', 'Silver', 'Bronze']].sort_values('Gold',
                                                                                      ascending=False).reset_index()

    x['total'] = x['Gold'] + x['Silver'] + x['Bronze']


    x['Gold'] = x['Gold'].astype('int')
    x['Silver'] = x['Silver'].astype('int')
    x['Bronze'] = x['Bronze'].astype('int')
```

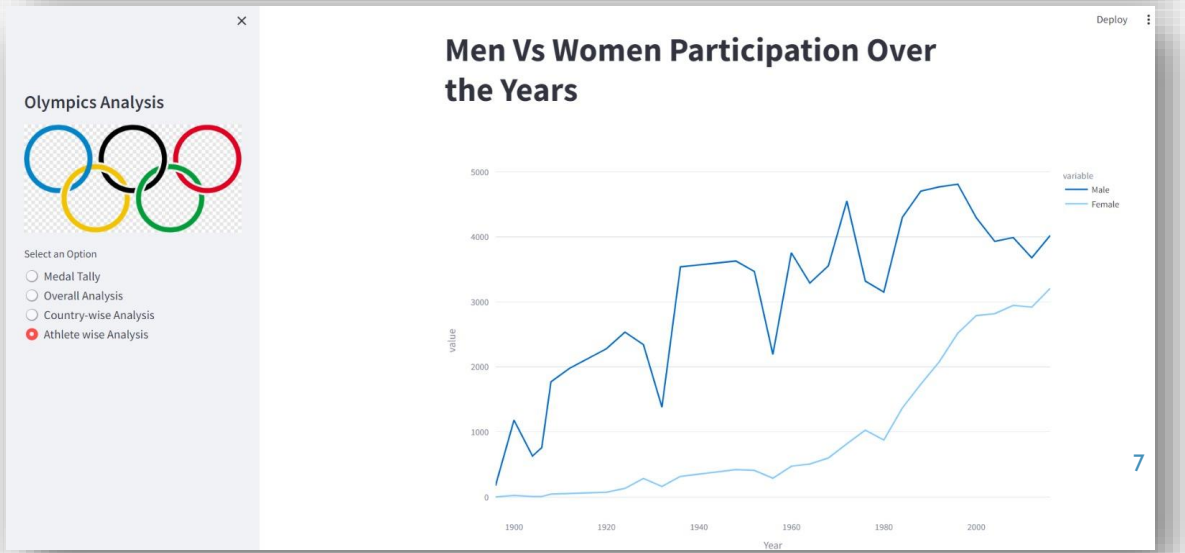
Final Outcomes



	region	Gold	Silver	Bronze	total
0	USA	1035	802	708	2545
1	Russia	592	498	487	1577
2	Germany	444	457	491	1392
3	UK	278	317	300	895
4	France	234	256	287	777
5	China	228	163	154	545
6	Italy	219	191	198	608
7	Hungary	178	154	172	504
8	Sweden	150	175	188	513
9	Australia	150	171	197	518
10	Japan	142	134	161	437
11	Finland	104	86	120	310
12	South Korea	90	85	89	264
13	Netherlands	88	97	114	299
14	Romania	88	95	120	303
15	Cuba	77	67	70	214
16	Poland	69	87	134	290







CONCLUSION AND FUTURE DIRECTIONS

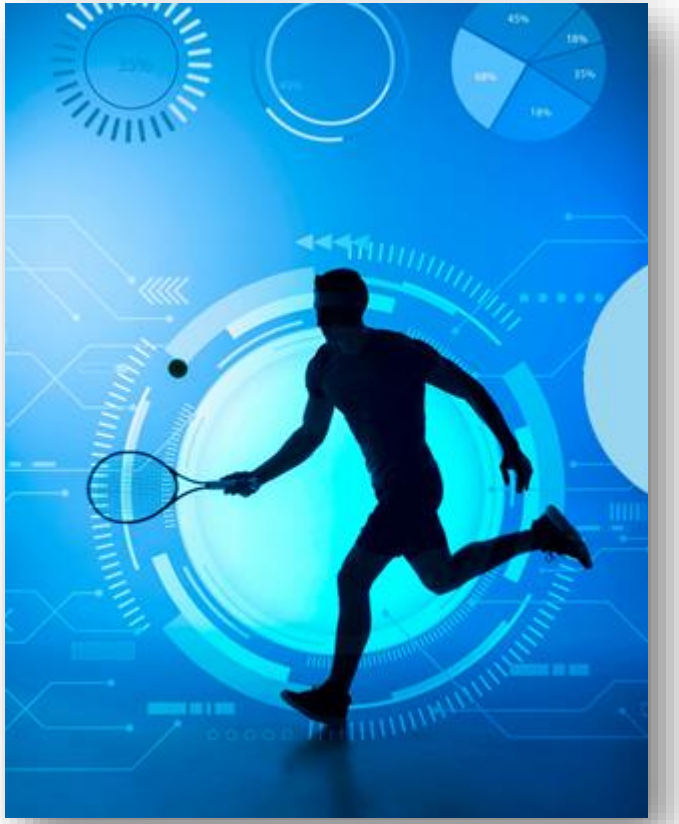
Future Work in Sports Analytics

- **Integration of real-time data analytics in Olympic sports**

Implementing real-time analytics to enhance performance monitoring and strategic decision-making.

- **Collaboration with sports organizations for data enhancement**

Partnering with sports entities to improve data collection methods and optimize data utilization in sports analytics.



CONCLUSION

In conclusion, machine learning provides powerful tools for analyzing Olympic data. As technology advances, the potential for deeper insights and enhanced athlete performance continues to grow. Future studies should focus on integrating real-time data for immediate decision-making in competitions.

LOVELY
PROFESSIONAL
UNIVERSITY

Thank you!