Analysis & Design Chapter

**Application Overview**

Provide a summary of your sports analytics platform, describing its purpose and key features. Mention that the project aims to analyze rugby player and team performance, integrating machine learning models and data visualization techniques to present useful insights.

**3.2 Potential Blockers and Considerations**

* **Data Availability and Quality**: Discuss the challenges of acquiring high-quality rugby data. Mention the potential issues in scraping reliable data or obtaining consistent statistics from different sources.
* **Dynamic Content Scraping**: Web scraping rugby statistics from various platforms will likely involve challenges with dynamically loaded content (e.g., JavaScript-rendered elements), which might require handling using tools like Selenium.
* **Computational Complexity**: Implementing machine learning models for predictive analysis can be computationally intensive. Highlight potential limitations due to processing power when handling large amounts of match data or training models.

**3.3 Scope**

Clearly define the boundaries of your project. The platform will focus on features such as:

* Visualizing match statistics, player performance, and predictive analysis.
* Tracking match data and player metrics, leaving out features like betting odds to maintain academic suitability.

**3.4 Prerequisites**

Mention the relevant background knowledge and tools you need for the project:

* **Web Scraping**: Knowledge of web scraping techniques and familiarity with tools like BeautifulSoup or Selenium.
* **Machine Learning**: Understanding of relevant machine learning models, such as classification (for player categorization) and regression models (for performance prediction).
* **Data Visualization**: Understanding of data visualization libraries such as Plotly or D3.js for presenting data effectively.

**3.5 Software Development Process**

* **Agile Methodology**: Describe how you'll adopt an agile approach, breaking down development into manageable sprints. Each sprint could focus on different parts of the project, such as data scraping, backend setup, machine learning implementation, and visualization.
* **Minimal Viable Product (MVP)**: Start with a streamlined version of the platform focusing on key features—match data visualization and basic player analysis—before expanding to predictive models.

**3.6 Tools and Frameworks Considered**

* **Backend Framework**: Python (Django) will be used as it allows for rapid development, scalability, and easy integration of machine learning models.
* **Frontend Framework**: React for an interactive user interface, suitable for handling dynamic and real-time data updates.
* **Database**: PostgreSQL or a similar relational database for structured storage of match data and player metrics.
* **Machine Learning Tools**: Use scikit-learn for initial model training and possibly TensorFlow or PyTorch for more advanced predictive capabilities.

**3.7 Dataset Considerations**

* **Data Sources**: Outline the rugby data sources you'll be using, whether from open datasets or through web scraping. Mention how the data will be preprocessed, such as transforming raw data into formats suitable for machine learning.

**3.8 System Architecture**

* **High-Level Design**: Include a diagram to show the interaction between the user interface, backend services, database, and machine learning modules.
* **Modular Design**: Mention that the system is designed in a loosely coupled manner, allowing future expansion to include more data sources or add features like mobile compatibility.

**3.9 Data Visualization and Exploration**

Discuss the approach for creating informative visualizations, including:

* Using graphs such as heat maps, bar charts, and line charts to convey match statistics.
* Integration of real-time match updates to provide dynamic visualizations similar to those seen in platforms like FotMob.

**3.10 Model Architecture Considered**

Explain the machine learning models you'll employ:

* **Classification Models** (e.g., Random Forest or Decision Trees) for categorizing player performance.
* **Time-Series Models** (e.g., LSTM networks) for predicting match outcomes based on historical data.

**3.11 Evaluation Metrics**

* **Performance Metrics**: Accuracy, precision, and recall for classification models.
* **Visualization Effectiveness**: User feedback on the clarity and usefulness of data visualizations.
* **Computational Efficiency**: How well the system handles large datasets in real-time.

A diagram of different colors

Description automatically generated

A diagram of a rugby diagram

Description automatically generated

List of use cases

Primary actor = User

* View Match Statistics
* Track Player Performance
* Compare Teams
* Predict Match Outcomes
* Analyse Team Form and Trends
* View Upcoming Fixtures and Predictions
* View Player Lineup and Head-to-Head Analysis
* Historical Data Analysis
* Visualize Set Play Analytics
* User Registration and Login ?????

Primary actor = System

* Real-Time Data Collection via Web Scraping

A diagram of a rugby diagram

Description automatically generated

A diagram of a rugby diagram

Description automatically generated

A diagram of a software development

AI-generated content may be incorrect.

Figure System Architecture Diagram

A screenshot of a computer

AI-generated content may be incorrect.

Figure Use Case Diagram

A diagram of a company

AI-generated content may be incorrect.

Figure Sequence Diagram

A screenshot of a computer

AI-generated content may be incorrect.

Figure Data Flow Diagram

A screenshot of a computer

AI-generated content may be incorrect.

Figure Entity-Relationship Diagram