Football Analytics

017536623 - Darpankumar Jiyani
017507100 - Dhruv Patel
017441359 - Kush Bindal
017552795 - Shobhita Agrawal

Abstract:

In today's fast-paced world of sports, making strategic moves using data analytics has become a game changer. Teams now understand the critical role of analytics in achieving competitive advantage. This project dives deep into football data analytics and explores strategies on and off the pitch. All the insights that can be found that are known or unknown to the football community can be highly impactful. Using extensive datasets from major leagues such as the UEFA Champions League, our analysis breaks down player performance in different game scenarios. By looking at players' dynamics along with environmental factors and opponent tactics, we reveal hidden insights that improve decision-making. Using advanced statistical techniques, we aim to insightful intelligence, revolutionizing player selection and tactical strategies. Ultimately, our goal is to show how soccer data analysis can succeed in the ever-evolving world of sports.

Motivation:

The Motivation for Football Analytics focuses on the improvement of the sport. The landscape of football analytics is currently reliant on manual computations, yielding limited insights such as goals scored, assists, and passing accuracy. To address this, we propose the development of a comprehensive model utilizing data from top football leagues over the past decade. By analyzing the data for individual players, strategy formation, and match statistics, teams can find fruitful insights to obtain strategies, improve player performance, and gain a competitive advantage. Traditional methods of gaining insights through repetitive SQL queries on extensive datasets are both cumbersome and time-consuming. Therefore, our project seeks to streamline this process by constructing a real-time performance model that harnesses historical data to deliver actionable insights promptly. By empowering coaches and team managers with timely and accurate performance predictions, our project aims to facilitate informed decision-making before matches, leading to enhanced team performance and strategic adjustments. This initiative marks a significant step towards optimizing football analytics and elevating the standard of performance assessment in the sport.

Literature Survey:

The research of (Klyuchka, Cherednichenko, Vasylenko, & Yakovleva, 2015) aimed to find the most important factors that are not confidential information and can be easily determined before the start of the football match. It presents that forecasting rules are used to increase the accuracy of predicting the results of football matches by identifying the winning team based on data retrieved from results of previous games championships, adding substantial factors, to understand the influence of results.

According to (Ali, 2011) and raising the bridge to arrive on the study subject, football is a complex sport, requiring the repetition of many disparate actions. For instance, there are several proofs of concepts that are

currently being used such as assessing the physical prowess of players, approaching simple running tests using monitor speed, agility proofs, and repeated sprint performance.

Joining what (Ali, 2011) concludes, (De Silva, et al., 2018) present that performance management of top football players is a complex system involving enhancement of physical performance, skill-based training, tactical training, minimization of injury risk, and psychological support. Managing practice is vital to allowing players to perform at an optimal level throughout a play season's length.

(Constantinou & Fenton, 2017) agrees with (Ali, 2011) and (FIFA, 2019) that Football is the most popular sport in the world and it leverages the inspiration of several researchers to use football activities as a real-world application field to test various statistical, probabilistic, and machine-learning techniques.

Methodology:

Experiment Design:

1. **Requirement Analysis:**
   - The project will be initiated by in-depth research about Football players' performance from major leagues such as the UEFA Champions League.
2. **System Design**:
   - The system will be designed on top of Python and SQL. The backend will be designed using Python as will provide efficient running of the application and can be easily scalable. SQL will be used for database management as using SQL we can create normalized tables from CSV datasets, and it will be robust.
3. **Database design:**
   - We will do data modeling for our project which will show our database structure, different tables, and their entities, relationships among tables, and their constraints.
4. **Development:**
   - Our project will be built on Python at the backend and SQL for database operations. Python APIs will be used for database connectivity, database operations, user interactions, and more.
5. **Documentation:**
   - We will create detailed end-to-end documentation of our project, which will help users in interacting with our application.
6. **Conclusion**:
   - Using the mentioned approach we will build a project that will meet user requirements and will be adherent to industry standards using SQL and Python as base technologies providing an efficient and scalable system.

Software to be Used:

Relational Database: MySQL
NoSQL Database: MongoDB
AWS tool: Amazon S3, IAM Roles
Data Analysis Tool: PowerBI, Python
ETL Tool: Power Query
Data Warehouse: SnowFlake
Data modeling techniques: ER diagrams

Evaluation Methods:

1. **Performance Testing:**
    - Analyze the efficiency of data retrieval and storage in MySQL and MongoDB databases, focusing on response times and query execution speeds across different data loads and query complexities.
2. **Scalability Assessment**:
    - Evaluate the scalability of the chosen databases by examining their performance under increasing data volumes and user loads, ensuring seamless operation and responsiveness as the system grows.
3. **Usability Testing**:
    - Gather feedback from end-users to assess the ease of data access, query execution, and visualization customization using the implemented methodology, ensuring user-friendliness and intuitive navigation.
4. **Security Analysis:**
    - Conduct a thorough security assessment of the entire methodology stack, including databases, data analysis tools, and cloud services, to identify and address potential vulnerabilities and ensure data confidentiality and integrity.

Deliverables:

1. **Final Report**:
    - A complete report that describes the proper methodology, analysis, ETL processes, and conclusions of the project.
2. **Database Implementation:**
    - Implement the architecture of the database that includes the structure of the table, schema design, indexing strategies, and data normalization.
3. **Data Exploration**:
    - Generate descriptive statistics of the data and also perform the exploratory data analysis to gain insight and distribution of the data.
4. **Data Visualization and Dashboards**:
    - Data Visualization, dashboards, or reports that show important insights and findings using tools like Power BI.
5. **Documentation:**
    - Prepare documentation on the design of the database, visualization techniques, ETL processes, and AWS configuration.
6. **Presentation:**
    - The purpose of the presentation is to summarize the key findings, fruitful insights, and recommendations of the whole project.

Milestones:

1. **Week 1-2: Getting Started:**
    - Define project goals and objectives.
    - Set up project management tools.
    - Collect football data from various sources.

2. **Week 3: Setting up Databases:**
   - Install **MySQL and MongoDB** databases.
   - Design database structures.
   - Start importing football data into databases.

3. **Week 4-5: Analyzing Data:**
   - Use **PowerBI** and **Python** to explore football data.
   - Create basic charts and graphs.
   - Identify important insights from the data.

4. **Week 6: Data Transformation:**
   - Use **Power Query** to clean and organize data.
   - Make sure data is ready for analysis.

5. **Week 7: Setting up AWS:**
   - Create storage buckets on **Amazon S3**.
   - Set up security measures with **IAM roles**.
   - Connect databases with Amazon S3 for backups.

6. **Week 8: Testing and Finalizing:**
   - Test database setups and **ETL processes**.
   - **Optimize performance**.
   - Document everything for future reference.

Team members and their roles:

1. **Darpankumar Jiyani**:
   - Project Proposal, Analysis stage 1, Evaluation of Results, Deployment phase(AWS), Final Report, Dashboard building

2. **Dhruv Patel:**
   - Finding the Dataset, Setting Evaluation Metrics, Setting up the Database in SQL, Cleaning data, Creating a project Blog, Final Data Analysis and Data Visualization, Final Report

3. **Kush Bindal:**
   - Setting on the database in NoSQL, Data cleaning, Creating Intermediate Status Report, Setting up the Database in SQL, Beta Testing, Final Report

4. **Shobhita Agrawal:**
   - Understanding Tools, Deployment Phase(AWS), Dashboard building and designing, Preparing Presentation Slides, Final Report

Relevance to the course:

The project on football data analytics aligns closely with the syllabus of the DATA-225 course, demonstrating relevance and integration within the course topics.

The course covers foundational concepts such as the structure of Database Management Systems (DBMS), transaction management, and data models.

Moreover, the exploration of ER and Enhanced ER Models, along with database architecture and applications, is mirrored in our project's scope. We delve into the modeling of football-related data using ER diagrams and mapping them into relational models, ensuring data integrity and efficient querying capabilities.

The course also emphasizes SQL, including data definition, manipulation, and complex queries. In our project, we extensively utilize SQL for querying & python libraries for data exploration of the football data to perform analytics, and generate insights to enhance decision-making processes.

The syllabus covers NoSQL databases like MongoDB, which directly relates to our project's exploration of non-relational data storage solutions for football data analytics.

Topics such as data warehousing, data cleansing, and ETL processes are integral components of both the course curriculum and our project scope. We leverage these concepts to design and implement a robust data warehousing solution for storing and analyzing football data, ensuring data quality and reliability.

Overall, the football data analytics project encompasses a broad spectrum of topics covered in the DATA-225 course, providing with practical hands-on experience in applying course concepts to real-world scenarios in the realm of sports analytics.

Technical Difficulty:

1. **Data Collection:**
   - For a collection of various datasets from multiple sources like goals, player information, stadium information, player's injury records, etc. It could be a challenging task as data may be inconsistent and in different data formats.

2. **Scalability and Performance:**
   - For designing scalability and handling the large amount of data of football, user interactions, and statistics tasks can be challenging. Database design and computational algorithms for efficiency and scalability is crucial to ensure proper performance and responsiveness.

3. **Security and Data Privacy:**
   - To ensure the security and privacy of the large football dataset, such as player information, team formation is critical.

4. **Data Cleaning and Preprocessing:**
   - When we try to clean a huge dataset then it would be difficult to interpret which columns are useful and which are not. In Football data it might be possible that it contains missing values, null values, data outliers and inconsistencies in the data.

Novelty and Uniqueness:

1. To obtain fresh insight to improve the game's decision-making, we are looking into different approaches to data analysis in football.

2. Our analysis offers new viewpoints on player performance, team tactics, and game conditions, providing a more comprehensive understanding of football.

3. We're analyzing football data more accurately and practically than ever before because of the modern methods and tools, which enhance the value of our outcomes.

4. Our objective is to make data simple to understand and use in real time, allowing coaches and teams to make better decisions during games.

5. We're setting up new standards for football managing data by utilizing various database types to ensure that our data is well-organized and easy to access.

Impact:

The project on football data analytics holds significant potential for impactful outcomes and chances of publication due to several key factors:

1. **Unique Approach:**
   - Our innovative methodologies and interdisciplinary collaboration offer a fresh perspective on sports analytics, increasing the likelihood of generating novel insights and findings.

2. **Practical Application:**
   - By focusing on real-world scenarios in football management, our project addresses tangible needs within the sports industry, enhancing its utility and relevance.

3. **Stakeholder Engagement:**
   - The project's potential to benefit stakeholders such as coaches and team managers increases its appeal for publication and dissemination, as it addresses pressing challenges in sports management.

4. **Methodological Rigor:**
   - Through the utilization of advanced database systems and SQL techniques, our project ensures robust data analysis and scholarly rigor, enhancing its credibility and potential for publication.

Overall, the project's innovative approach, practical relevance, stakeholder engagement, and methodological rigor contribute to its potential impact, publication opportunities, and utility in the field of sports analytics.

Heilmeier Catechism:

1. **What is the primary aim of our project?**
   - Our primary goal is to conduct in-depth analysis of football data utilizing advanced database systems and SQL techniques, with a focus on facilitating strategic decision-making in sports management, avoiding technical language.

2. **How is the current landscape of sports analytics, and what are its constraints?**
   - Current methodologies often rely on manual processes, limiting the scope of comprehensive data analysis. Our project introduces innovative techniques to address these limitations and enhance analytical capabilities.

3. **What novel approaches does our project introduce, and why are we confident in its success?**
   - We employ cutting-edge database systems and SQL methodologies to streamline data processing and analysis, anticipating significant advancements in sports analytics effectiveness.

4. **Who stands to gain from our accomplishments, and what implications will they have?**
   - Various stakeholders in football, including coaches and team managers, will benefit from the insights derived from our thorough data analysis, ultimately leading to improved team performance.

5. **What risks do we anticipate encountering?**
   - Potential challenges include issues related to data quality and technical complexities in database management, necessitating proactive measures to mitigate risks effectively.

6. **What financial considerations have we made for the project?**
   - Our budget encompasses personnel, software, hardware, and data-related expenses, requiring meticulous planning and resource allocation to ensure project success.

7. **What is the anticipated timeline for project completion?**
   - We anticipate the project to span several weeks, with clearly defined milestones for data acquisition, analysis, and reporting, ensuring adherence to the projected timeline.

8. **How do we intend to monitor progress and gauge success?**
   - Progress will be monitored through mid-term assessments focusing on database setup and preliminary analysis, followed by comprehensive evaluations of project deliverables and actionable insights to ensure successful project outcomes.