

Structured and Unstructured Data - Football Analysis

Ashwin Ajoy Dharmavaram - 1RV22AI011

Questions:

1. Choose an industry or application area such as healthcare, automotive or finance.
2. Describe one example of structured data and one example of unstructured data generated or used in that domain
3. Explain how each type of data is stored and processed technically (e.g. databases, data lakes, big data tools).
4. Discuss the engineering challenges involved in integrating and analyzing these two types of data to produce useful insights or automation.
5. Suggest technologies, algorithms, or architectures that can help overcome these challenges (e.g., SQL, NoSQL, Hadoop, AI/ML).
 1. Football Analysis
 2. **Structured Data:** Player statistics: Goals, assists, minutes played, xG, etc.
Unstructured Data: Player popularity, public opinion of player performances.
 3. **Structured Data:** This type can be stored in RDBMs, where neat relations can be made between players, clubs, competitions, etc, and data can be queried efficiently.
Unstructured Data: Since this consists of massive amounts of paragraphs with only metadata as structure, they can be stored in **NoSQL** databases, or data lakes.
 4. **Data Integration:** Clubs would need data on both player statistics, and how well they fit in with their club in order to sign them. This needs combination of structured and unstructured data.
Scalability: Larger the clubs get, more amount of analysis would be required. So, addressing scalability issues are vital.
 5. For structured data, regular SQL queries should be enough for analysis. For unstructured data:
 - (a) NLP for sentiment analysis of social media posts
 - (b) ML models for outlier analysis on under the radar players.
 - (c) Computer Vision for video analysis.