

**Project Title:** Player Performance Index (PPI) – A Context-Aware Cricket Player Ranking System

**Summary:** This study aims to develop a cricket player ranking system by leveraging ball-by-ball match data from Cricsheet. Existing ranking methods often overlook game context, player contributions, and critical performances. This project introduces a Player Performance Index (PPI) that integrates statistical analysis, machine learning models, and contextual factors to produce a more refined player ranking mechanism.

**Research Question:** How can a contextual Player Performance Index (PPI) be developed to assess cricket players' performance using match data and advanced computational techniques?

**Project Objectives:**

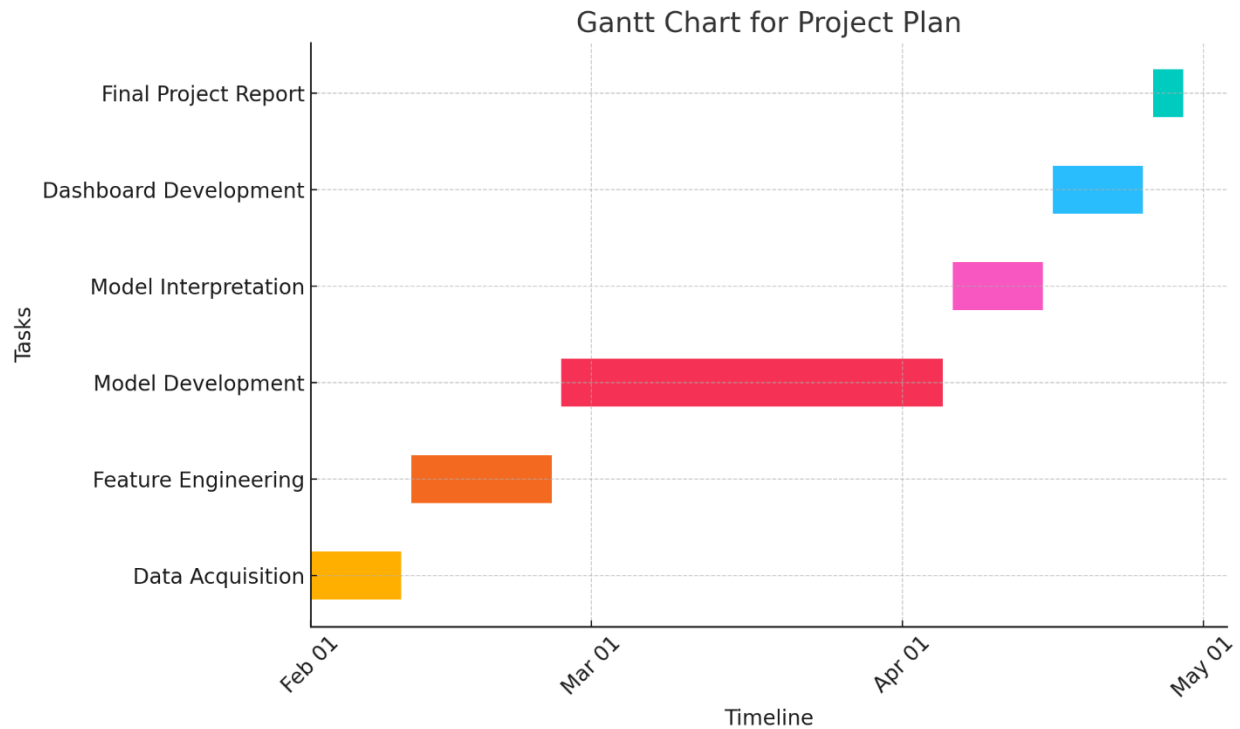
1. Gather and preprocess cricket match data while maintaining data integrity.
2. Develop feature extraction techniques focusing on key performance metrics.
3. Implement and evaluate machine learning models (Random Forest, XGBoost, Neural Networks) to determine player rankings.
4. Utilize model explainability tools, such as SHAP values, for interpretability.
5. Design a user-friendly dashboard for visualizing ranking insights.

**Reference List:**

- Deep Prakash, C., Patvardhan, C., & Singh, S. (2016). "A New Machine Learning-Based Deep Performance Index for Ranking IPL T20 Cricketers." *International Journal of Computer Applications*, 137(10), 42-49
- Bharadwaj, F., Saxena, A., Kumar, R., Kumar, R., Kumar, S., & Stević, Ž. (2024). "Player Performance Predictive Analysis in Cricket Using Machine Learning." *Revue d'Intelligence Artificielle*, 38(2), 449-457

---

**Project Plan: Task List and Timeline**



### 3. Data Management Plan

**Dataset Overview:** The dataset comprises ball-by-ball match records for men's international and domestic cricket, sourced from Cricsheet. It includes player actions, match settings, and performance indicators.

#### Data Collection:

- **Source:** Cricsheet ( <https://cricsheet.org/> )
- **Format:** CSV files
- **Estimated Dataset Size:** Around 50,000+ match records

#### Metadata:

- **File Type:** CSV
- **Record Count:** Varies per dataset
- **Storage Requirements:** ~5GB (compressed format)

#### Version Control:

- **GitHub Repository:** <https://github.com/Sh624-web/Player-Performance-Index-PPI-A-Data-Driven-Contextual-Cricket-Ranking-System>
- **Commit Frequency:** Weekly commits for version tracking
- **File Naming Convention:** match\_data\_YYYYMMDD.csv, ppi\_model\_vX.py

#### ReadMe File Inclusions:

- Project summary
- Installation and usage guide
- Dataset structure and attributes
- Code implementation details
- Licensing information

**Storage and Security Measures:**

- Data and code will be securely stored in OneDrive and GitHub.
- Restricted access will be applied to protect data integrity.

**Ethical Considerations:**

1. **GDPR Compliance:** The dataset consists of publicly available cricket statistics and does not contain personally identifiable information.
2. **UH Ethical Policies:** The project adheres to the ethical guidelines set by the University of Hertfordshire.
3. **Permission for Data Use:** The dataset is openly licensed for research purposes.
4. **Data Collection Ethics:** The data is obtained from public sources and follows ethical data-sharing practices.