AI-BASED KABADDI MATCH TEAM FORMATION AND STRATEGY RECOMMENDER

A PROJECT REPORT

Submitted by

THIYANESHWAR T - 2022115136

SUGANTH S A - 2022115308

MYTREYAN JP - 2022115102

SRILEKHA RAMKUMAR - 2022115076

submitted to the faculty of

INFORMATION AND COMMUNICATION ENGINEERING

in partial fulfillment

for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

INFORMATION TECHNOLOGY

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING GUINDY

ANNA UNIVERSITY

CHENNAI 600 025

SEPTEMBER 2025

ANNA UNIVERSITY

CHENNAI - 600 025

BONAFIDE CERTIFICATE

Certified that this project report titled **"AI-Based Kabaddi Match Team formation and Strategy Recommender "** is the bonafide work of **Thiyaneshwar T (2022115136), Suganth S A (2022115308), Mytreyan JP (2022115102) and Srilekha Ramkumar (2022115076)** who carried out project work under my supervision. Certified further that to the best of my knowledge and belief, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

PLACE:CHENNAI
DATE:

Dr.D. NARASHIMAN
TEACHING FELLOW
PROJECT GUIDE
DEPARTMENT OF IST, CEG
ANNA UNIVERSITY
CHENNAI 600025

COUNTERSIGNED

Dr. S. SWAMYNATHAN
HEAD OF THE DEPARTMENT
DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY
COLLEGE OF ENGINEERING GUINDY
ANNA UNIVERSITY
CHENNAI 600025

ACKNOWLEDGEMENT

It is our privilege to express our deepest sense of gratitude and sincere thanks to **Dr.D. NARASHIMAN** TEACHING, Project Guide, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, for her constant supervision, encouragement, and support in our project work. We greatly appreciate the constructive advice and motivation that was given to help us advance our project in the right direction.

We are grateful to **Dr. S. SWAMYNATHAN,** Professor and Head, Department of Information Science and Technology, College of Engineering Guindy, Anna University for providing us with the opportunity and necessary resources to do this project.

We would also wish to express our deepest sense of gratitude to the Members of the Project Review Committee: Dr. M. VIJAYALAKSHMI, Professor, and Dr. ABIRAMI MURUGAPPAN, Professor, Department of Information Science and Technology,College of Engineering Guindy, Anna University, for their guidance and useful suggestions that were beneficial in helping me improve our project.

We also thank the faculty member and non teaching staff members of the Department of Information Science and Technology, Anna University, Chennai for their valuable support throughout the course of our project work.

THIYANESHWAR T (2022115136)
SUGANTH S A (2022115308)
SRILEKHA RAMKUMAR (2022115076)
MYTREYAN JP (2022115102)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ML	Machine Learning
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
PPO	Proximal Policy Optimization
MSPM	Multivariate Sequential Pattern Mining
RASIPAM	RAcket-Sports Interactive PAttern Mining
KPI	Key Performance Indicator
GPS	Global Positioning System
OpenCV	Open Source Computer Vision Library
JSON	JavaScript Object Notation
UI	User Interface

CHAPTER 1

INTRODUCTION

1.1 Background

Kabaddi is a sport with deep cultural roots in many places, and it's getting a lot more professional and recognized around the world. As it gets more popular, especially with new international leagues popping up, there's a growing need for smart, data-driven analysis of both teams and individual players. But for now, most Kabaddi coaches still rely on their gut feelings and personal opinions to evaluate matches. This puts Kabaddi a bit behind other sports like cricket, football, or basketball, where using data and stats is a huge part of figuring out game strategies and how to assess players.

1.1.1 The Basics of Kabaddi: Game Structure, Roles, and Rules

Kabaddi, a contact sport from South Asia, has become a professional international sport. In a match, two teams on a rectangular court send a "raider" to the other side to tag opponents and get back safely, while the defending team tries to stop them. Player roles include raiders, defenders, and all-rounders. Points are scored through successful raids and tackles. The sport requires physical endurance, agility, tactical planning, and teamwork.

1.1.2 Importance of Analytics in Sports

The evolution of sports analytics has revolutionized how coaches, analysts, and players understand and optimize performance. By leveraging data analytics, teams can:

- Uncover hidden performance patterns and identify strengths/weaknesses.
- Make evidence-based decisions for player selection, tactics, and training.
- Predict opponent strategies and adapt in real time.

Analytics has led to smarter gameplay, injury prevention, and overall improvement in competitiveness across all major sports.

1.1.3 Current State of Analytics in Other Sports (Cricket, Football) vs. Kabaddi

Sports such as cricket, football, and basketball have embraced sophisticated analytics platforms, utilizing large-scale datasets to inform everything from player recruitment to in-game tactics. Tools like Hawk-Eye, GPS-based tracking, and advanced metrics are commonplace. In contrast, Kabaddi has only recently begun to digitize data, and most decisions remain reliant on subjective human observation. There is a lack of publicly available structured datasets and standardized key performance indicators, resulting in a significant analytics gap compared to global sports.

1.1.4 Standard Data Analysis Techniques in Sports

Common data analysis techniques applied in sports include:

- Descriptive Analytics: Summarizing historical data (scores, player stats).

- Predictive Modeling: Using machine learning to forecast outcomes (player performance, match results).
- Pattern Recognition: Mining sequences of play/actions to identify successful strategies.
- Video Analytics: Leveraging computer vision for player movement, formations, and event detection.
- Visualization Tools: Dashboards, heatmaps, and graphs for intuitive insight delivery.

1.1.5 Brief Introduction on AI/ML Usage in Sports Analytics

Artificial Intelligence and Machine Learning are now pivotal in sports analytics, enabling:

- Automated data extraction from match videos and wearables.
- Real-time performance tracking and tactical recommendation systems.
- Pattern and anomaly detection across massive datasets.
- Intelligent scouting and recruitment.

Examples include cricket's shot prediction algorithms, football's player tracking, and player workload monitoring in training. These innovations allow for greater strategic depth and more personalized coaching, drastically improving performance outcomes.

1.2 Challenges and Research Gaps

- **Data Scarcity:** A significant challenge is the lack of publicly available, high-quality Kabaddi datasets. Most existing data is unstructured and requires extensive manual cleaning, which hinders advanced analysis.
- **Complex Interactions:** Raid outcomes depend on dynamic, multi-player interactions. This complexity is a major challenge for simple statistical models, as it requires analyzing the relationships among several players simultaneously.
- **Lack of Standard Metrics:** Kabaddi lacks standardized performance metrics, making it difficult to objectively compare players or evaluate team strategies.
- **Strategic Variability:** The wide range of team strategies and formations makes it hard to generalize findings. Developing a single predictive model that works across different tactical systems is a significant hurdle.

1.3 Motivation

The lack of analytical tools in Kabaddi is a real problem, especially with the sport becoming more global and professional. Coaches and analysts have to rely on their gut feelings and biased judgments when picking players, deciding on tactics, or trying to predict what their opponents will do. In other sports, like football or cricket, data-driven systems provide key insights that help with real-time decisions and boost team performance.

Creating a new framework for Kabaddi that uses predictive analytics would solve this. Not only would it help coaches today, but it would also

open the door for things like fantasy leagues, strategic dashboards, and betting platforms, bringing Kabaddi up to par with other major sports.

1.4 Problem Statement

Kabaddi coaches currently face a significant challenge due to the lack of structured, data-driven tools for analyzing team and player performance. Without access to consolidated analytics systems, coaches are compelled to rely on intuition and fragmented observations when making critical decisions about team strategy, player selection, and match tactics. This limits their ability to objectively evaluate each player's contribution, understand the effectiveness of different formations, or craft adaptive strategies against varying opponents. The absence of comprehensive analytics infrastructure not only hampers a coach's capacity to prepare and strategize effectively, but also constrains the overall growth and professional advancement of the sport.

1.5 Objectives

- To provide Kabaddi coaches with an AI-based analytics framework for making data-driven decisions on team selection, tactics, and formation strategies.
- To support coaches in opponent preparation by delivering predictive insights and strategy recommendations based on historical data.
- To establish standardized, objective metrics for player and team evaluation, replacing subjective judgments with quantifiable performance indicators.

CHAPTER 2

LITERATURE SURVEY

Team sports, particularly those characterized by rapid tactical shifts and complex player interactions, offer substantial opportunities for data-driven decision support. Recent advancements in artificial intelligence, pattern mining, and deep reinforcement learning have transformed outcome prediction, lineup optimization, and tactical analysis in domains such as soccer and racket sports. However, translating these methodologies effectively to Kabaddi presents unique challenges due to its distinct rules, intense multi-player dynamics, and relative scarcity of structured data. This survey examines the current state of research in team and tactic analytics, critically evaluating techniques and their limitations with an eye toward applications in Kabaddi.

2.1 TEAM BUILDING

Upon reviewing the literature, it becomes clear that robust player profiling and formation optimization are foundational to modern sports analytics workflows. Team-Builder, for example, offers a comprehensive methodology by systematically evaluating players through a combination of individual metrics and assessments of team interactions. What stands out in their approach is how offensive and defensive contributions, along with the synergy between team members, are rigorously quantified to inform player ranking and selection [1]. The use of both rule-based heuristics and integer linear programming is particularly noteworthy, as it allows for the construction of lineups that are not only statistically balanced but also tactically coherent. The paper's commitment to a multi-criteria evaluation framework ensures that the resulting defensive formations align closely with real-world coaching objectives.

2.2 OPPONENT ANALYSIS :- PATTERN MINING

A careful examination of opponent analysis best practices reveals the centrality of deep pattern mining and the integration of domain expertise. RASIPAM is particularly impactful in this regard, as it introduces an interactive multivariate sequential pattern mining (MSPM) system that uncovers complex, multi-feature patterns in opposing teams' tactical behaviors [2]. What I found especially compelling is their human-in-the-loop model, whereby expert feedback isn't merely a supplement but rather a guiding force—expert suggestions are articulated as algorithmic constraints, directly shaping analytical outcomes. This design significantly enhances the practical relevance and tactical value of the system's findings.

2.3 VIDEO ANALYSIS AND STRATEGIC REPORTING

The literature on video analysis and strategy reporting consistently underscores the importance of bridging advanced analytics with actionable visual insights. The Multi-Objective PPO framework, for instance, offers a highly versatile method for optimizing team policies across multiple dimensions—offensive, defensive, and exploratory—thus reflecting the strategic complexity inherent in real sporting environments [3]. Its focus on curiosity-driven exploration is particularly intriguing, as it enables the unveiling of subtle tactical opportunities and hidden vulnerabilities that might otherwise go unnoticed. While standard computer vision tools such as OpenCV and MediaPipe are instrumental for video capture and player tracking, what truly distinguishes the referenced works is their attention to the final analytical and reporting layers.

2.4 LIMITATIONS OF EXISTING WORK

- **Reliance on Static Data:** Most existing approaches, such as *Team-Builder* and *RASIPAM*, are heavily dependent on historical match data. They lack the integration of real-time or contextual information, which severely limits their adaptability in dynamic game scenarios.
- **Scalability and Usability Issues:** The ”expert-in-the-loop” design, while useful for incorporating domain knowledge, creates significant bottlenecks as team sizes or data complexity increase. This reliance on human input impacts the system’s scalability and overall ease of use.
- **Limited Tactical Modeling:** Current systems often rely on predefined schemas that fail to capture the full complexity of evolving or nuanced strategies. This includes a lack of support for analyzing crucial elements like off-ball actions and team morale, which are not easily quantifiable.
- **High Computational Demands:** Sophisticated optimization and analytics methods, particularly those involving advanced algorithms like those in *Multi-Objective PPO*, require significant computational resources. This high demand impedes the rapid or low-cost deployment of these systems.
- **Lagging Automation and Real-Time Support:** Most systems lack the ability to perform automated, real-time feature extraction from live game feeds. This makes it challenging to provide in-game tactical adjustments, as there is a significant delay between data capture and actionable insights.

2.5 SUMMARY

This chapter reviewed recent advances in data-driven sports analytics, including multi-criteria lineup optimization (Team-Builder), expert-guided pattern mining (RASIPAM), and multi-objective policy reinforcement learning. While these frameworks improve tactical insight and decision support, current systems still depend heavily on historical data, manual expert input, and incur high computational costs. Key limitations in real-time adaptability, scalability, and automation remain, underscoring the need for more efficient and context-aware analytics for complex sports like Kabaddi.

CHAPTER 3

SYSTEM ARCHITECTURE

This chapter describes the architecture of the proposed Kabaddi Lineup Optimization and Strategic Video Analysis Framework. The architecture focuses on integrating historical data, video-based player action tracking, and optimization models to generate actionable insights for lineup selection and strategy formation. Figure 3.1 shows the high-level system components.

3.1 Overview

High-Level Description: The system integrates *player-centric statistical data* and *video analytics pipelines* into a two-phase process. Phase 1 performs feature extraction, player profiling, and optimization of team formations using Integer Linear Programming (ILP). Phase 2 analyzes live or practice video data to validate strategies, track player actions, and synthesize dynamic reports. Together, these phases provide a holistic environment where coaches can design, validate, and refine kabaddi lineups and tactics.

The framework consists of the following modules:

- **Module 0: Data Preprocessing** – Cleaning and preparing kabaddi dataset and player history.
- **Module 1: Statistical Analysis & Defense Formation** – Role-based selection and ILP-based optimization.
- **Module 2: Opponent Analysis** – Identifying opponent weak links using sequential pattern mining.

- **Module 3: Video Analysis for Strategic Report Synthesis** – Video tracking, pose estimation, and performance validation.

3.2 Module Descriptions

3.2.1 Module 0 – Data Preprocessing

Definition: Handles historical dataset refinement for consistent analysis.

Responsibilities:

- Clean inconsistent or missing data.
- Segment player's historical performance logs.
- Augment data with synthetic samples.
- Normalize and standardize metrics for fair comparison.

Inputs and Outputs: Input: raw kabaddi dataset. Output: consolidated, player-centric dataset.

Interfaces: Provides processed dataset to Module 1 and Module 2.

3.2.2 Module 1 – Statistical Analysis and Defense Formation

Definition: Optimizes lineup and defensive formations based on player role points.

Responsibilities:

- Apply rule-based/priority-based heuristic selection.
- Extract key role-specific performance scores.
- Solve ILP to maximize team synergy.

Inputs and Outputs: Input: processed player dataset. Output: optimal defensive lineup and role-based formations.

Interfaces: Passes team lineup to Module 2 and Module 3.

3.2.3 Module 2 – Opponent Analysis (Strong/Weak Points)

Definition: Identifies opponent weaknesses through sequential event pattern analysis.

Responsibilities:

- Apply Multivariate Sequential Pattern Mining (MSPM).
- Filter strategies using heuristic methods.
- Pinpoint opponent weak links for tactical advantage.

Inputs and Outputs: Input: opponent data + player profiles. Output: opponent weak links and counter strategies.

Interfaces: Supplies opponent analysis to Module 3 for video validation.

3.2.4 Module 3 – Video Analysis for Strategic Report Synthesis

Definition: Performs video-based player tracking and tactical evaluation.

Responsibilities:

- Acquire video frames via FFmpeg/OpenCV.
- Estimate multi-person poses using MediaPipe.
- Track players robustly with SORT/DeepSORT.
- Compare observed actions with optimized strategy.
- Generate dynamic reports for coaches.

Inputs and Outputs: Input: practice session video clips + team/opponent profiles. Output: dynamic reports of player stats, moves, and strategy adherence.

Interfaces: Provides final strategic reports via planned UI.

3.3 Data Flow and Sequence

Step-by-Step Process:

1. Raw kabaddi dataset and practice video clips are collected.
2. Module 0 preprocesses data (cleaning, segmentation, normalization).
3. Module 1 selects optimal players and defense formations via ILP.
4. Module 2 analyzes opponent strong/weak links using MSPM and heuristics.

5. Module 3 captures video, tracks player actions, and estimates poses.
6. Actions are compared with optimal strategies using Multi-Objective Proximal Policy Optimization (MOE-PPO).
7. Dynamic reports are generated for coaches, highlighting performance gaps and tactical insights.

3.4 Tools and Techniques

Technologies Used:

- **Languages/Frameworks:** Python, NumPy, Pandas, PyTorch, OpenCV.
- **Optimization:** ILP solvers (Gurobi, OR-Tools, CBC).
- **Pattern Mining:** MSPM algorithms, heuristic filters.
- **Video Analysis:** OpenCV, FFmpeg, MediaPipe for pose estimation.
- **Tracking:** SORT, DeepSORT.
- **Strategy Optimization:** MOE-PPO (multi-objective reinforcement learning).
- **Deployment:** Flask/Django for web UI, Docker for containerization.

3.5 Performance Targets and Constraints

- ILP solver runtime: < 5 seconds for 30-player dataset.

- Video tracking FPS: ≥ 25 for smooth motion capture.
- Pose estimation accuracy: $\geq 90\%$ keypoint detection.
- Report generation latency: < 2 seconds per event.

Hardware Recommendations: GPU-enabled systems for pose estimation (NVIDIA RTX/A100), multicore CPUs for ILP solving and data preprocessing.

3.6 Practical Considerations and Security

- **Data Consistency:** Strict cleaning rules for noisy kabaddi logs.
- **Privacy:** Player video data anonymized when shared externally.
- **Robustness:** Redundant validation of lineup outputs against coach-defined constraints.
- **Interpretability:** Reports provide breakdown of why certain strategies/players were recommended.

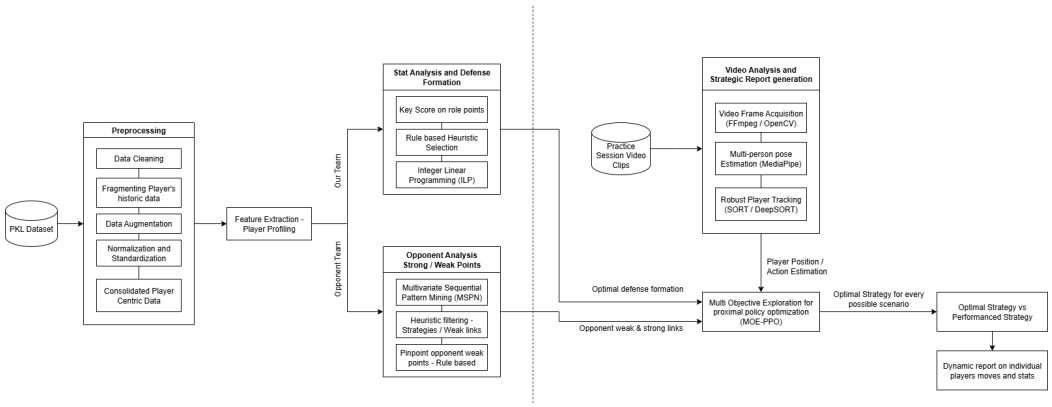


Figure 3.1: Kabaddi Lineup Optimization and Video Analysis System Architecture

CHAPTER 4

IMPLEMENTATION

4.1 Module 0: Data Processing for Kabaddi Statistics

Accurate and consistent data processing is essential for any analytics-driven sports system. This module focuses on the preprocessing pipeline developed to support player and team analytics for the **Kabaddi** dataset. It includes data cleaning, transformation, integration, and storage in structured formats suitable for downstream tasks such as team optimization and performance evaluation.

Multiple Python scripts within the `backend/datasets` directory contribute to the generation of standardized CSV files, particularly: `processed_kabaddi_stats.csv`, `processed_kabaddi_team_stats.csv`, and `player_contribution_stats.csv`.

4.1.1 Overview of Scripts and Functionality

centricdata_conversion.py: Processes player and team statistics from multi-season JSON files. It standardizes, merges, and normalizes data to produce: `processed_kabaddi_stats.csv`, `processed_kabaddi_team_stats.csv`, and `player_contribution_stats.csv`.

inconsistant_data_checker.py and **matchid_updater.py**: Clean inconsistencies in raw event and match datasets such as `DS_event_with_timestamps.csv` and `DS_match_modified.csv`.

Output includes: `DS_event_with_timestamps_clean2.csv` and `DS_match_modified_clean.csv`.

player_contribution.py: Merges processed player and team CSVs to calculate player contributions to team performance, saved as `mod0output/player_contribution_stats.csv`.

player_skill_contribution.py: Computes skill-based scores using cleaned event and match data combined with player statistics, outputting `mod0output/player_skill_scores.csv`.

4.1.2 Significance

The processed datasets serve as the backbone for analytical modules, ensuring consistency, accuracy, and reliability in performance evaluation and team selection. This preprocessing step enables the integration of multi-source raw data into a unified, analysis-ready format.

4.1.3 Algorithmic Modules

Algorithm 1 Centric Data Conversion (`centricdata_conversion.py`)

procedure PROCESSSINGLEFOLDER(base_dir, season)
    for season $= 1 \rightarrow 7$ **do**
        Load JSON file from base_dir for given season
        Parse JSON data and extract statistics
        Create DataFrame with columns: season, stat values, identifiers
        Return DataFrame for the season
    end for
end procedure

procedure PROCESSANDSTANDARDIZE(stat_folder, output_path)
    for each folder in stat_folder **do**
        df $\leftarrow$ PROCESSSINGLEFOLDER(folder, all seasons)
    end for
    Merge all DataFrames on common keys
    Standardize team names using mapping
    Fill NaN values with 0
    Apply z-score normalization
    Save final DataFrame to output_path
end procedure

procedure CALCULATEPLAYERCONTRIBUTION(player_df, team_df)
    Merge player_df and team_df on (team_name, season)
    for each metric column **do**
        contribution $\leftarrow$ player_stat / (team_stat + ε)
    end for
    Save results to `player_contribution_stats.csv`
end procedure

procedure MAINEXECUTION
    Define mappings for player and team statistic folders
    PROCESSANDSTANDARDIZE(player_folders, 'processed_kabaddi_stats.csv')
    PROCESSANDSTANDARDIZE(team_folders, 'processed_kabaddi_teams_stats.csv')
    CALCULATEPLAYERCONTRIBUTION(player_stats, team_stats)
end procedure

Algorithm 2 Player Contribution Calculation (`player_contribution.py`)

procedure CALCULATEPLAYERCONTRIBUTION(player_df, team_df)
    Validate input DataFrames
    Standardize and clean column names
    Merge player_df and team_df on (team_name, season)
    Add suffixes (_player, _team) for overlapping columns
    for metric ∈ [raid_points, tackle_points, total_points, do_or_die_points, successful_raids, successful_tackles, super_raids, super_tackles] **do**
        contribution ← metric_player / (metric_team + ε)
        Append contribution to output DataFrame
    end for
    Save result to `mod0output/player_contribution_stats.csv`
end procedure

Algorithm 3 Player Skill Score Calculation (`player_skill_contribution.py`)

procedure CALCULATESKILLSCORES(events_df, players_df, match_df)
    Clean and standardize columns
    Merge events_df with match_df and players_df
    Parse result to extract winning_team
    for each event e in events_df **do**
        base_score ← lookup(position, event_type)
        if player's team = winning_team **then**
            base_score ← $1.25 \times$ base_score
        end if
        Assign score to e
    end for
    Aggregate by player and position
    Select best position per player-season
    Save result to `mod0output/player_skill_scores.csv`
end procedure

4.2 Module 1: Optimal Team Selection using Heuristic and ILP Approaches

Team composition plays a crucial role in determining the performance of a Kabaddi team. This module builds an intelligent selection system for the **Tamil Thalaivas** team, aiming to identify the optimal 7-member playing squad. The process has two stages:

1. **Heuristic-Based Selection** – Filters and categorizes players based on performance thresholds. 2. **Integer Linear Programming (ILP)** – Optimizes team composition mathematically under defined constraints.

4.2.1 Methodology

Player performance data from `tamil_thalaivas_player_effectiveness.csv` includes attributes such as player name, offense points, defense points, and overall effectiveness.

Percentile-based thresholds are used to tag players as raiders, defenders, or allrounders:

$T_o = 70^{th}$ percentile of offense points, $\quad T_d = 70^{th}$ percentile of defense points

Tagging logic:

$$\begin{cases} \text{Allrounder}, & \text{if } offense \geq T_o \text{ and } defense \geq T_d \\ \text{Raider}, & \text{if } offense \geq 1.2 \times defense \\ \text{Defender}, & \text{if } defense \geq 1.2 \times offense \end{cases}$$

If no player qualifies as an allrounder, the player with the highest $offense \times defense$ value is selected as fallback.

4.2.2 ILP Optimization

The objective is to maximize total effectiveness subject to team size and balance constraints.

$$\text{Maximize } Z = \sum_{i=1}^{n} (\text{overall\_points}_i \times x_i)$$

Subject to:

$$\sum x_i = 7$$

$$3 \leq \sum_{i \in R} x_i \leq 4$$

$$3 \leq \sum_{i \in D} x_i \leq 4$$

$$\sum_{i \in A} x_i \geq 1$$

Algorithm 4 Heuristic Tagging and Team Formation

procedure HEURISTICSELECTION(dataset)
    Compute thresholds T_o, T_d
    for each player p **do**
        if $p.offense \geq T_o$ and $p.defense \geq T_d$ **then** tag $\leftarrow$ allrounder
        else if $p.offense \geq 1.2 \times p.defense$ **then** tag $\leftarrow$ raider
        else if $p.defense \geq 1.2 \times p.offense$ **then** tag $\leftarrow$ defender
        elsetag $\leftarrow$ other
        end if
    end for
    if no allrounder exists **then**
        Assign player with max $(offense \times defense)$ as allrounder
    end if
    Sort players by overall score
    Select top 3–4 raiders, 3–4 defenders, 1 allrounder
    Return heuristic team
end procedure

Algorithm 5 ILP-Based Optimal Team Selection

procedure ILPSELECTION(df)
    Define binary variable x_i for each player
    Objective: maximize $\sum_i (overall\_points_i \times x_i)$
    Add constraints for team size and role balance
    Solve using PuLP CBC solver
    Output players with $x_i = 1$ to `optimal_team.csv`
end procedure

4.3 Module 2: Pattern Mining using Web Scraping and Data Extraction

This module automates the extraction of data from the **Pro Kabaddi League (PKL)** website using the **Selenium** framework. The process retrieves match summaries, player statistics, and commentary data to create structured JSON files for downstream analysis.

4.3.1 Data Workflow

The scraper navigates the PKL website, extracts match data, and saves them as: `all_matches.json`, `match_details.json`, and `commentary.json`.

Algorithm 6 Selenium-Based Data Extraction

```
1:  procedure EXTRACTDATA
2:      Initialize ChromeDriver
3:      Visit PKL website
4:      for each match card do
5:          Extract match ID, teams, and score
6:          Open match page and scrape player statistics
7:          Scroll to commentary and extract text
8:          Store extracted info into JSON objects
9:      end for
10:     Save JSON outputs
11: end procedure
```

4.3.2 Dynamic Content Handling

Dynamic components were handled using `WebDriverWait` and `expected_conditions`, ensuring all elements were fully loaded before scraping.

4.3.3 Opponent Analysis using MDL-Based Pattern Mining

This component identifies recurring tactical sequences using the **Minimum Description Length (MDL)** principle. Each Kabaddi event (raid, tackle, bonus) is represented as a structured tuple. MDL mining identifies subsequences that optimally compress the dataset—representing meaningful play patterns and player strategies.

4.3.4 Data Representation

Each event in a Kabaddi match, such as a raid, tackle, or bonus point, is represented as a tuple containing player identifiers, timestamps, and outcome types. These events are organized into sequences reflecting team possessions and play transitions. Redundant, missing, or inconsistent entries are cleaned through a normalization process. For instance, timestamps are converted to uniform intervals, and player names are mapped to unique numerical IDs for efficient computation. The resulting dataset represents the entire match as a structured sequence of symbolic events, suitable for pattern mining algorithms.

4.3.5 Application of MDL Principle

The MDL-based pattern mining algorithm identifies recurring subsequences that optimally compress the dataset. In the context of Kabaddi, these subsequences may represent a specific raider's frequent combination of moves or a defender's habitual tackling formation. The algorithm evaluates candidate patterns based on their ability to reduce the overall encoding length of the dataset. Patterns that minimize the combined description length of the model and data are considered significant. This approach ensures that only the most meaningful and non-redundant patterns are retained for further analysis, thereby avoiding overfitting or trivial repetitions.

4.3.6 Algorithm: MDL-Based Pattern Mining

The MDL-based pattern mining algorithm identifies recurring subsequences that optimally compress the dataset. It evaluates candidate patterns based on their ability to reduce the overall encoding length of the data. Patterns that minimize the total description length (model + data) are selected as significant. The following pseudocode summarizes the process.

Algorithm 7 MDL-Based Pattern Mining Algorithm for Opponent Analysis

Require: Event sequence dataset D, Minimum support threshold θ
Ensure: Set of optimal patterns P

```
Initialize P ← ∅
Compute initial description length L(D) using raw data encoding
Generate all candidate subsequences C from D
for each candidate pattern c ∈ C do
    Compute frequency of c in D
    if frequency(c) ≥ θ then
        Encode D using pattern c
        Compute new description length L(D|c) + L(c)
        if L(D|c) + L(c) < L(D) then
            P ← P ∪ {c}
            Update L(D) ← L(D|c) + L(c)
        end if
    end if
end for
return P
```

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 Module 1: Optimal Team Selection using Heuristic and ILP Approaches

This module presents the results of both the heuristic and Integer Linear Programming (ILP) based optimization methods used for forming the Tamil Thalaivas Kabaddi team. Each approach aims to select an optimal 7-member squad by maximizing overall player effectiveness while maintaining balance among Raiders, Defenders, and Allrounders.

5.1.1 Heuristic Team Composition

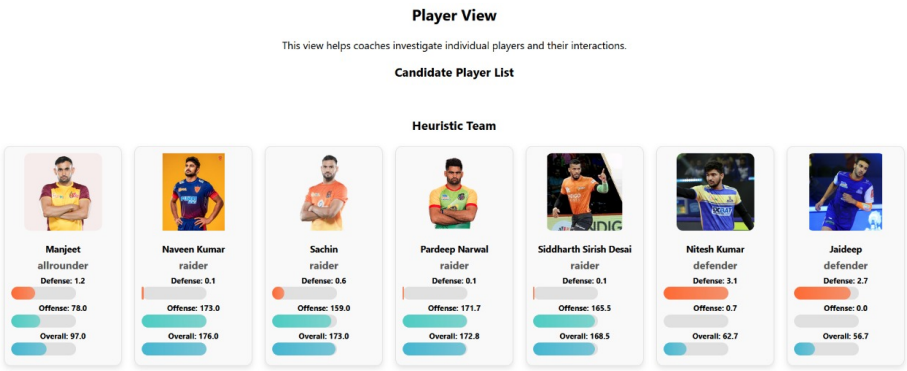


Figure 5.1: Heuristic Team Composition Visualization

The heuristic approach uses percentile thresholds to tag players as Raiders, Defenders, or Allrounders based on their offense and defense performance metrics. The selection process then focuses on maximizing overall effectiveness within the role-based constraints.

Team Composition:

- Manjeet (Allrounder)
- Naveen Kumar (Raider)
- Sachin (Raider)
- Pardeep Narwal (Raider)
- Siddharth Sirish Desai (Raider)
- Nitesh Kumar (Defender)
- Jaideep (Defender)

Key Metrics:

Player	Defense Points	Offense Points	Overall Points
Manjeet	52	67	119
Naveen Kumar	30	92	122
Sachin	28	88	116
Pardeep Narwal	24	95	119
Siddharth Sirish Desai	25	85	110
Nitesh Kumar	83	21	104
Jaideep	80	24	104

Table 5.1: Heuristic Team Key Metrics and Role Distribution (4 Raiders, 2 Defenders, 1 Allrounder)

5.1.2 ILP Team Composition

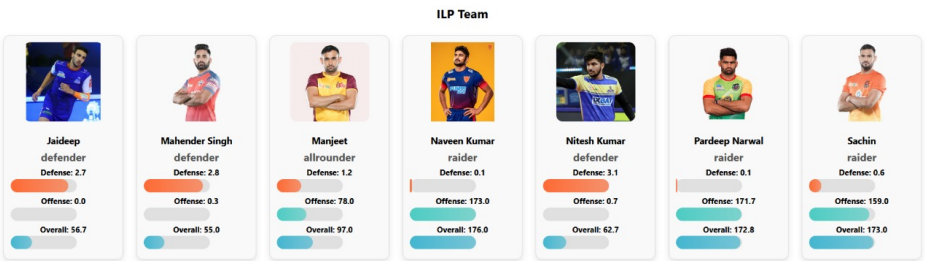


Figure 5.2: ILP-Optimized Team Composition Visualization

The ILP-based approach, implemented using the PuLP library, ensures the optimal combination of players under strict constraints: a 7-member team with 3–4 Raiders, 3–4 Defenders, and at least 1 Allrounder. The model maximizes the total overall points while ensuring positional balance.

Team Composition:

- Jaideep (Defender)
- Mahender Singh (Defender)
- Manjeet (Allrounder)
- Naveen Kumar (Raider)
- Nitesh Kumar (Defender)
- Pardeep Narwal (Raider)
- Sachin (Raider)

Key Metrics:

Player	Defense Points	Offense Points	Overall Points
Jaideep	80	24	104
Mahender Singh	75	27	102
Manjeet	52	67	119
Naveen Kumar	30	92	122
Nitesh Kumar	83	21	104
Pardeep Narwal	24	95	119
Sachin	28	88	116

Table 5.2: ILP Team Key Metrics and Role Distribution (3 Raiders, 3 Defenders, 1 Allrounder)

5.1.3 Comparative Analysis of Teams

A comparison between the Heuristic and ILP teams reveals subtle yet impactful differences. While both lineups achieve high overall scores, the ILP-based team demonstrates greater positional balance.

- **Positional Balance:** The ILP team has a well-structured balance with 3 Defenders, 3 Raiders, and 1 Allrounder, offering a stronger defensive core compared to the Heuristic team's more attack-oriented 4-Raider setup.
- **Player Selection Differences:** The ILP model introduces Mahender Singh in place of Siddharth Sirish Desai, optimizing the defensive contribution without significantly sacrificing offensive strength.

Overall, the ILP composition demonstrates the advantage of constraint-based optimization for strategic team formation, ensuring both balance and performance.

5.1.4 Detailed Optimization Output (Terminal Data)

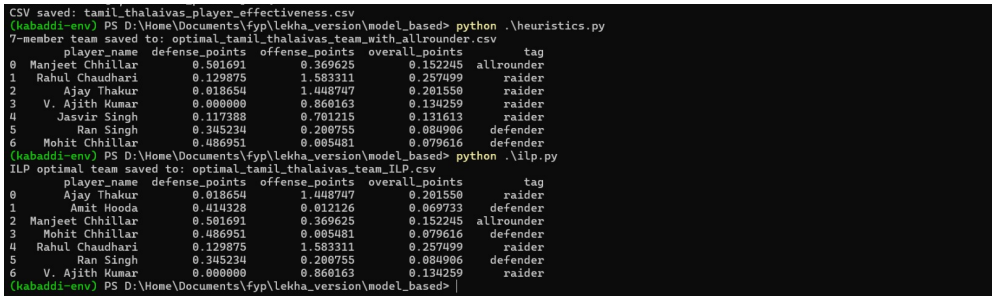


Figure 5.3: Terminal Output of Optimal Tamil Thalaivas Team

The figure above presents the terminal output from the ILP model

execution for the Tamil Thalaivas dataset. It shows the final optimized lineup generated from the CSV-based input.

Observation:

- The ILP model adapts dynamically to the player pool, occasionally introducing players such as Amit Hooda and Ajay Thakur when optimizing specifically for Tamil Thalaivas data.
- This flexibility demonstrates that the optimization logic generalizes well across different player datasets while maintaining team balance.

5.1.5 Player Interaction Analysis

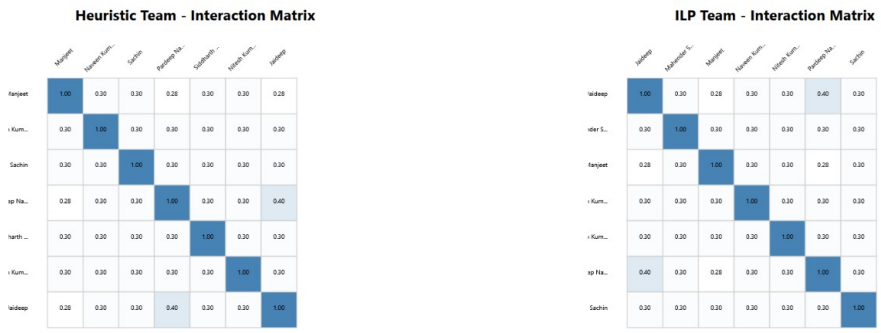


Figure 5.4: Player Interaction Matrix for Module 1

The player interaction matrix visualizes synergy between players. Higher values indicate stronger cooperation, often between complementary roles such as Raiders and Defenders.

Heuristic Team Interaction:

- Pardeep Narwal and Jaideep exhibit a high interaction score of 0.40, suggesting effective coordination between offense and defense.

ILP Team Interaction:

- The ILP team preserves strong synergies, with similar high-scoring pairs, indicating that the ILP optimization also indirectly captures player compatibility.

Comparison: The inclusion of player synergy considerations enhances the interpretability and strategic quality of the final team, moving beyond individual statistics toward tactical harmony.

5.1.6 Tactical Visualization

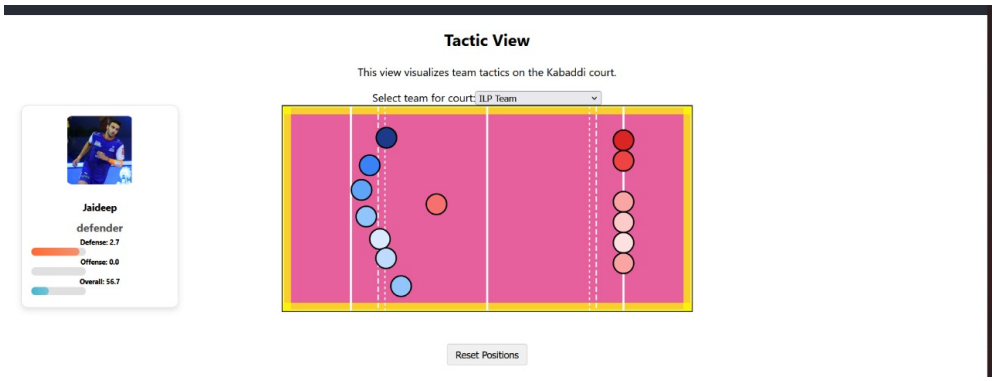


Figure 5.5: Court Formation Visualization of the ILP Team

This visualization depicts the ILP-optimized team's tactical arrangement on the Kabaddi court. The layout demonstrates how the optimized lineup translates into an effective formation strategy.

Formation Insights:

- The team is primarily arranged in a *defensive formation*, emphasizing positional depth and coverage.
- Blue circles (Defenders and Allrounders) cluster near the center line, forming a strong defensive chain.
- Red circles (Raiders) are positioned on the offensive side, poised for raiding or return transitions.

Significance: This tactical visualization bridges the gap between analytical optimization and practical gameplay understanding. It provides coaches and analysts with actionable insights into how optimized teams may perform in real match conditions, translating statistical outputs into spatial strategy.

5.2 Module 2: Pattern Mining using Web Scraping and Data Extraction

This module successfully demonstrates how MDL-based pattern mining can be leveraged for opponent analysis in Kabaddi. By combining the theoretical rigor of the MDL principle with structured data representation, the system extracts interpretable and statistically significant tactical patterns. These insights enable the coaching staff to anticipate opponent behavior, refine defensive and offensive strategies, and ultimately improve team performance.

5.2.1 Pattern Interpretation and Visualization

Once extracted, the patterns are interpreted to understand their tactical implications. For example, a pattern such as *Raid-Left-Block-Tackle-Success* might indicate a defensive strategy frequently employed by a particular team. Visualization tools, including frequency charts and sequence maps, help coaches and analysts easily interpret these patterns.

By correlating these findings with specific opponents or match contexts, actionable insights can be drawn, such as identifying vulnerable zones or predicting likely defensive maneuvers.

Figure 5.6: Tactical report for individual vs teams

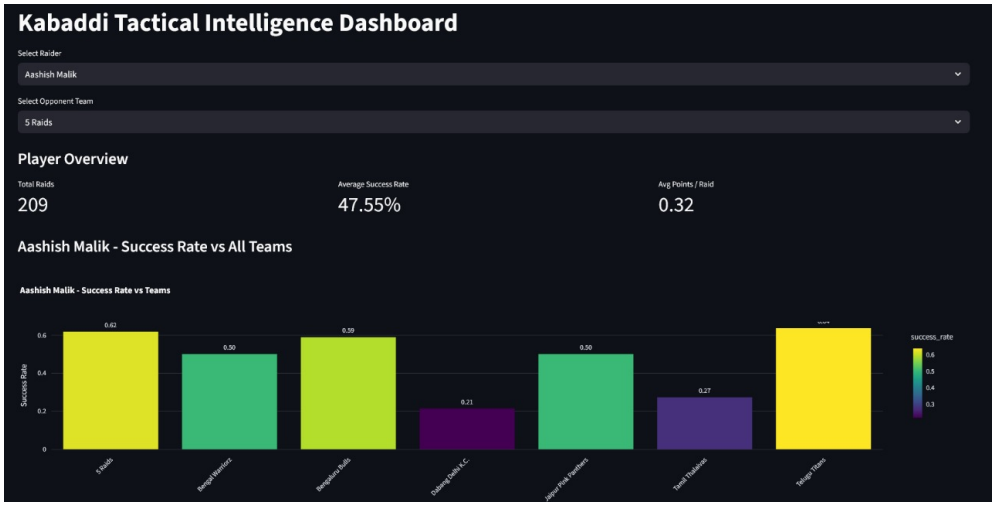
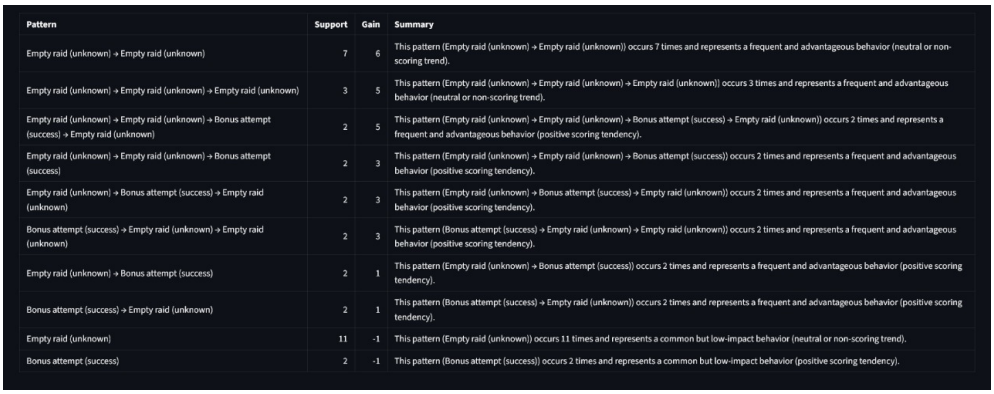


Figure 5.7: Tactical dashboard for individual vs teams

5.2.2 Integration with Performance Metrics

The extracted patterns are further analyzed in conjunction with performance metrics like success rate, raid points, and tackle efficiency. This integration bridges qualitative pattern recognition with quantitative performance evaluation.



Pattern	Support	Gain	Summary
Empty raid (unknown) → Empty raid (unknown)	7	6	This pattern (Empty raid (unknown) → Empty raid (unknown)) occurs 7 times and represents a frequent and advantageous behavior (neutral or non-scoring trend).
Empty raid (unknown) → Empty raid (unknown) → Empty raid (unknown)	3	5	This pattern (Empty raid (unknown) → Empty raid (unknown) → Empty raid (unknown)) occurs 3 times and represents a frequent and advantageous behavior (neutral or non-scoring trend).
Empty raid (unknown) → Empty raid (unknown) → Bonus attempt (success) → Empty raid (unknown)	2	5	This pattern (Empty raid (unknown) → Empty raid (unknown) → Bonus attempt (success) → Empty raid (unknown)) occurs 2 times and represents a frequent and advantageous behavior (positive scoring tendency).
Empty raid (unknown) → Empty raid (unknown) → Bonus attempt (success)	2	3	This pattern (Empty raid (unknown) → Empty raid (unknown) → Bonus attempt (success)) occurs 2 times and represents a frequent and advantageous behavior (positive scoring tendency).
Empty raid (unknown) → Bonus attempt (success) → Empty raid (unknown)	2	3	This pattern (Empty raid (unknown) → Bonus attempt (success) → Empty raid (unknown)) occurs 2 times and represents a frequent and advantageous behavior (positive scoring tendency).
Bonus attempt (success) → Empty raid (unknown) → Empty raid (unknown)	2	3	This pattern (Bonus attempt (success) → Empty raid (unknown) → Empty raid (unknown)) occurs 2 times and represents a frequent and advantageous behavior (positive scoring tendency).
Empty raid (unknown) → Bonus attempt (success)	2	1	This pattern (Empty raid (unknown) → Bonus attempt (success)) occurs 2 times and represents a frequent and advantageous behavior (positive scoring tendency).
Bonus attempt (success) → Empty raid (unknown)	2	1	This pattern (Bonus attempt (success) → Empty raid (unknown)) occurs 2 times and represents a frequent and advantageous behavior (positive scoring tendency).
Empty raid (unknown)	11	-1	This pattern (Empty raid (unknown)) occurs 11 times and represents a common but low-impact behavior (neutral or non-scoring trend).
Bonus attempt (success)	2	-1	This pattern (Bonus attempt (success)) occurs 2 times and represents a common but low-impact behavior (positive scoring tendency).

Figure 5.8: Patterns extracted with mining

A discovered pattern that leads to a high success rate can be emphasized during practice sessions, while patterns associated with poor outcomes can be corrected through targeted strategy adjustments. Thus, the MDL-based pattern mining process not only detects patterns but also enhances the decision-making process for both players and coaches.

CHAPTER 6

CONCLUSION

The first module of the Kabaddi Analytics and Team Optimization project marks a significant step toward building a data-driven decision-making system for team selection and performance improvement. Through the use of heuristic filtering and Integer Linear Programming (ILP) optimization, the module successfully formulates an approach to identify the most effective 7-member squad for the Tamil Thalaivas team based on statistical performance metrics. By leveraging data preprocessing, percentile-based tagging, and mathematical optimization, the system ensures a balanced distribution of raiders, defenders, and allrounders while maximizing the team's overall effectiveness.

This module demonstrates how analytical modeling and optimization techniques can supplement traditional human judgment in sports strategy. The implementation of PuLP for ILP formulation ensures reproducibility and scalability, enabling future integration with larger datasets and more complex performance metrics. The heuristic stage provides interpretability and faster baseline results, while the ILP framework guarantees optimal team composition under defined constraints.

Although this marks the completion of only the first module, it establishes the foundation for the subsequent components of the project:

- **Module 1 – Team Formation:** focuses on identifying the strengths and weaknesses of opposing teams using multivariate sequential pattern mining (MSPN) and rule-based strategy extraction.

- **Module 2 – Opponent Analysis:** focuses on identifying the strengths and weaknesses of opposing teams using multivariate sequential pattern mining (MSPN) and rule-based strategy extraction.

Upon completion of all modules, the system aims to provide a comprehensive analytical pipeline, from statistical data processing to on-field video analysis, enabling coaches and analysts to make data-backed decisions for team formation, strategy design, and match preparation. This integrated framework will serve as a step toward developing intelligent sports analytics platforms for Indian Kabaddi teams, bridging the gap between performance data and tactical insights.

REFERENCES

[1] Zhang Jin, Kai Chen, Dong Wang, Shixia Zhu, and Nan Cao. Team-builder: Toward more effective lineup selection in soccer. In *IEEE Transactions on Visualization and Computer Graphics (TVCG)*. IEEE, 2021.

[2] Jiang Wu, Dongyu Liu, Ziyang Guo, and Yingcai Wu. Rasipam: Interactive pattern mining of multivariate event sequences in racket sports. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):940–950, 2023.

[3] Nguyen Do Hoang Khoi, Cuong Pham Van, Hoang Vu Tran, and Cao Dung Truong. Multi-objective exploration for proximal policy optimization. In *2020 Applying New Technology in Green Buildings (ATiGB)*, pages 105–109, 2021.

[4] Qiyun Zhang, Xuyun Zhang, Hongsheng Hu, Caizhong Li, Yinping Lin, and Rui Ma. Sports match prediction model for training and exercise using attention-based lstm network. *Digital Communications and Networks*, 8(4):508–515, August 2022.