

Name: Lokeshkannan U

Roll number: 713322AD058

Dept: AiDS

Task number:4

Task: Data Science Lifecycle Example

I have chosen sports from task 1

1.Problem Definition

Objective: The goal is to improve player performance and overall team success by leveraging data analytics.

Tasks:

- Identify specific performance metrics to track (e.g., distance covered, passes completed, shots on target).
- Set clear goals (e.g., increase passing accuracy by 10%).

2. Data Collection

Objective: Gather relevant data from various sources.

Tasks:

- Collect data from match statistics, wearable devices, training sessions, and player health records.
- Ensure data is gathered in a consistent and accurate manner.

3. Data Cleaning

Objective: Prepare the data for analysis by removing inaccuracies and inconsistencies.

Tasks:

- Handle missing values, outliers, and data duplication.
- Standardize data formats and units.

4. Data Exploration and Analysis

Objective: Understand the data, identify patterns, and generate insights.

Tasks:

- Perform exploratory data analysis (EDA) to summarize main characteristics of the data.
- Use statistical methods to identify relationships between performance metrics and outcomes (e.g., correlation between fitness levels and match performance).

5. Feature Engineering

Objective: Create new features or variables that can improve model performance.

Tasks:

Derive new metrics (e.g., calculating player's work rate by combining distance covered and high-intensity runs).

- Select relevant features for the model based on their predictive power.

6. Model Building

Objective: Develop machine learning models to predict and optimize player performance.

Tasks:

- Choose appropriate algorithms (e.g., linear regression, decision trees, neural networks).
- Train the models using historical data and evaluate their performance.

7. Model Evaluation

Objective: Assess the accuracy and effectiveness of the models.

Tasks:

- Use metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE) to evaluate model accuracy.
- Perform cross-validation to ensure model generalization.

8. Model Deployment

Objective: Implement the models into production systems for real-time use.

Tasks:

- Integrate the models into the team's analytics platform or dashboard.
- Ensure models are continuously updated with new data and monitored for performance.

9. Monitoring and Maintenance

Objective: Continuously monitor and maintain the models to ensure they remain accurate and effective.

Tasks:

- Track model performance over time and make necessary adjustments.
- Address any issues or anomalies that arise in real-time data.

Real-World Scenario

Imagine a football team using this lifecycle to improve their midfielders' passing accuracy. By defining the problem, collecting and analyzing data from matches and training sessions, and building predictive models, the team can identify key factors influencing passing accuracy. They can then implement data-driven strategies to enhance training programs, monitor player performance in real-time, and make informed decisions to optimize team tactics.