



---

## Final Project Abstract

1 message

---

**Omar Akram-1189** <omar\_a11189@cic-cairo.com>

Wed, Dec 13, 2023 at 1:57 AM

To: Asmaa AbuBakr <asmaa\_fahmy@cic-cairo.com>

Cc: Khaled Abdo\_0314 <khaled\_a10314@cic-cairo.com>, hager ali <hager\_m11241@cic-cairo.com>, Abdelrahman Mohamed <abdelrahman\_m11081@cic-cairo.com>

This project presents a comprehensive analysis and application of big data and machine learning techniques to a dataset of FIFA players. The dataset, encompassing a wide range of player attributes, including links to player card images, is subjected to a multi-phase approach to unlock insightful information and predictive capabilities.

### Phase 1: Data Analysis and Cleaning

In the initial phase, the dataset undergoes rigorous analysis and cleaning. This involves identifying and rectifying inconsistencies, missing values, and outliers. Standard data cleaning techniques, along with exploratory data analysis, are employed to ensure data quality and readiness for subsequent phases.

### Phase 2: Data Insertion

The second phase focuses on the efficient insertion and storage of the cleansed data. After evaluating various big data tools, a suitable platform such as Hive or MongoDB. This phase ensures optimal data organization for efficient retrieval and analysis.

### Phase 3: Data Visualization

Utilizing advanced data visualization tools and techniques, this phase aims to represent the data in an intuitive and informative manner. The visualizations aid in uncovering patterns, trends, and correlations within the players' data, providing both a comprehensive overview and detailed insights into specific attributes.

### Phase 4: Machine Learning Analysis

The final and most diverse phase involves applying a range of machine learning algorithms to the dataset. This includes:

Classification Algorithms to categorize players based on specific attributes.

Regression Algorithms for predicting continuous variables like player performance and market value.

Clustering Algorithms to identify natural groupings and patterns within the data.

Gender Detection Algorithm, utilizing CNNs for image-based analysis of player card images, to classify players by gender.

### Group Members:

- Omar Akram 202101189

- Abdelrahman Mohamed 202101081

- Khaled Abdo 202100314

- Hager Magdy 202101241