# Lambton College Mississauga
# BDM 2053 - Big Data Algorithms and Statistic
# Professor. Dhruwal Shah

**Assignment 4**
**(Project Proposal)**

Henry Jones Inbaraj - C0863081
Meenakshi - C0864515
Javier A. Melo – C0871987
Amal Vandananikkal - C0863255
Kelvin Simon - C0866577

**Table of Contents**

# Project Title

# Big Data Analysis for Football Team Formation

# Project objective

- To examine individual performance data for football using big data algorithms.
- To determine the top players for each position based on their performance metrics and data.
- To assemble the most effective squad based on the findings of the analysis.
- To assess the success of the team's composition based on performance information.
- To Implement the data science workflow (Ask, Get, Explore, Model, Share)

# Project Purpose

The Big Data Analysis for Football Team Formation is an exciting initiative that analyses player performance data to determine the top players for each position using big data algorithms. The main goal of the project is to build the best team feasible using player statistics and performance metrics.

The project team will create a model that, using performance measures, can determine who is the best player for each position.

Once the model is created, the project team can use it to choose the best candidate for each role to create the best team possible. The team can then make data-driven choices during a game by using this information to improve their game plan.

The team can create game plans that concentrate on delivering the ball to the striker in the final third of the field, for instance, if they are aware that their striker has a high shot accuracy rate.

As a result of the analysis, the best team will be created using the player statistics, but it will also provide information on how to create the most effective game plans.

# Project Scope

- The team will involve the football players dataset of 2022 to address our objectives.
- To find the most pertinent metrics to use in the analysis, the project team will explore their knowledge in both big data analysis and football.
- The team will use the analysis results to form the best possible team based on player statistics and performance metrics.
- The Analysis will help to form the best game plan.
- This will be best model to use it to form the best team and create a game plan before the game.
- This will also be useful to analyze the team after the game.

# Project Methodology

Project team will use descriptive statistics method to examine FIFA 22 statistics and learn more about the distribution and characteristics of the data. Here are the methods for using descriptive statistics on data from FIFA 22.

- Team will extract the data from Kaggle
- Team will clean the data by removing any missing values or outliers that may skew the results.
- Team will the features that will help our analysis such as mean, median, mode, standard deviation, range, or frequency distribution.
- Team will calculate, interpret, and visualize the key findings.

Project Team will also use the best model to predict the best players for each position with their metrics. Here are the steps to do it.

- Team will Load the data into the data structures like Pandas data frame or NumPy array.
- Team will preprocess the data which involves Data profiling, Data Cleansing, Data reduction, Data transformation, Data enrichment and data Validation.
- Team will use descriptive statistics results and correlation technique to select the metrics to use in the model for prediction.
- Team will split the data into two parts like training data and testing data for the model to predict the best players out of the selected metrics.
- Team will also evaluate and optimize the model for the prediction of best player for each position.

## Risk Assessment

| **Category** | **Risk** | **Mitigation** |
|---|---|---|
| High risk | Loss of data | Back up option\Kaggle website |
| Medium risk | Incorrect algorithm selection | Accuracy and evaluation of the model will be recorded |
| Low risk | Lack of stakeholder engagement | Collaborating stakeholders throughout the project. |

## Project Resources

Project Team will use Jupyter Notebook in Vs-code to execute the methodologies. Project Team will work with Python libraries like Pandas, NumPy, Matplotlib, Scikit-learn and seaborn and team might use other resources for benefit of the project.

| **Resources** | **Usage** |
|---|---|
| Jupyter Notebook in Vs-code | Team's work environment |
| Kaggle | To get the data |
| Github | To collaborate the work |
| Python libraries | Whole Project |
| Pandas | To import and explore the data |
| NumPy | To explore the data |
| Matplotlib | To create charts, graphs, and other visualizations. |
| Scikit-learn | To preprocess the data, feature selection, model selection, and evaluation. |
| Seaborn | To create advanced visuals and statistical graphics |

Dataset Link: https://www.kaggle.com/datasets/stefanoleone992/fifa-22-complete-player-dataset?select=players_22.csv

## Project Timeline

| **Milestone** | **Due date** |
|---|---|
| Project Title | 16th March 2023 |
| Finding the data | 23rd March 2023 |
| Project objective and planning | 30th March 2023 |
| Explore the data | 4th April 2023 |
| Model Selection | 10th April 2023 |
| Train and Test the Data | 11th April 2023 |
| Model Evaluation | 12th April 2023 |
| Final Report | 13th April 2023 |

## Project Budget

| **Category** | **Budget** |
|---|---|
| Internet cost | $500 |
| Hardware cost | $10,000 |
| Staff cost | $25,000 |
| Other cost | $1000 |
| Total project budget | **$36,500** |

## Project Team

Meenakshi - Project Manager
Javier A. Melo - Project Co-ordinator
Amal Vandananikkal Thankachan- Football expert
Henry Jones Inbaraj - Machine Learning Expert
Kelvin Simon - Data Analyst

## Project Stakeholders

Dhruwal Shah – Professor – Lambton College

## Conclusion

In conclusion, by using big data algorithms to assess the performance of individual players, our project plan seeks to fundamentally alter the way football teams are constructed. The best players in each position will be determined by our team of experts using a thorough analysis of the FIFA 22 full player dataset using an algorithm. In doing so, we will be able to offer football managers and coaches priceless insights that will help them put together the best squad possible and improve their chances of winning games.

We are adamant that this initiative has the potential to revolutionize the football business and establish higher benchmarks for player analysis.

Make a big influence on the football world by travelling with us on this journey. Let's collaborate to bring football players' skills to the fore of the game and unlock their untapped potential.