



# BASKETBALL TOURNAMENT PREDICTION SYSTEM

We already know the outcome—why just play?



UNIVERSIDAD DISTRITAL  
FRANCISCO JOSÉ DE CALDAS

By: Nicolás Romero Rodríguez - Carlos Andrés Celis Herrera -Xiomara Salome Arias Arias

Professor: Carlos Andrés Sierra

## INTRODUCTION

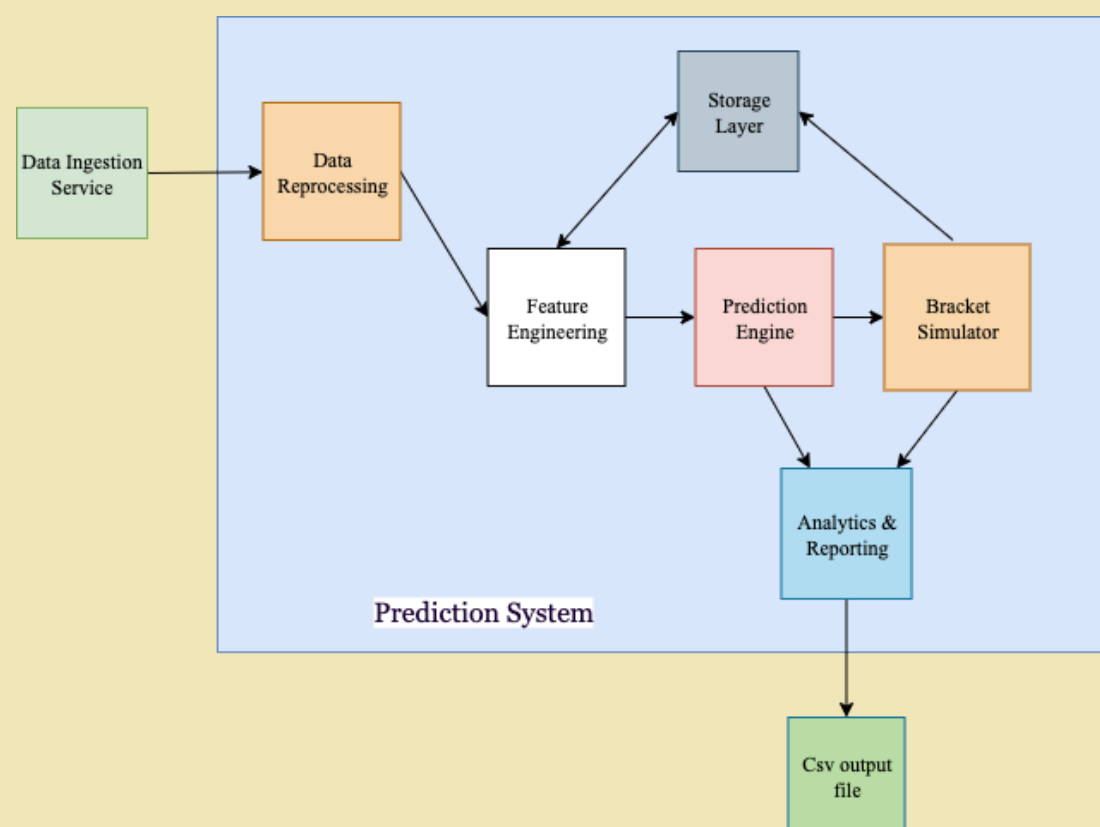
A basketball tournament can be seen as a complex system and as such is made of components, inter-relationships, emergent behaviors and interactions with its environment. Understanding these elements allow us to abstract the system, identify patterns on its behavior and finally generate predictions.

By applying this knowledge its possible to design of a machine learning model capable of predicting the result of any given basketball match in the context of this tournament by aplying data cleaning techniques, features engineering and systems engineering principles.

## OBJECTIVE

To design and implement a system capable of predicting the results of the march madness basketball tournament using machine learning techniques and systems engineering principles

## METHODOLOGY



The architecture consists of a series of independent modules connected through a defined data flow, which will enable the processing of historical performance data and the generation of predictions for tournament matches.

- Careful selection of the most relevant features such as win/lose ratio and average score in every match ensure accuracy in predictions
- Taking into account the most recent data in the dataset is important for making much more accurate predictions and avoiding any values that could skew the model.
- Modular architecture allows flexibility for future improvements, such as integrating new machine learning models or adding real-time data sources without needing to rework the entire system structure.

## PROPOSED SOLUTION

Raw data given to generate the predictions must go through a preprocessing process in order to filter out irrelevant information, later the remaining information is rearranged to keep and organize the most important variables, this new dataset is then fed to a prediction module that uses logistical regression to generate the predictions of the tournament matches' results in terms of probabilities.

## EXPERIMENTS & RESULTS

- It was possible to produce around 9 thousand predictions applying the proposed model,
- it was not possible to identify all possible matchchups within the given dataset and therefore was not possible to introduce an entry for the kaggle competition.
- The metrics realized within the system show that predictions tended to be accurate

## CONCLUSIONS

- The development of a prediction system for a basketball tournament has made evident the inherent complexity of this type of events, which operate as dynamic systems composed of multiple interdependent elements.
- Systems thinking allows understanding complex systems; if its principles are applied, it is easier to build solutions orplans for their implementation in simulation and prediction environments

## BIBLIOGRAPHY

- [1] Kaggle, "March Machine Learning Mania 2025,"
- [2] NCAA, "What is March Madness? The NCAA tournament explained".
- [3] University of Waterloo, "What are complex systems?"