

CS185 Group Project Proposal

Summary

This document serves as the proposal for the group project for CS185. Please fill it out, save as a PDF, and post to canvas.

Project description

In this section, provide a description of the following.

1. Data set (e.g. movie lens, million song data set, 20 news groups, Wikipedia, spam assassin, data.gov, kaggle.com, synthetic, ...)

Our team wants to focus on a dataset pertaining to sports, specifically on either the NBA or NFL. For this proposal, we selected a handful of datasets that interest us from both the NBA and NFL that we want to focus on and find potential topics of interest.

<https://www.kaggle.com/sebastianmantey/nba-free-throws>

<https://www.kaggle.com/ronaldjgrafjr/nfl-draft-outcomes>

<https://www.kaggle.com/kendallgillies/nflstatistics>

2. Goal of project (e.g. anomaly detection, recommendation, clustering, ...). Include the outcomes you wish to produce.

Based on the dataset selected from above, we have the following ideas as potential outcomes we wish to produce. Here are a few as an example,

- Correlations and patterns of where a player is drafted
- Predicting a players free throw ability based on past stats
- Predicting player longevity and potential injury

3. Tools (e.g. MapReduce, Spark, Drill, ...)

Below are a few tools we want to investigate/experiment and start with,

- Tableau
- SQL
- Python/R
- Notebooks (Jupyter/Zeppelin)
- Spark

Group members and responsibilities

In this section, identify each group member and their respective responsibilities for this project. Note that every student is required to deliver an equally timed part of the presentation to reflect their contribution to the project.

1. Leeladithya Avala

- Help analyze data
- Come up with visuals to represent findings

2. Vishaal Prasad

- Help analyze data
- Research into tools viable to process/analyze datasets

The above is a tentative plan for each group member and is not representative of finalized responsibilities for the project.