

ENSF 444 - Final Project Proposal
Group 111

Faheem Mazhar - 30140922
Moiz Bhatti - 30163705

Submitted On: March, 21, 2025

In this project we aim to create a machine learning model to predict and classify any given poker hand from a regular 52 card deck and place it in 10 possible poker hand categories. This model will address a couple different key problems, the first being the ability to ensure fair play in casinos and online platforms, as casinos and online platforms require automated systems to track card distributions and detect unusual gameplay patterns that may indicate cheating or unfair card dealing. The second problem being addressed is helping to create a learning tool for new poker players, as many beginners struggle to recognize the strength of poker hands. This model could be integrated into poker training applications to classify hands in real time, helping players improve their decision-making. Our specific client would be to game organizations and casinos in efforts to ensure tracking for fair gameplay.

The dataset we will be using in this project is: <https://archive.ics.uci.edu/dataset/158/poker+hand>
Some details about the dataset are as follows:

- Dataset Size
 - 1,000,000 training samples, 1,000,000 testing samples
 - Can be randomly sampled to shrink dataset still allowing for accuracy but decrease in training time for some models
- Features
 - 10 input variables representing the suit and rank of five drawn poker cards.
- Target Variables
 - One of 10 poker hand categories, ranging from "No Pair" to "Royal Flush".

The models that will be used for the model will be:

- Logistic Regression (Linear model)
 - This will be used as the simplest baseline model to provide an initial understanding of the relationships between the input features and poker hand categories.
- Gradient Boosted Trees (Non-linear model)
 - This tree based model will be used as it is expected to perform well in handling the complex relationship between card ranks and suits, and using the gradient boosting version of trees will allow for sequential improvement.
- SVM (Non-linear model)
 - As the dataset is of a high dimensionality this model is great in handling high dimensional datasets, and can help determine optimal decision boundaries between different poker hand categories.
- Naive Bayes (Non-linear model)
 - As the dataset is quite large, using a Naive Bayes model will provide a fast and computationally efficient method for classification.