Project Contribution: **Feature Evaluation using cooperative game theory and voting.**

Venkat Aashish Mathsyaraja, Apoorva Patil.

The main goal of the project is to use coalitional game theory methods to solve the problem of feature selection, so we had to look into the previous research first. After getting to know exactly what has been done, we wrote down our objectives for the project clearly, which are to

1) Apply game theory to solve the problem of feature selection
2) Evaluate the importance of each feature by forming coalitions and calculating contribution functions.
3) Improve the performance of the classifier.
4) Try to obtain better results than Feature selection done by just Shapley value analysis (Which is termed as one of the best methods for feature selection)

As we have our targets and we had to complete them sequentially, we assigned deadlines to each task and we divided the work in each task. First, we divided the literature review, as I looked into applying Banzhaf power index for calculating the contribution functions of features. Apoorva looked into applying Shapley value analysis for the contribution and then we met to discuss whatever we have figured. Once we theoretically solved how to apply them in machine learning context we started writing the algorithms for both by stating our specifications. Once we are done with the algorithm design we started to write code in python 3.0 by considering a small data set from UCI's machine learning repository. I coded the algorithm for Banzhaf power index for feature selection and Apoorva coded the Shapley value analysis.
Next, we worked on the other two approaches which evaluates the importance of each feature. One among the methods was using Banzhaf Power Index to filter the most important features by mutual information, then we use Shapley value to obtain the final feature set which has only the features with higher importance. I implemented this method by using Banzhaf Power Index to initially reduce the feature set then further reducing the feature set by using Shapley value analysis. Banzhaf Power Index selects the feature based on the power of each feature. And it mainly selects the critical voters i.e., the necessary features. Banzhaf value refines our feature set. When we later apply Shapley value analysis to this refined set we get very refined feature set with very few or no redundant features. The other approach we implement was combining the votes between features selection from all these approaches and including some additional approaches like randomized decision trees and Pearson correlation. Apoorva implemented this voting on all the features based on the four models, and the features with highest number of votes was chosen as the final feature set. Our voting approach selects the number of features in the final set as the minimum number of features selected by all the four algorithms. Making this approach to select only the most important features or the most necessary features.

We proposed our unified approach mainly to overcome the problem faced by the Shapley value in case of larger data sets. Shapley value analysis tends to fail if it is needed to form large coalitions to improve classification. But, our unified approach overcomes this problem as voting of features is considered between both filter and embedded methods to obtain the final feature set.