Feature Engineering Applied on UCP Dataset

Using Extra Tree Classifier

## Features taken from the dataset

1. Simple Actors
2. Average Actors
3. Complex Actors
4. Simple Use Cases
5. Average Use Cases
6. Complex Use Cases
7. UAW
8. UUCW
9. TCF
10. ECF
11. Effort

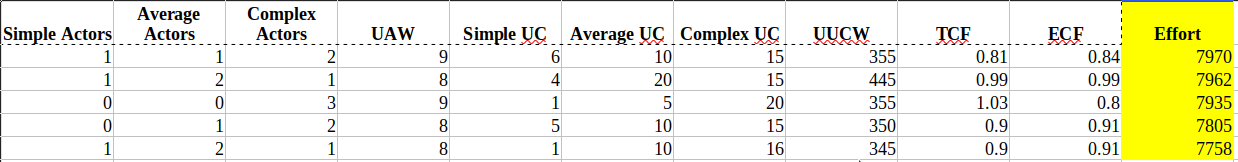
***Effort*** *will be the dependent variable*

Total number of features is 11

Independent variables = 10

Dependent variable = 1

First 5 rows of the dataset



*Dependent column(Effort) is colored yellow*

Applying Extra Tree Classifier

Python libraries used:

**scikit-learn, matplotlib, numpy, pandas, sklearn.ensemble.ExtraTreesClassifier**

After training the classifier on the data, we get a list of features order by their weight

Following are the features and their weight

Simple Actors 0.0149

Average Actors 0.0156

Complex Actors 0.0371

UAW 0.0419

Simple UC 0.0422

Average UC 0.0468

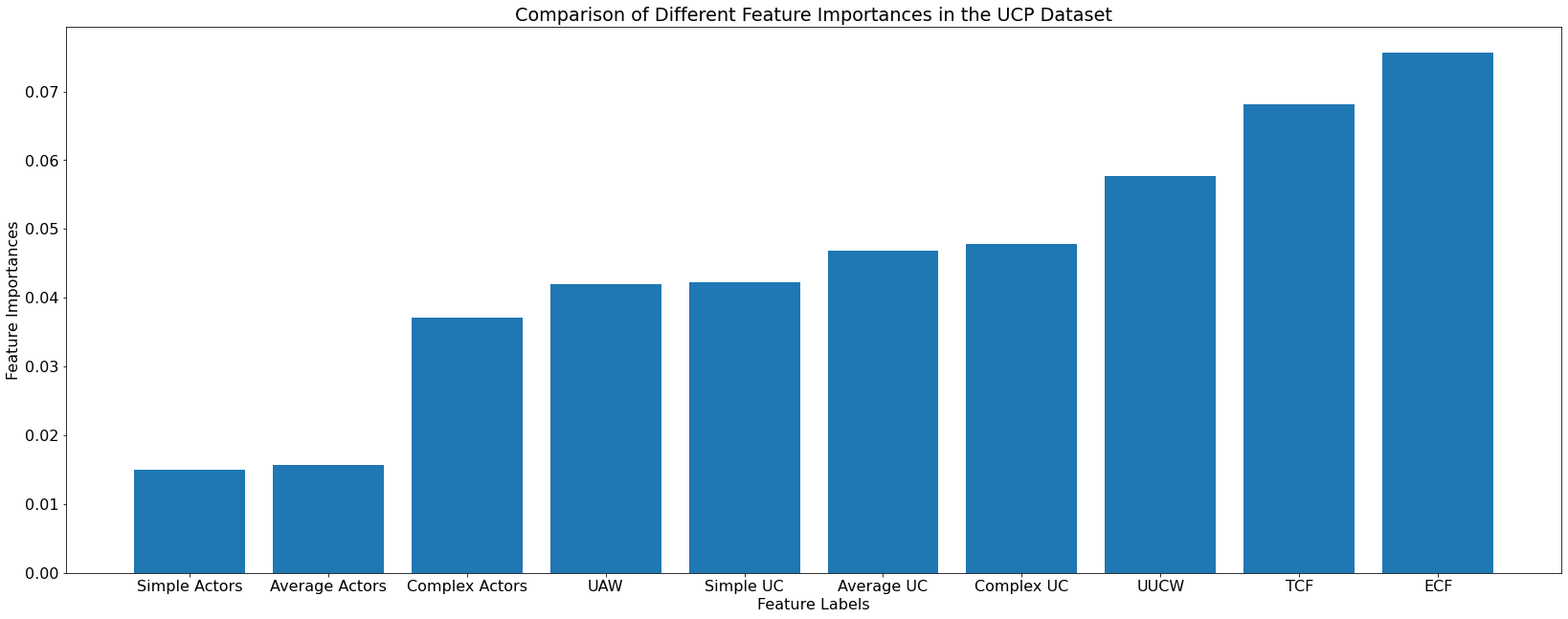
Complex UC 0.0478

UUCW 0.0577

TCF 0.0680

ECF 0.0756

Here is a bar graph of these features



**Summary:**

As show in the graph, the three **most** important factors are **UUCW, TCF, ECF**

and the **least** important factors are **Simple Actors, Average Actors, Complex Actors,**