**FUTURE WORK**

The project here has some limitations and it can be extended further. The first limitation is that we considered a small data set that contains 5000 URLs, and there are 17 features for each URL. The second limitation is that all features are discrete. Often, classifiers such as decision trees, Random Forest, and rule-based systems are more suitable when features are discrete. Furthermore, we used features that were already extracted from URLs.

The present work can be extended as be we can evaluate classifiers using a large data set that contains a few thousands of URLs and extract more number of features that may be significant in decision making. Larger data sets are available in public domain.

We can generate associative rules using the frequent item data sets with the minimum support and confidence values and build a rule-based system using associative rules to classify URLs. The rule-based classifier then can be compared with other classification methods. another approach for generating classification rules from data samples is to divide the feature space using fuzzy membership functions and extract and optimize classification rules. The extracted rules can be used to build a fuzzy inference system that can classify URLs.

In order to avoid the problem of overfitting a classifier, we need to include a pre-process stage. In processing, we can use clustering to find out outliers or noisy data samples. Such samples should not be used in the training set data.