Author – Ranjith Dhanaraj.

Evaluation of Classifiers and Ensemble Techniques

Dataset regarding salary is utilized to train the model. The goal is to predict the salary of the person based on a number of features that are available in the dataset.

	J-48	1-NN
Method	Cross-Validation = 10	Cross-Validation = 10
Accuracy	83.3172	79.0171
Mean Absolute error	0.2151	0.21
Average TP Rate	0.833	0.79
Average FP Rate	0.358	0.361
Average Recall	0.833	0.79
Average F-Measure	0.826	0.79

We can observe that the J-48 classifier is better than the 1-NN classifier for the given dataset. The dataset also seems to be skewed with most of the people earning more than 50k a year, 2337 after resampling to just 756 after resampling who earn less than 50k a year. When a dataset is skewed it is better to consider the F-Measure to be the evaluation metric. If we just go by accuracy, then we might be overestimating the actual performance of the classifier in predicting if an individual earns more than 50k p.a. Since F-measure takes into consideration the precision and recall of the classifier, it is better equipped to deal with irregularities in the dataset.

Evaluation:

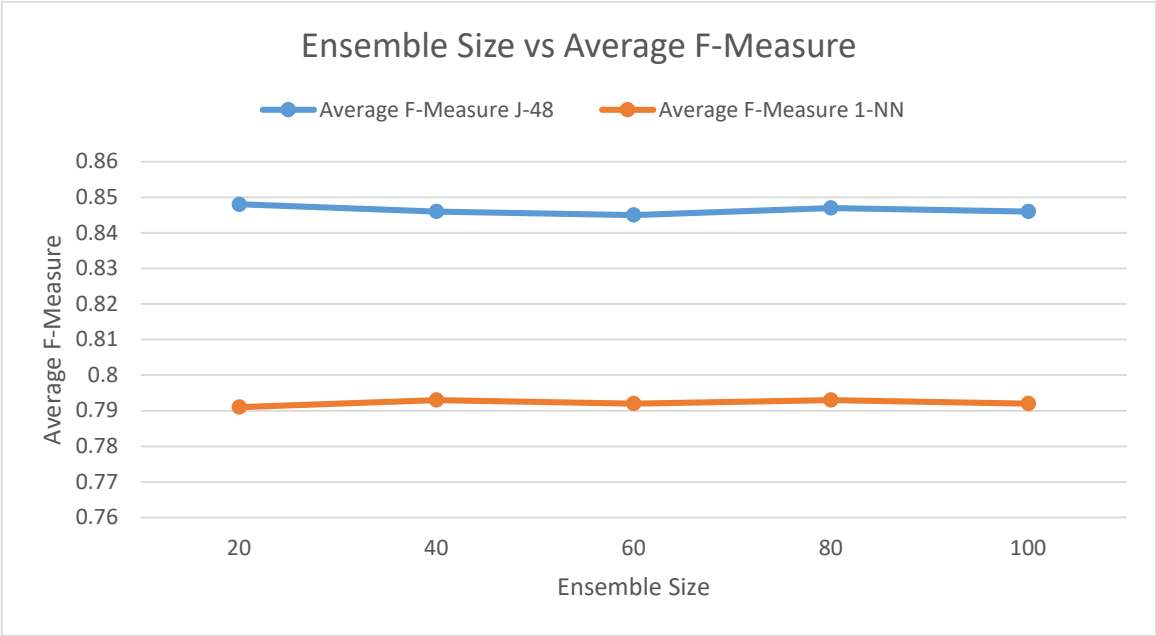
Bagging –

Ensemble Size	Average F-Measure	
	J-48	1-NN
20	0.848	0.791
40	0.846	0.793
60	0.845	0.792
80	0.847	0.793
100	0.846	0.792

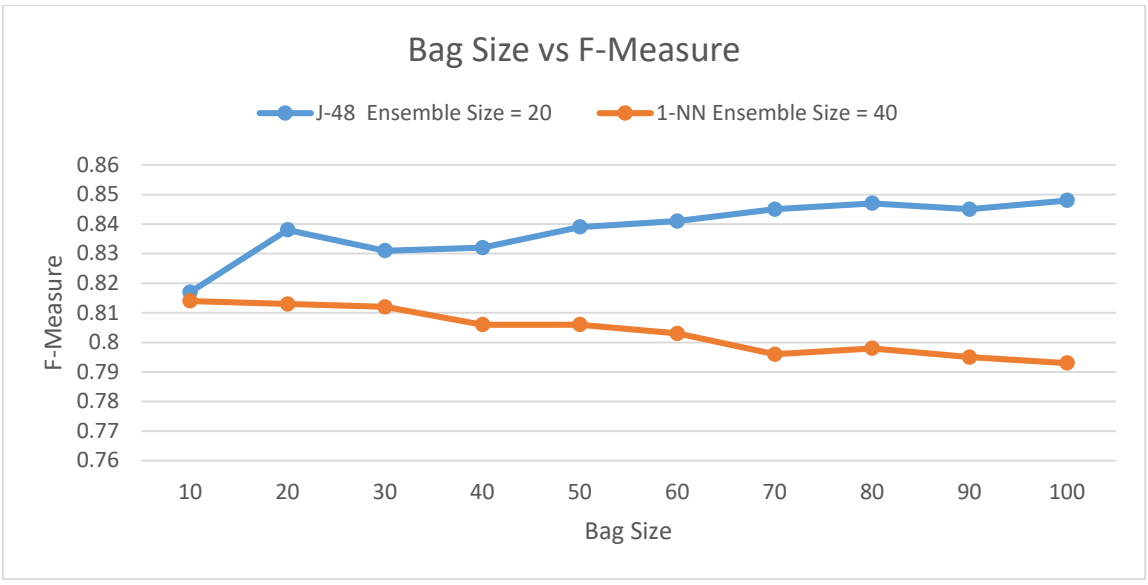
In Bagging we train different classifiers using different subsets of the training data. We use random sampling with replacement technique where in each data point has an equal opportunity to be selected and with replacement, a subset can have the same data point more than once.

Ensemble is a technique in which we combine individual prediction from multiple classifier. This in-turn helps us obtain results with higher accuracy.

We can see from the table that the optimal ensemble size for J-48 classifier is 20 and for 1-NN classifier, it is 40. There is a tie between ensemble size 40 and 80, this issue was resolved by selecting the one with the least Mean Absolute Error.



Bag Size	J-48 Ensemble Size = 20	1-NN Ensemble Size = 40
10	0.817	0.814
20	0.838	0.813
30	0.831	0.812
40	0.832	0.806
50	0.839	0.806
60	0.841	0.803
70	0.845	0.796
80	0.847	0.798
90	0.845	0.795
100	0.848	0.793

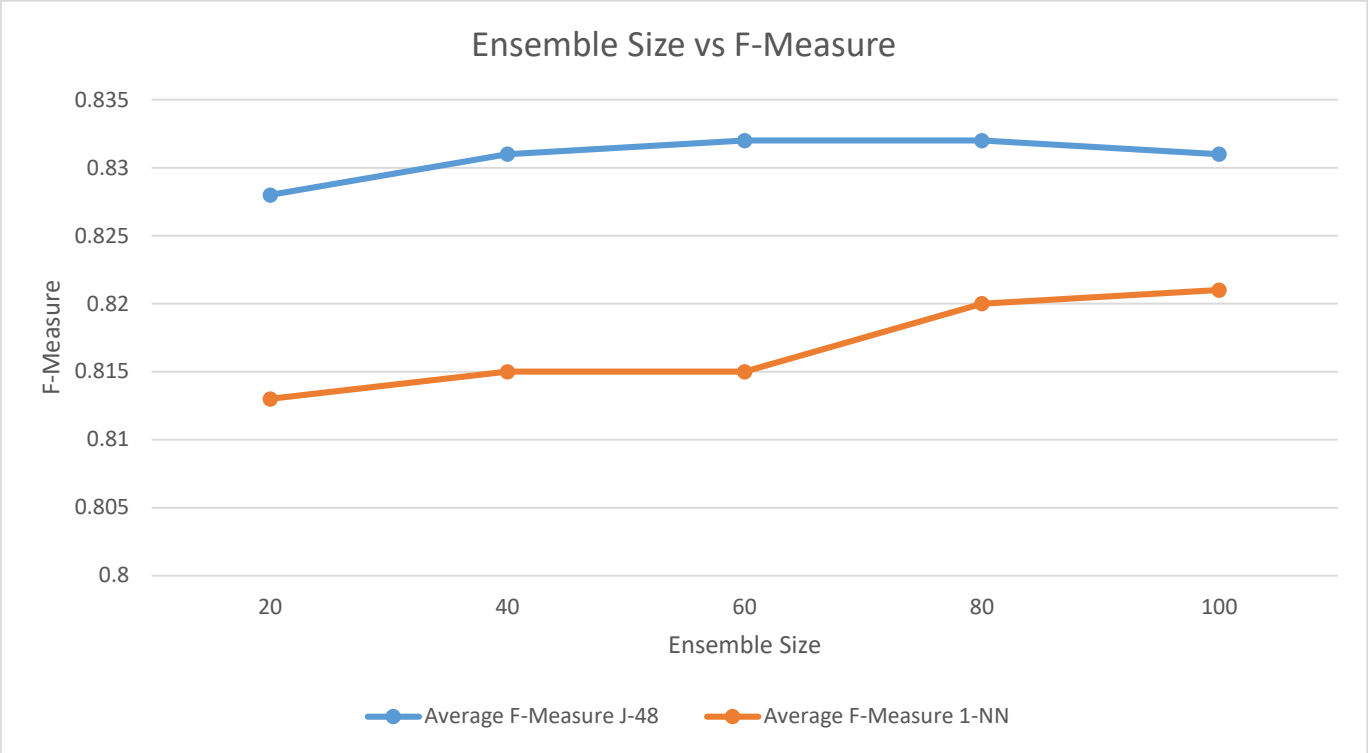


The optimal bag size for J-48 classifier with an ensemble size of 20 is 100 whereas for 1-NN classifier with an ensemble size of 40 is 10. It can be observed that there is a decrease in F-Measure value of the 1-NN classifier from the beginning

of the experiment with the highest value being obtained right at the beginning. It is the complete opposite for J-48 classifier wherein the F-Measure value keeps on increasing with increasing bag size.

Random Subspace Sampling –

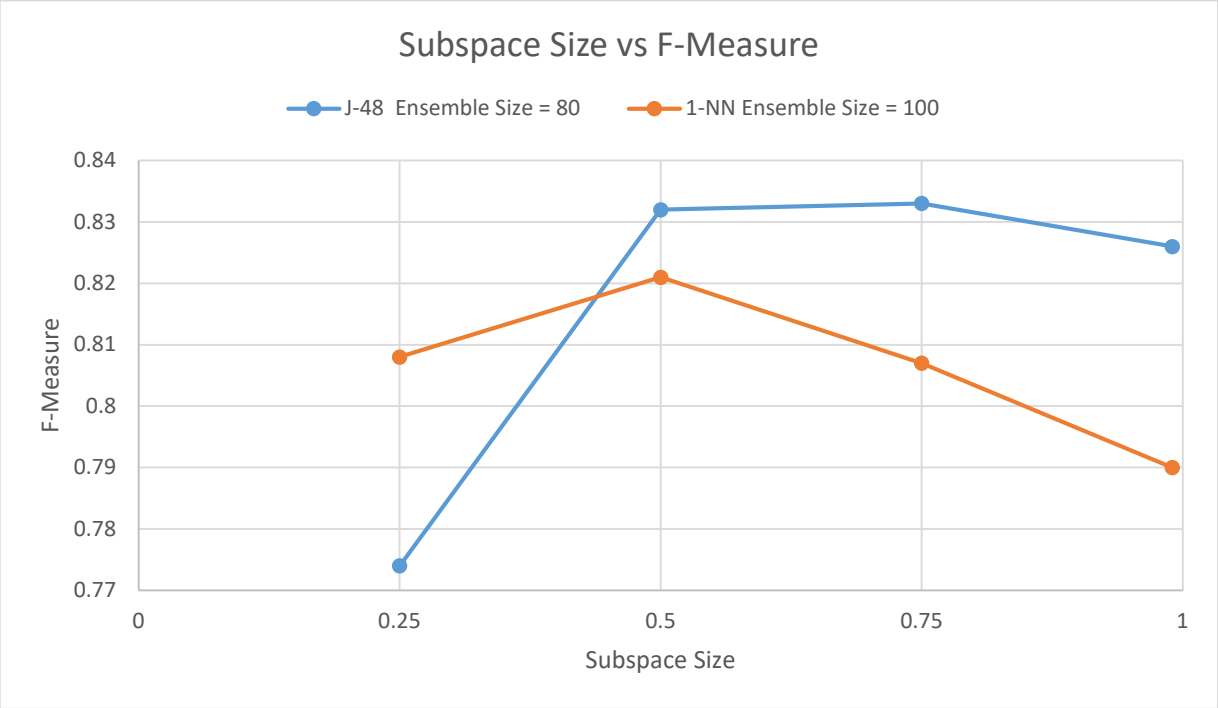
Ensemble Size	Average F-Measure	
	J-48	1-NN
20	0.828	0.813
40	0.831	0.815
60	0.832	0.815
80	0.832	0.82
100	0.831	0.821



The Random Subspace Method, also called Attribute Bagging or Feature Bagging, is an ensemble learning method that attempts to reduce the correlation between estimators in an ensemble by training them on random samples of features instead of the entire feature set. In ensemble learning one tries to combine the models produced by several learners into an ensemble that performs better than the original learners

The ensemble size was again varied from 20-100 in steps of 20. The F-measure values for J-48 classifier was the highest when the ensemble size was 80 and 60. The optimal value was selected to be 80 since the Mean Absolute Error was lower when compared to 60. 1-NN classifier on the other hand did not have any such overlaps with respect to F-measure value and it was the highest when the ensemble size was 100.

Subspace Size	J-48 Ensemble Size = 80	1-NN Ensemble Size = 100
0.25	0.774	0.808
0.5	0.832	0.821
0.75	0.833	0.807
0.99	0.826	0.79



Since, subspace size takes percentage of the attributes into consideration during classification, just four intervals were considered. The default subspace size while classifying is 0.5 (50% of the attributes). In this experiment we change the percentage of the attributes that were selected for learning and try to figure out the optimal amount of attributes that each classifier requires to produce the best results.

It can be observed that the F-measure increase until 75% of the attributes are selected beyond which the value of F-measure decreases rapidly for the J-48 classifier with an ensemble size of 80. The subspace size which produces the best result is 0.75 (75% of the attributes being selected). For 1-NN classifier, the value of F-measure peaks out at 0.5 of the subspace size and decreases slowly beyond that point.

Results –

The performance of each classifier is dependent on the dataset that is has to 'learn'. We can see that the F-measure of the decision tree is favorable with respect to bagging technique. Without the utilization of bagging technique, the F-measure of the J-48 classifier was 0.826 whereas, after the application of bagging techniques, the F-measure rose up to 0.848. When trying to evaluate and determine which classifier benefitted the most from the application of Sub-spacing techniques, it gets quite tricky since there is an increase in F-measure for both the classifiers. The F-measure for the decision tree increases by with the ensemble size but the increase isn't as quite significant as that of 1-NN wherein the F-measure before the application of Sub-spacing techniques was 0.79 and after the application, it jumped up to 0.814 and maintained a high value consistently. Thus, bagging techniques favors the Decision Tree, whereas, Sub-spacing techniques favors 1-NN classifier.

In the dataset provided, it is clear that the J-48 decision tree classifier is better at predicting if an individual is earning more than 50k a year for both the methods, i.e. with Ensemble with Bagging and with Subspace.