Machine Learning Project Essay

This essay will discuss the algorithms and the two datasets that were used in the project. The first being the Adult dataset (Kohavi and Becker, 2018). The second being the Ionosphere dataset (Sigillito, 2018). Both of the datasets are used to train and test machine learning algorithms that are used for binary classification as both of these datasets are about sorting data into two categories. The Adult dataset has <=50K or >50K and the Ionosphere dataset has g(ood) or b(ad). The machine learning model that was used was a voting ensemble made up of three different algorithms. The first algorithm is Naïve Bayes, the second is Sequential Minimal Optimization and the final one is Multilayer Perceptron.

The reason that these algorithms were chosen, was that after research and talking to lecturers about binary classification algorithms and the ones available on WEKA these three were the top recommended ones. As well as this each of the algorithms have different strengths and weaknesses as well as having different levels of complexity. Naïve Bayes is the simplest of the three algorithms however it is good at classification with a small amount of training data which was especially good for the ionosphere dataset as it doesn't have that much data, however due to its relative simplicity the Naïve Bayes is not as accurate as the algorithms. The second algorithm, Sequential Minimal Optimization is the most complex out of the three algorithms and as a result is the most accurate of the three algorithms, however it has a severe disadvantage which is that it takes a lot longer to process the data given to it, sometimes taking as long as 15 minutes to run on a test set. Finally, there is Multilayer Perceptron. Multilayer Perceptron is the second most complex of the algorithms and its main advantage is that it is very good at working with complex and random datasets in order to get the correct result this was especially good for both of the datasets chosen as they're both complex datasets as each has 10+ instances. Now these three algorithms were put into a voting ensemble. They were combined as each of them had different advantages which worked well with the chosen datasets. As a result, when they were combined together and trained on the datasets they did very well together as can be seen in Appendix 4, for the adult dataset the ensemble managed to get a respectable 84.4595% accuracy on the Adult dataset and a very respectable 88.8889% accuracy on the Ionosphere dataset.

The best way to discuss the model architecture is to split it up into the 3 different algorithms that make up the model, discuss them individually and then talk about the voting ensemble model as a whole. The algorithms themselves were each trained about 15 times, after they were trained each algorithm went through experiments each of which was designed to change one parameter and leave the other ones alone. Once each algorithm had each of its individual parameters changed, each setting of each parameter which did the best job was put in and tested. One problem presented doing this technique was that even though some of the parameters when changed in isolation had a positive effect when they were combined it caused problems, this can be best seen with the Multilayer Perceptron in Appendix 3. Overall this technique was successful at improving the algorithms individually as well as when they were part of the ensemble.

For the Naïve Bayes algorithm there were not that many parameters that were changeable. As can be seen in Appendix 1 the ones that were changeable were put into an excel spreadsheet and changed incrementally. The parameters that were the ones that made a difference to the F measure average and the correctly identified instances were the kernel estimator and the supervised discretisation. However, the supervised discretisation made a negative impact on both the F measure and the correctly identified instances. So ultimately the only parameter that was changed from the default Naïve Bayes was making the kernel estimator true.

For the Sequential Minimal Optimization algorithm, it was the opposite of the Naïve Bayes as it had so many different parameters that could be changed and had sub algorithms that had parameters that could also be altered. The only problem was that the Sequential Minimal Optimization algorithm took 15+ minutes to run on my machine and as a result none of the sub-algorithms were changed. So the only tested parameter that changed the F measure and the correctly identified instances was the C value as can be seen in Appendix 2.

For the Multilayer Perceptron there were a lot of different parameters that could be changed and the good thing was it took a fraction of the time of the Sequential Minimal Optimization algorithm which meant all of the different parameters could be experimented with. So the parameters which had an effect on the F measure and the correctly identified instances are: hidden layers; learning rates; momentum; nominal to binary filter; normalise attributes and the seed. However the learning rates, momentum, nominal to binary filter and normalise attributes parameters all had negative impacts on the F measure and the correctly identified instances. The hidden layers and the seed were the only parameters which had a positive impact on the F measure and the correctly identified instances, as can be seen in Appendix 3.

For the voting ensemble algorithm there was only two things that could reasonably be changed and that was the combination rule and the seed. However, both of them only caused a negative impact on the F measure and the correctly identified instances so they were left to their defaults.

The data preparation will be broken down per each dataset. The first one to be discussed is the Adult dataset. The first thing that was done was to download the dataset, which was in the .data format which WEKA would not use, and even WEKA's built in converter wouldn't convert the dataset to the .arff format, so the dataset was copied into excel and all of the attributes and their different values were added to the top of the file and the data was identified as having missing values. It was then saved into the .csv format and then converted into a .arff file. Now that the dataset was in a WEKA readable file it underwent pre-processing. Using an algorithm, the missing data was replaced using the mean and modes from each attribute. The data was then normalised to aid with binary classification. Now that the dataset had underwent pre-processing the final step was to split the dataset into a training, test and validation subsections. This was done using the Resample function built into WEKA. The dataset was split into 80%/10%/10% with 80% being used for the training, and 10% going to both the test and validation subsections respectively, as well as splitting the dataset like this it was also important to make sure that none of the data was duplicated in the datasets as this would lead to false readings, so one of the options of the Resample method allows for it to be specified that none of the data is duplicated among the subsections. The second dataset to be discussed is the Ionosphere dataset. WEKA already had the Ionosphere dataset within it's example datasets. As a result it was already in the .arff format which meant there was no need to do anything to the file format. As well as this the ionosphere dataset doesn't have any missing data which means that during pre-processing it wasn't required to fill in the missing data like it was in the Adult dataset. However just like the Adult dataset an algorithm, the data was then normalised to aid with binary classification. Now that the dataset had finished pre-processing the final step was to split the dataset into a training, test and validation subsections. This was done using the Resample function built into WEKA just as it was with the Adult dataset. The dataset was split into 80%/10%/10% with 80% being used for the training, and 10% going to both the test and validation subsections respectively, as well as splitting the dataset like this it was also important to make sure that none of the data was duplicated in the datasets as this would lead to false readings, so one of the options of the Resample method allows for it to be specified that none of the data is duplicated among the subsections. With the datasets now ready for use by algorithms the three that were chosen were Naïve Bayes, Sequential Minimal Optimization and Multilayer Perceptron. Each one of these algorithms had the training sets ran at least 15 times on it. After this each algorithm was then ran through a series of optimization experiments as was discussed above. Due to the fact that both datasets were binary classification tasks the need for the parameters or the algorithms themselves to be changed for each dataset was not there as the same parameters and algorithms work best for binary classification tasks, this is part of the reason why all of the experiments were done with the Adult dataset, the other part of the reason is so that there was an accurate comparison between the algorithms as they were running under the same amount of training and were running the same training and test sets.

In conclusion the voting ensemble with Naïve Bayes, Sequential Minimal Optimization and Multilayer Perceptron was a good machine learning model to use for both of these datasets as they are binary classification datasets, and those algorithms are designed for use on binary classification problems. The experimental procedure was also the best way to do this as standardising the algorithms to 15 training runs followed by them being run on the test datasets for the experiments allowed for them to be easily compared to one another.

References

Kohavi, R. and Becker, B. (2018). UCI Machine Learning Repository: Adult Data Set. [online] Archive.ics.uci.edu. Available at: http://archive.ics.uci.edu/ml/datasets/adult [Accessed 29 Apr. 2018].

Sigillito, V. (2018). UCI Machine Learning Repository: Ionosphere Data Set. [online] Archive.ics.uci.edu. Available at: https://archive.ics.uci.edu/ml/datasets/ionosphere [Accessed 1 May 2018].

Appendix 1

NaïveBayes Batch Size Experiments

Algorithm	Batch Size	Number of Decimal Places	Kernal Estimator (True/False)	Supervised Discretisation (True/False)	F Measure average	Correctly Identified Instances (%)
NaïveBayes	10	1	FALSE	FALSE	0.826	83.4203
NaïveBayes	20	1	FALSE	FALSE	0.826	83.4203
NaïveBayes	30	1	FALSE	FALSE	0.826	83.4203
NaïveBayes	40	1	FALSE	FALSE	0.826	83.4203
NaïveBayes	50	1	FALSE	FALSE	0.826	83.4203
NaïveBayes	60	1	FALSE	FALSE	0.826	83.4203
NaïveBayes	70	1	FALSE	FALSE	0.826	83.4203
NaïveBayes	80	1	FALSE	FALSE	0.826	83.4203
NaïveBayes	90	1	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	1	FALSE	FALSE	0.826	83.4203

NaïveBayes Number of Decimal Places Experiments

Algorithm	Batch Size	Number of Decimal Places	Kernal Estimator (True/False)	Supervised Discretisation (True/False)	F Measure average	Correctly Identified Instances (%)
NaïveBayes	100	1	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	2	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	3	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	4	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	5	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	6	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	7	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	8	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	9	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	10	FALSE	FALSE	0.826	83.4203

NaïveBayes Kernal Estimator Experiments

Algorithm	Batch Size	Number of Decimal Places	Kernal Estimator (True/False)	Supervised Discretisation (True/False)	F Measure average	Correctly Identified Instances (%)
NaïveBayes	100	2	FALSE	FALSE	0.826	83.4203
NaïveBayes	100	2	TRUE	FALSE	0.856	86.0608

NaïveBayes Supervised Discretisation Experiments

Algorithm	Batch Size	Number of Decimal Places	Kernal Estimator (True/False)	Supervised Discretisation (True/False)	F Measure average	Correctly Identified Instances (%)
NaïveBayes	100	1	FALSE	TRUE	0.846	84.0958

NaïveBayes Number of Decimal Places with Kernal Esimator = TRUE, Experiments

Algorithm	Batch Size	Number of Decimal Places	Kernal Estimator (True/False)	Supervised Discretisation (True/False)	F Measure average	Correctly Identified Instances (%)
NaïveBayes	100	2	TRUE	FALSE	0.845	85.043
NaïveBayes	100	3	TRUE	FALSE	0.845	85.043
NaïveBayes	100	4	TRUE	FALSE	0.845	85.043
NaïveBayes	100	5	TRUE	FALSE	0.845	85.043
NaïveBayes	100	6	TRUE	FALSE	0.845	85.043
NaïveBayes	100	7	TRUE	FALSE	0.845	85.043
NaïveBayes	100	8	TRUE	FALSE	0.845	85.043
NaïveBayes	100	9	TRUE	FALSE	0.845	85.043
NaïveBayes	100	10	TRUE	FALSE	0.845	85.043

Appendix 2

SMO Batch Size Experiments

Algorithm	Batch Size	C	Calibrator	Filer Type	Kernel	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
SMO	10	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	20	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	30	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	40	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	50	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	60	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	70	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	80	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	90	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	100	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097

SMO C Experiments

Algorithm	Batch Size	C	Calibrator	Filer Type	Kernel	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
SMO	100	1	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	100	1.5	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0097
SMO	100	2	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0404
SMO	100	2.5	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.843	85.0373
SMO	100	0.5	weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4	None	weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007	1	1	0.842	84.9513

Appendix 3

MultilayerPerceptron Batch Size Experiments

Algorithm	Batch Size	Hidden Layers	Learning Rate	Momentum	Nominal To Binary Filter (True/False)	Normalise Attributes (True/False)	Normalise Numeric Class (True/False)	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
MultilayerPerceptron	10	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	20	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	30	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	40	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	50	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	60	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	70	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	80	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	90	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973

MultilayerPerceptron Hidden Layers Experiments

Algorithm	Batch Size	Hidden Layers	Learning Rate	Momentum	Nominal To Binary Filter (True/False)	Normalise Attributes (True/False)	Normalise Numeric Class (True/False)	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	100	2	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.826	83.7888
MultilayerPerceptron	100	3	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.825	83.5431
MultilayerPerceptron	100	4	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.832	83.973
MultilayerPerceptron	100	5	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.827	83.666
MultilayerPerceptron	100	6	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.832	84.1572
MultilayerPerceptron	100	7	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.828	83.8809
MultilayerPerceptron	100	8	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.835	84.2186
MultilayerPerceptron	100	9	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.832	83.973
MultilayerPerceptron	100	10	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.839	84.4642
MultilayerPerceptron	100	12	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.832	84.0037
MultilayerPerceptron	100	14	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.829	83.8809

MultilayerPerceptron Learning Rates Experiments

Algorithm	Batch Size	Hidden Layers	Learning Rate	Momentum	Nominal To Binary Filter (True/False)	Normalise Attributes (True/False)	Normalise Numeric Class (True/False)	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	100	1	0.2	0.1	TRUE	TRUE	TRUE	1	0	0.82	83.2054
MultilayerPerceptron	100	1	0.3	0.1	TRUE	TRUE	TRUE	1	0	0.823	83.1747
MultilayerPerceptron	100	1	0.4	0.1	TRUE	TRUE	TRUE	1	0	0	75.7753
MultilayerPerceptron	100	1	0.5	0.1	TRUE	TRUE	TRUE	1	0	0.822	82.8984
MultilayerPerceptron	100	1	0.6	0.1	TRUE	TRUE	TRUE	1	0	0.81	82.5913
MultilayerPerceptron	100	1	0.7	0.1	TRUE	TRUE	TRUE	1	0	0	75.7753
MultilayerPerceptron	100	1	0.8	0.1	TRUE	TRUE	TRUE	1	0	0.806	82.3764
MultilayerPerceptron	100	1	0.9	0.1	TRUE	TRUE	TRUE	1	0	0.81	82.4378
MultilayerPerceptron	100	1	1	0.1	TRUE	TRUE	TRUE	1	0	0.81	82.008

MultilayerPerceptron Momentum Experiments

Algorithm	Batch Size	Hidden Layers	Learning Rate	Momentum	Nominal To Binary Filter (True/False)	Normalise Attributes (True/False)	Normalise Numeric Class (True/False)	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.2	TRUE	TRUE	TRUE	1	0	0.827	83.6967
MultilayerPerceptron	100	1	0.1	0.3	TRUE	TRUE	TRUE	1	0	0.827	83.6045
MultilayerPerceptron	100	1	0.1	0.4	TRUE	TRUE	TRUE	1	0	0.828	83.666
MultilayerPerceptron	100	1	0.1	0.5	TRUE	TRUE	TRUE	1	0	0.824	83.4817
MultilayerPerceptron	100	1	0.1	0.6	TRUE	TRUE	TRUE	1	0	0.823	83.3896
MultilayerPerceptron	100	1	0.1	0.7	TRUE	TRUE	TRUE	1	0	0.822	83.1133
MultilayerPerceptron	100	1	0.1	0.8	TRUE	TRUE	TRUE	1	0	0.798	82.0387
MultilayerPerceptron	100	1	0.1	0.9	TRUE	TRUE	TRUE	1	0	0.82	82.5913
MultilayerPerceptron	100	1	0.1	1	TRUE	TRUE	TRUE	1	0	0	75.7753

MultilayerPerceptron Nominal To Binary Filter Experiments

Algorithm	Batch Size	Hidden Layers	Learning Rate	Momentum	Nominal To Binary Filter (True/False)	Normalise Attributes (True/False)	Normalise Numeric Class (True/False)	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	FALSE	TRUE	TRUE	1	0	0.831	83.666

MultilayerPerceptron Normalise Attributes Experiments

Algorithm	Batch Size	Hidden Layers	Learning Rate	Momentum	Nominal To Binary Filter (True/False)	Normalise Attributes (True/False)	Normalise Numeric Class (True/False)	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	FALSE	TRUE	1	0	0	75.7753

MultilayerPerceptron Normalise Numeric Class Experiments

Algorithm	Batch Size	Hidden Layers	Learning Rate	Momentum	Nominal To Binary Filter (True/False)	Normalise Attributes (True/False)	Normalise Numeric Class (True/False)	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	FALSE	1	0	0.83	83.973

MultilayerPerceptron Number of Decimal Places Experiments

Algorithm	Batch Size	Hidden Layers	Learning Rate	Momentum	Nominal To Binary Filter (True/False)	Normalise Attributes (True/False)	Normalise Numeric Class (True/False)	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	2	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	3	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	4	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	5	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	6	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	7	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	8	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	9	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	10	0	0.83	83.973

MultilayerPerceptron Seed Experiments

Algorithm	Batch Size	Hidden Layers	Learning Rate	Momentum	Nominal To Binary Filter (True/False)	Normalise Attributes (True/False)	Normalise Numeric Class (True/False)	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	0	0.83	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	1	0.829	83.6352
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	2	0.819	81.1483
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	3	0.834	84.0651
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	4	0.831	83.8809
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	5	0.824	83.3589
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	6	0.834	83.5738
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	7	0.826	83.6045
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	8	0.811	80.043
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	9	0.836	83.973
MultilayerPerceptron	100	1	0.1	0.1	TRUE	TRUE	TRUE	1	10	0.814	83.1133

MultilayerPerceptron Optimised Experiments

Algorithm	Batch Size	Hidden Layers	Learning Rate	Momentum	Nominal To Binary Filter (True/False)	Normalise Attributes (True/False)	Normalise Numeric Class (True/False)	Number of Decimal Places	Seed	F Measure average	Correctly Identified Instances (%)
MultilayerPerceptron	100	10	0.1	0.1	TRUE	TRUE	TRUE	1	3	0.829	83.8195
MultilayerPerceptron	100	11	0.1	0.1	TRUE	TRUE	TRUE	1	2	0.835	84.0344
MultilayerPerceptron	100	9	0.1	0.1	TRUE	TRUE	TRUE	1	2	0.833	83.7581
MultilayerPerceptron	100	8	0.1	0.1	TRUE	TRUE	TRUE	1	2	0.839	84.2186

Appendix 4

Voting Ensemble Optimisation Experiments **Voting Ensemble Optimisation Experiments**

Algorithm	Batch Size	Classifiers	Combination Rule	Number of Decimal Places	Seed	Pre-built Classifiers	F Measure average	Correctly Identified Instances (%)
Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Average of Probabilities	2	1	NONE	0.906	85.2925
Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Product of Probabilities	2	1	NONE	0.836	84.398
Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Majority Voting	2	1	NONE	0.845	84.9509
Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Minimum Probability	2	1	NONE	0.836	84.398
Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Maximum Probability	2	1	NONE	0.836	84.3673

Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Average of Probabilities	2	1	NONE	0.906	85.2925
Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Average of Probabilities	2	2	NONE	0.837	84.4595
Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Average of Probabilities	2	5	NONE	0.837	84.4595
Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Average of Probabilities	2	10	NONE	0.837	84.4595

Voting Ensemble Test Experiments

Algorithm	Batch Size	Classifiers	Combination Rule	Number of Decimal Places	Seed	Pre-built Classifiers	Dataset	F Measure average	Correctly Identified Instances (%)
Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Average of Probabilities	2	1	NONE	Adult	0.837	84.4595
Voting Ensemble	100	Optimised NaiveBayes, SMO and MLP	Average of Probabilities	2	1	NONE	Ionosphere	0.888	88.8889