Characteristics Behind the Selection of Base Classifiers in Multiple Classifier System

MSc Research Project
Data Analytics

Vikas Tomer
x150136777

School of Computing
National College of Ireland

Supervisor: Dr. Simon Caton

National College of Ireland
Project Submission Sheet – 2015/2016
School of Computing

Student Name:	Vikas Tomer
Student ID:	x150136777
Programme:	Data Analytics
Year:	2016
Module:	MSc Research Project
Lecturer:	Dr. Simon Caton
Submission Due Date:	12/12/2016
Project Title:	Characteristics Behind the Selection of Base Classifiers in Multiple Classifier System
Word Count:	6489

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	21st December 2016

PLEASE READ THE FOLLOWING INSTRUCTIONS:
1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Characteristics Behind the Selection of Base Classifiers in Multiple Classifier System

Vikas Tomer
x150136777
MSc Research Project in Data Analytics

21st December 2016

Abstract

The pace of generating data in all areas is extremely high. This pace has been mounting the pressure on data scientists to make the advancement continuously, in the selection of machine learning algorithms which are very specific and precise. As there is a range of classifier families available, it is hard to choose which classifier or family of classifiers is appropriate for a particular dataset. This difficulty hinders the advancement of the Multiple Classifier System (MCS). A specified system from several machine learning algorithms will not only be time saving, but also will give more definite and detailed results. There are many articles and a lot of research available in the literature which discuss this problem. However the job of selecting the best classifier from several, for a particular classification task on any dataset, is still incomplete. There is no concrete solution available for this complex problem in the collected works. This project focuses on selecting the appropriate base classifiers in the Multiple Classifier System for a particular dataset. The primary motive of this project is to make some helpful contribution in this direction by finding out those characteristics which decide the selection method of base classifiers. This project advocates for the absence of some certain classifier in the base layer of the MCS, as the absence of these classifier increase the accuracy of pridiction in the MCS. Furthermore, this project corroborate the efficacy of Vote Meta learner more than the Stacking Meta learner.

Keywords: Multiple Classifier System (MCS), Base Classifier, Ensemble of Classifier (EoC), Stacked Generalization.

1 Introduction

The last two decades have seen a rapid development in the area of machine learning. Different machine learning algorithms have been introduced over the time to solve the complicated tasks of prediction. Numeric Prediction, Classification, Clustering and Regression are some examples of the complicated tasks of Data Mining. Data scientists have been introducing a wide range of new classifiers into the families of classifiers to make the prediction task easier. Each family has a range of classifiers for doing the classification task and new classifiers continue to be added to these classifier families. Due to this wide range of available classifiers with huge capability for solving different problems, new domains are adapting these classifiers every day. Email spam filtering, credit card fraud detection, remote sensing of images, text based analysis, sentiment analysis, pattern recognition, and so on, are some of the key examples of these new domains. All these tasks come under the umbrella of supervised machine learning. Machine learning is basically categorized into two types- Supervised and Unsupervised learning (Ayodele, Taiwo Oladipupo; (2010). Classification and Regression are the components of Supervised learning, as they require the data for training before testing the data. On the other hand, in Unsupervised learning, there is no target feature to achieve. It searches the interesting pattern from the untrained dataset. Clustering and Association Rules are the components of unsupervised learning.

The concept of MCS is described as multiple machine learning algorithms applied on some classification task for getting better predicting accuracy in comparison to any single algorithm (Arruti, A and Mendialdua, I and Sierra, B and Lazkano, E and Jauregi, E; (2014). It is also seen as Supervised Learning, as it trains the data before doing the classification (predicting outcome). So prediction is done on the premise of some hypotheses. Of course, MCS provides the better accuracy by using various classifiers on the base layer of MCS (Tulyakov, Sergey and Jaeger, Stefan and Govindaraju, Venu and Doermann, David; (2008). In MCS, the selection of base classifiers is the heart of the entire system. These base classifiers are then fed into the Meta learner. The Meta layer aggregates the results from base classifiers into one unit. Different Meta learners have their own methodology of aggregating the results.

This project mainly focuses on Meta Learners- Stacking and Voting. It has been observed in previous research that accuracy of prediction is highly dependent on the base classifier (Fernández-Delgado, Manuel and Cernadas, Eva and Barro, Senén and Amorim, Dinani and Amorim Fernández-Delgado, Dinani; (2014). The better the selection of these base classifiers and the combination of these base classifiers with the Meta layer, the better will be the accuracy in the prediction. However the selection of base classifiers is the subjective task and requires a huge amount of time to unravel this entire complex dilemma. This quandary encourages the research question of this project i.e. which characteristics facilitate the selection method of base classifiers in the Multiple Classifier System for Meta layers (like stacking or vote)?

The research question contains some key terms, which should be described before moving towards the solution. What are the base classifiers? Why do we need them in the multiple classifier system? Why is the good selection of these base classifiers so important?

The first thing to clarify is why just using a single classifier is insufficient. Is it due to unawareness about the perfect fit of a classifier to a particular dataset or is there any other reason? Choosing a single classifier for an entire dataset has problems; for instance, the selection of a classifier which is appropriate for a particular dataset is time consuming and a complicated task and expecting the highest accuracy from a single classifier is not possible every time. In other words, selection of an expert learner for a different dataset every time is not possible (Giorgio Valentini , Francesco Masulli; (2002).

The second element to consider is the diversity in the selection. How can diversification be achieved between the classifiers? Furthermore, Diversity is not only an issue within a single family of classifiers, but it is also an issue between the different families of classifiers. So, the question arises, whether the diversity should be retained within a single family of classifiers or should it exist within the entire families of classifiers? An example of diversity in base classifier, which is very popular in many articles of data mining- is the game show (Who wants to be a millionaire), in which an audience poll has always been better than an expert opinion. It can be interpreted that several different weak learners can work more effectively in a team rather than one strong expert learner. Diversity is always helpful in getting the higher accuracy in the prediction task.

The third factor, which creates confusion in the mind of the user is, if more than one classifier is always beneficial then what should be the size of the base classifier in the MCS? What should be the exact number of base classifiers in a single operation of experiment by the Multiple Classifier System? How many classifiers are required by Diversity at the least and at the most? All these questions are sorted out by a huge set of experiments in this project.

Having all these three major problems above, the data scientists have formulated some optimal solutions, which can be derived with the help of the Ensemble of Classification (EoC) (Cavalin, Paulo R. and Sabourin, Robert and Suen, Ching Y.; (2013) in the MCS. The MCS basically works in the following way:

The combination of different base classifiers are fed into one Meta layer for accomplishing a particular classification task. This combination of base classifiers in the base layer varies every time, and depends on what situation and / or a dataset demands at that particular time (Giorgio Valentini , Francesco Masulli; (2002). Diversification is needed to acquire higher accuracy from the Multiple Classification System. The combination between the base layer and meta layer also plays a vital role in any set of experiments. The base classifiers from different families do not make the collaboration with all meta learners.

The project work is composed of five themed sections including the Introduction as section 1. Section 2 describes the related work, this section covers the theoretical and experimental aspects of the research question. Section 3 presents the methodology of the project, focusing on the three key themes i.e. Framework, Experiments and Analysis. Section 4 is concerned with the implementation for this project, and includes the details of the working procedure. Section 5 Evaluation, describes the analysis of the results.

2 Related Work

Research into the classifiers has not a long history, still there is a significant amount of literature that has been published about the base classifiers in the MCS. The entire study of related work is grouped into the following four major categories:

2.1 Advantage of the multiple classifier system over single classifiers

In the observational study of the MCS, many articles advocate the use of the MCS over single classifiers because of several advantages. For instance; (Son, Hyojoo and Kim, Changmin and Hwang, Nahyae and Kim, Changwan and Kang, Youngcheol; (2014) advocate the use of the multiple classifier system in the automatic detection of construction material in images over a single classifier. In this particular work, the performance of the MCS is complex to understand but represents a higher accuracy in comparison to the single classifier. Whereas (Marqués, A. I. and García, V. and Sánchez, J. S.; (2012) compare different machine learning algorithms in the ensemble classifier in order explore the credit score through base classifiers. The final result shows the dominance of C4.5 decision tree over many other classifiers in ensemble. A novel microarray based predicting model presented by (Chen, Minjun and Shi, Leming and Kelly, Reagan and Perkins, Roger and Fang, Hong and Tong, Weida; (2011) also achieved better results in comparison to the single optimized model.

From this category of the review it is clear that the multiple classifier system can be used in a variety of analyses and there are several advantages of using the MCS instead of a single classifier. Feedback from the MCS is required in a specific real life application of data mining for substantiating the fact that it is advantageous in every domain. So, the next category of review is on the use of the MCS in a specific real life application of data mining, which is the remote sensing of images.

2.2 Use of the multiple classifier system in the remote sensing of images

Remote sensing of images is one of the domains, in which several novel approaches of machine learning have been revolutionized by using the MCS. In a new approach, for sensing the medium-high remote images, (Yang, Haibo and Zhao, Hongling and Wang, Zongmin; (2010) use a hybrid MCS approach which is a combination of the largest confidence algorithm, an ensemble method Bagging and an optimal set of sub-classifiers. The results show an improvement in the accuracy. More than two ensemble methods are also used in the Remote Sensing of images, as represented by the kind of work conducted by (Yang, Bin and Cao, Chunxiang and Xing, Ying and Li, Xiaowen; (2014). In this approach the objective is to find the land use type reflec ted. There are three types of ensembles used in this approach, comparative major voting (CMV), Bayesian Average (BA) and the new proposed method WA-AHP. The new method WA-BHP performed comparatively better than both CMV and BA.

Remote sensing of images is the field in which prediction can be achieved by both methods of machine learning, i.e. supervised and unsupervised learning (Mukhopadhyay, Anirban and Maulik, Ujjwal and Bandyopadhyay, Sanghamitra and Coello, Carlos Artemio Coello; (2014). Particularly, the study about this area provides the procedures of better optimization and procedure of recovery, if laser segmentation fails to operate correctly. There are some issues where the multi objective optimization is required. The subsequent category of literature review provides a deep analysis about dominance of Stacked Generalization over other ensemble methods in the MCS.

2.3 Dominance of Stacked Generalization over other ensemble methods

At the begining days of Stacked Generalisation as a new technology, it solved only those issues which could not be efficiently solved by the other ensemble methods. For example (Ting, Kai Ming and Witten, Ian H.; (1999) present the solution of higher level learning issues by combining the three learning algorithms by implementing stacked generalization. Stacked generalization proves better in the comparison of arcing and bagging in the results. Then, as seen in (Seewald, Alexander K.; (2002), the quality of stacked generalization is improved by adding the meta-learner layer (MTL). Earlier the stacked generalization worked only to an average level in the case of multiclass classification in comparison to two class classification. That is the reason for the higher number of classes; Stacking with MTL impairs this weakness for dealing with the high dimensionality of the data.

More recently, many new approaches have made this ensemble method more attractive, such as (Chen et al.; (2014) who propose a novel approach of Ant Colony Optimization (ACO) based on stacking ensemble configuration searching. This stacking configuration is based on a pool of base classifiers and then it is collaborated with the meta-classifier. (Mendialdua, Iñigo and Arruti, Andoni and Jauregi, Ekaitz and Lazkano, Elena and Sierra, Basilio; (2015) propose the approach of a layered architecture of classifiers. In this architecture, Layer 0 consists of base classifiers, in which there are a number of classifiers used which are appropriate for the dataset. Then after that, output of each classifier of this layer ensemble by stacked generalization, which is further used by the meta-learner layer classifier for the final result.

Stacked generalization of the ensemble method can be labelled as one of the best methods, as this method has the capability to provide a better compatibility between the different layers of the MCS. This method also provides the facility to sort out the problem of dealing with the high dimensionality of the data. It performs comparatively better if it uses the meta-classifier in the top layer.

2.4 Progression in the Multiple Classifier System for resolving the existing issues

The Multiple Classifier System (MCS) have increased the accuracy level of classification by continuously adding new features regularly, but there are some existing issues which have been associated with the MCS. There are some articles in the literature which raise different issues and provide the solution of those issues by their own capability. (Beitia, Mendialdua; (2015) introduces the technique to reduce the number of base classifiers in a MCS named as one-vs-one using the Undirected Cyclic Graph. The results represent the pros and cons of using machine learning algorithms like: SVM, Ripper, and C4.5. Spam filtering is one of the most common examples of intelligent adversary, which adapts the patterns. (Biggio, Battista and Fumera, Giorgio and Roli, Fabio; (2009) raise a very interesting issue of the adverse problems of classification, in which a classifier has to deal with the intelligent adversary who adaptively modifies the pattern to evade the system. This project resolves the issue by using the architecture of MCS. The result represents the architecture of MCS, which can increase the hardness to evasion. Another significant issue is the criteria of the selection of features from the dataset as raised by (Basu, Tanmay; (2012), who provides the automatic feature selection model to solve this issue. In this model, there is a procedure to find the similarity between the term and class and provide a score over the other classes. The result shows an improvement in accuracy by using this model.

MCS has improved a lot from its early days, still there are so many problems which either remain unsolved like the criteria behind the selection of base classifiers in MCS. Improvement in this area will open many doorways of rapid advancement in data mining.

2.5 Summary of literature review

After scrutinising the related work, the first task is to verify and validate the different hypotheses associated with this issue. There are so many articles in the related work that, do not advocate the single classifier for getting better results. Although there are some specific circumstances where a few of the classifiers give comparatively better results, these cannot be a benchmark for all. Overall following observations may support for the following hypotheses:

Choosing A single classifier for performing the best result for an entire dataset is almost impossible (Mendialdua, Iñigo and Arruti, Andoni and Jauregi, Ekaitz and Lazkano, Elena and Sierra, Basilio; (2015). A combination of base classifiers and Meta classifier is more efficient (Mendialdua, Iñigo and Arruti, Andoni and Jauregi, Ekaitz and Lazkano, Elena and Sierra, Basilio; (2015). The MCS performs better if the top layer is the Meta-Learner layer (Seewald, Alexander K.; (2002).

Hence, this project is carried out all the experimentation by considering above hypotheses.

3 Methodology

To determine the characteristics behind the selection of base classifiers in the MCS, the methodology is designed in the following figure by three elementary stages:

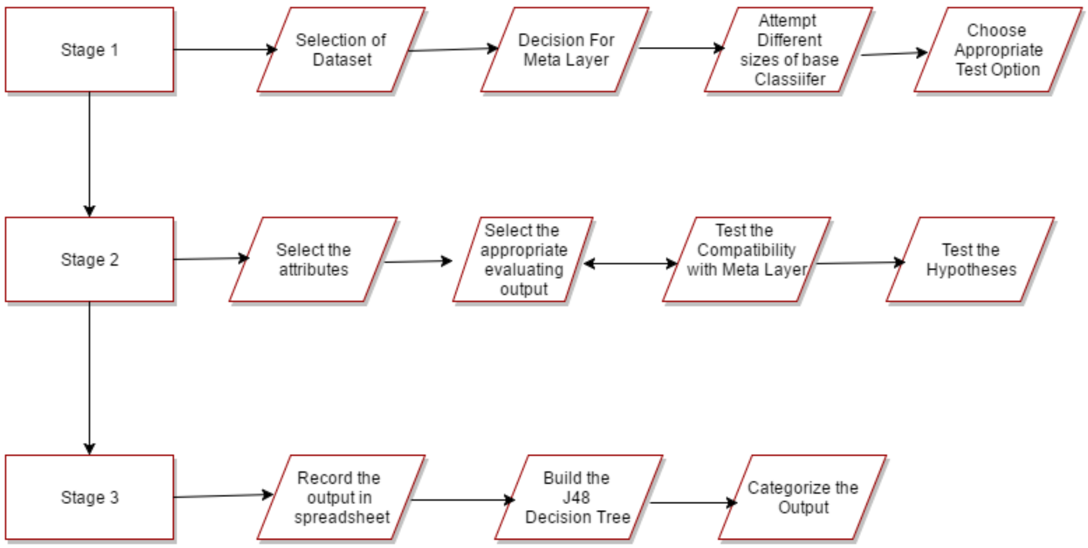

Figure 1: Formulation of Methodology

3.1 Methodology for the setup of the framework for experimentation

Wisely select the datasets for the final experimentation. Each dataset should be distinct from another. This project has three distinct datasets, each one belongs to a different category:

1. The IRIS dataset, this dataset belongs to multiclass classification, as is has more than two classes. This dataset is a multivariate dataset from life science. Hence, the MCS require to deal with some specific challenges to classify this dataset.

2. Medical dataset, a cardiac data is used to classifying the Stress Echocardiography. This classification provide lots of effective information about the working of the MCS during the experiments.

3. Credit Dataset is a detail of loan aspirants, this dataset consist of the transactional data from a bank. It is used for classifying the defaulters from all customers. This dataset is interesting because there is a good mix of attributes, like; continuous, nominal with small numbers of values, and nominal with larger numbers of values. There are also a few missing values.

There are Two meta layers are used in this project- Vote, and Stacking for each particular dataset. First of all, Dissemination of the task into the subtasks for the first layer classifiers (Ko, Albert Hung Ren and Sabourin, Robert; (2013) take place. Then, assign the size of base classifier in the base layer during each experiment. It starts with a single classifier, then the combination of base classifiers for each subset of dataset (Cavalin, Paulo R. and Sabourin, Robert and Suen, Ching Y.; (2013). Continue the experiments by assigning different combinations of classifier up to three, with in a family of classifier. After that prudently select the test options for each classification. In this project, two of the most general options are used- Training set and Cross validation. For Cross validation, 10 folds value is set for each operation. Finally, Appropriately select the parametric or nonparametric methods that are going to be applied on particular datasets (Demšar, Janez; (2006).

After setting up the framework for experimentation, the next task is to determine, what experiments are going to be done? How are these experiments going to be done? And what should be the order of these experiments? For answering all these questions, there is another stage of methodology.

3.2 Methodology for operations to be performed

The Process starts by engineering the features for the model design and implementation. Better feature selection provides the better split in individual decision tree and contributes a lot in the final accuracy (Domingos; 2012). Then select the attributes and check for missing attributes and other statistics- Standard Deviation and mean. Carry out the experiments for evaluating, Output Model, Output per Class Stats, Output Confusion Matrix and Output Predictions for Visualization. In this project, the main evaluation used from the output of each classification are- Correctly Classified Instances, Incorrectly Classified Instances, Kappa statistic, Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error and Total Number of Instances. During the experiments, tune the parameters of those particular models based on the subset of data assigned to them (Domingos; 2012) and test the compatibility of the first layer of the MCS to the next layer of the classifier (Beitia, Mendialdua; (2015), (Arruti, A and Mendialdua, I and Sierra, B and Lazkano, E and Jauregi, E; (2014). Finally,test the hypotheses, whether these are statistically significant or not (Demšar, Janez; (2006).

3.3 Methodology to construct the analysis of the results

This stage of methodology is build for analysing the results, as the result contain a huge amount of data, so first of all construct the spreadsheets for recording the outputs and other details of experiments done. Then for getting the useful pattern from this data, make decision trees on each Meta layer of different spreadsheets. This project opt J48 Decision tree to evaluate the results, as it is applicable on any dataset either nominal or numerical. These spreadsheets consist of of both of the attributes. Also J48 Decision trees are easy to interpret. The performance measures are categorised in Outstanding, Excellent, Good, Average, and Poor to make the user interface more easy.

4 Implementation

This project works on all available base classifiers families i.e. Bayes, Function, MI, Rules, and Trees. All the possible permutations of the base classifiers of maximum size 3, are fed into the Meta Layer. There are two Meta layers used in this project- Stacking and voting. The main tool which has been used in this project for maximum tasks is Weka. R studio, Rapid Miner, Tableau, and Excel are other tools, which help in different tasks of this project.

The procedure of the implementation starts with the selection of datasets. Each dataset should be different from another for making the project versatile. Initially, each dataset has its own limitations. They have to be cleaned and converted into an uniform format and conversion should be done in arff format by Excel97-20032ArffConverter tool. Then import the first dataset into Weka through the explorer and go for classifying the task. In the next step, select the attributes, which are required for the operation. After that, select the option "Using Training set" from the different classifier evaluation options. If Cross validation is used for evaluation, then set the 10 folds value for each operation. Next step is to select the attributes, check for missing attributes and other statistics- Standard Deviation and mean. Afterwords go for Classify and Select the Meta layer for the Multiple Classification System, i.e. either Stacking or Voting. Subsequently, decide the number of sizes of the base classifiers and select the base classifiers according to their size for operation from different classifier families and for saving all the parameters, click OK. For performing the operation, click Start and observe the performance measures from different output window. Different performance measures, which comes out from the output are: Correctly Classified Instances, Incorrectly Classified Instances, Kappa statistic, Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error and Total Number of Instances. For analysing the results, Prepare two different spreadsheets for each dataset for each Meta layer- Stacking and Voting. By doing this, total six spread sheet is prepared, each has almost 100 sets of experiments on almost 50 different base classifiers. Finally, construct the decision trees for each single performance measure on almost 400 experiments for analyzing the results.

5 Evaluation

This project evaluates the final results by making a separate spreadsheet for each Meta layer for each different dataset. All the spread sheets are included in the appendix. For analysing the total outcome, Decision trees are built on different performance measures. The left hand side branches of these decision trees represent the absenteeism of any particular classifier, whereas right hand side braches represent the presence of classifiers. Each of the following case studies represent analysis of the result of each dataset.

5.1 Case Study 1: IRIS dataset

This dataset represents the polynomial categorical class. In this project, the IRIS dataset generates the following output through decision trees on the different performance measures respectively.

Decision Tree for Kappa Statistics on Voting meta layer

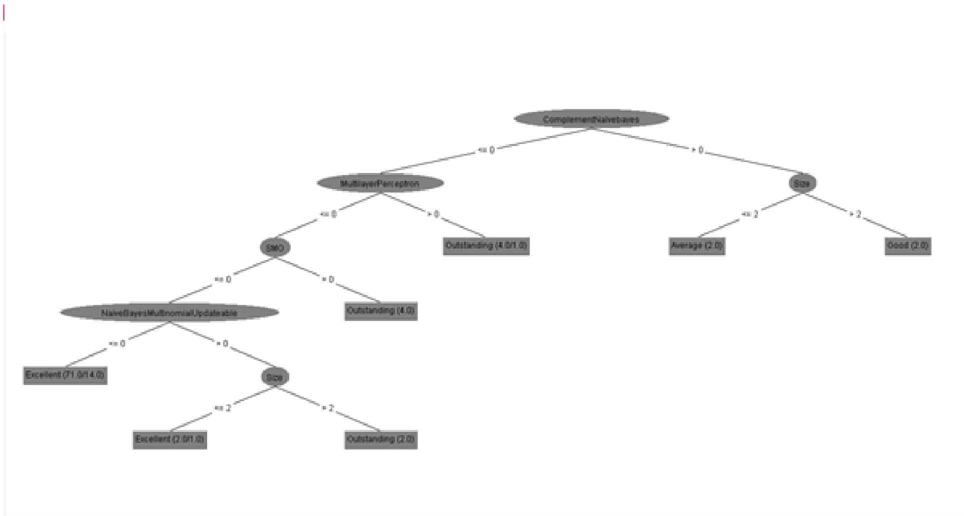

Figure 2: Decision Tree on Kappa Statistics on Voting meta layer

Interpretation: As the result (Figure 2) represents that the absence of Complement Naive Bayes, Multi-layer Perceptron, and Naive Bayes Updateable respectively, probably gives the Excellent result. However, it gives the Outstanding result in the presence of any single classifier from them. It also represents that the presence of complement nave bayes with the size of less than equal to two generates the Average result.

Decision Tree for Correctly Classified Instances on Voting meta layer

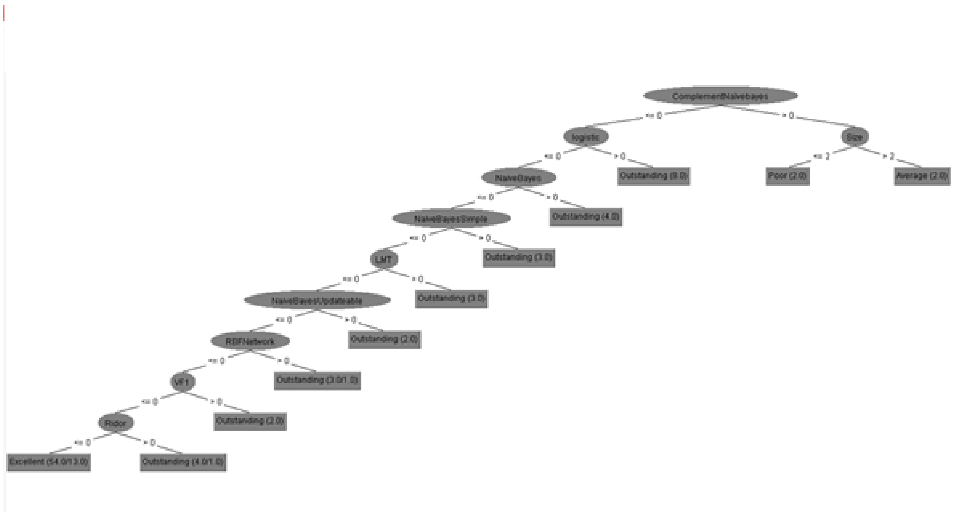

Figure 3: Decision Tree on Correctly Classified Instances on Voting meta layer

Interpretation: As the result (Figure 3) represents that the absence of Complement Naive Bayes, Logistic, Naive Bayes, Naive Bayes Simple, LMT, Naive Bayes Updateable, RBF Network, VF1 and Ridor respectively, probably gives the Excellent result. However, it gives the Outstanding result in the presence of any single classifier from them. It also represents that the presence of complement naive bayes with size of less than equal to two generates the Poor result.

Decision Tree for Root Mean Squared Error on Voting meta layer

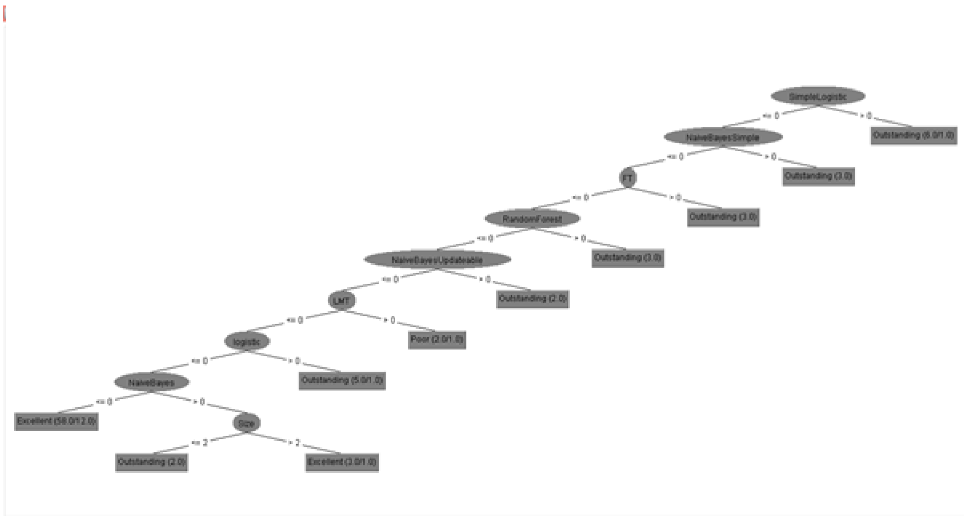

Figure 4: Decision Tree on Root Mean Squared Error on Voting meta layer

Interpretation: As the result (Figure 4) represents that the absence of Simple Logistic, Naive Bayes Simple, FT, Random Forest, Naive Bayes Updateable, LMT, Logistic, and Naive Bayes respectively, probably gives the Excellent result. However, it gives the Outstanding result in the presence of any single classifier from them except LMT. In the presence of LMT, it gives the Poor result.

Decision Tree for Time Taken on Stacking meta layer

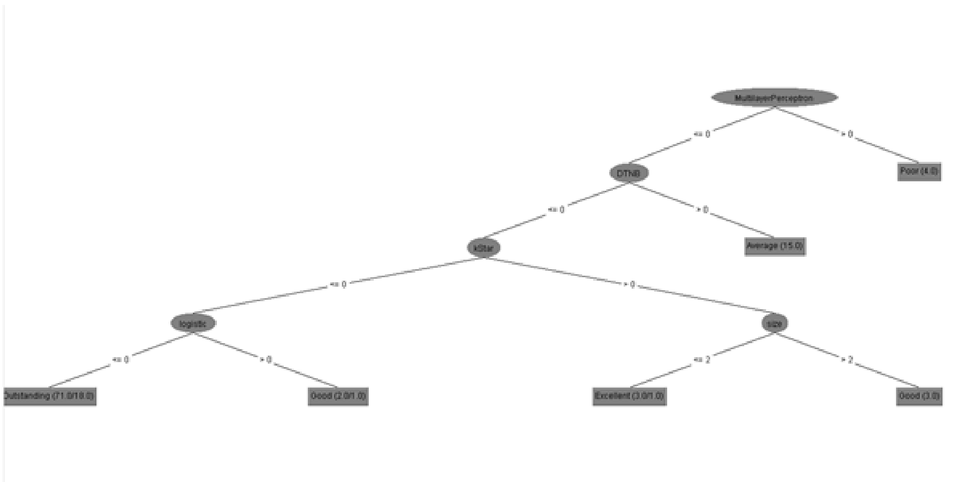

Figure 5: Decision Tree on Time Taken on Stacking meta layer

Interpretation: As the result (Figure 5) represents that the absence of Multi-layer Perceptron, DTNB, Kstar, and Logistic respectively, probably gives the Outstanding result. However, it gives the Poor result in the presence of Multi-layer perceptron only. It also represents that in the presence of DTNB it provides the Average result.

It has been observed from the analysis of the IRIS dataset that it provide either poor or average results quite often when it uses base classiifers from Bayes family with size less than two.

5.2 Case Study 2: Credit dataset

This dataset represents the polynomial categorical class. In this project, the Credit dataset generates the following decision trees on the different performance measures respectively.

Decision Tree for Kappa Statistics on Voting meta layer

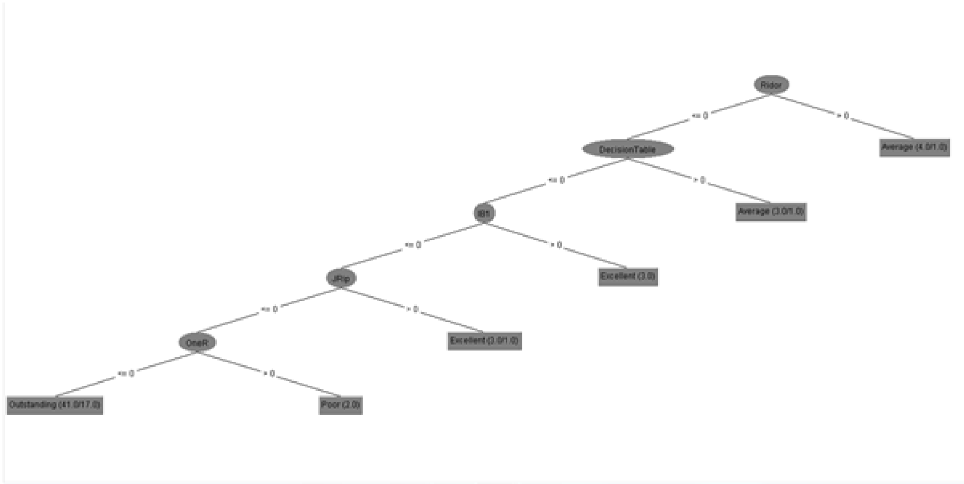

Figure 6: Decision Tree on Kappa Statistics on Voting meta layer Credit Dataset

Interpretation: As the result (Figure 6) represents that the absence of Ridor, Decision Stump, IB1, JRip, OneR respectively, probably gives the Outstanding result. However, in the presence of OneR, it gives the Poor result. Absence of classifiers- IB1 and JRip give the Excellent result.

Decision Tree for Correctly Classified Instances on Voting meta layer

Interpretation: As the result (Figure 7) represents that the absence of DTNB, Logistic, SMO, Naive Bayes Simple, IB1, Decision Stump, Ridor, NNge, IBK, and BayesNet respectively, probably gives the Poor result. However, it gives the Outstanding result in the presence of any of these single classifiers- Logistic, SMO, BayesNet, and IBK. In the presence of Naive Bayes Simple, it gives either the Outstanding or Excellent result based on the size.

Decision Tree for Root Mean Squared Error on Voting meta layer

Interpretation: As the result (Figure 8) represents that the absence of Hyperpipes, logistic, IB1 respectively, probably gives the Excellent result. However, it gives the Outstanding result in the presence Logistic only. It also represents that the presence of IB1, probably generate the Poor result.

Decision Tree for Time Taken on Stacking meta layer

Interpretation: As the result (Figure 9) represents that the absence of Hyperpipes, BayesNet, Nave Bayes, RBF Network respectively, probably gives Outstanding result. However, it gives the Outstanding result in the presence HyperPipes only. It also represents that the presence of BayesNet and Naive Bayes respectively, probably generates the Excellent result.

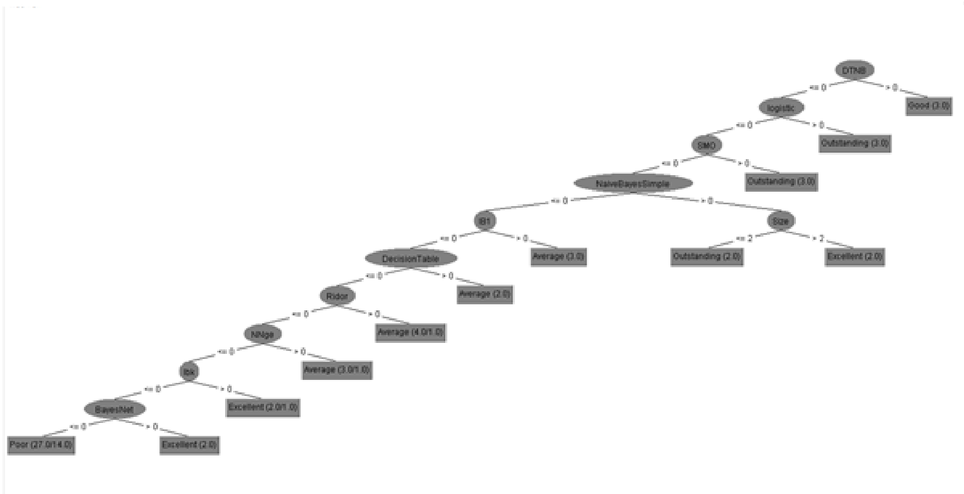

Figure 7: Decision Tree on Correctly Classified Instances on Voting meta layer of Credit Dataset

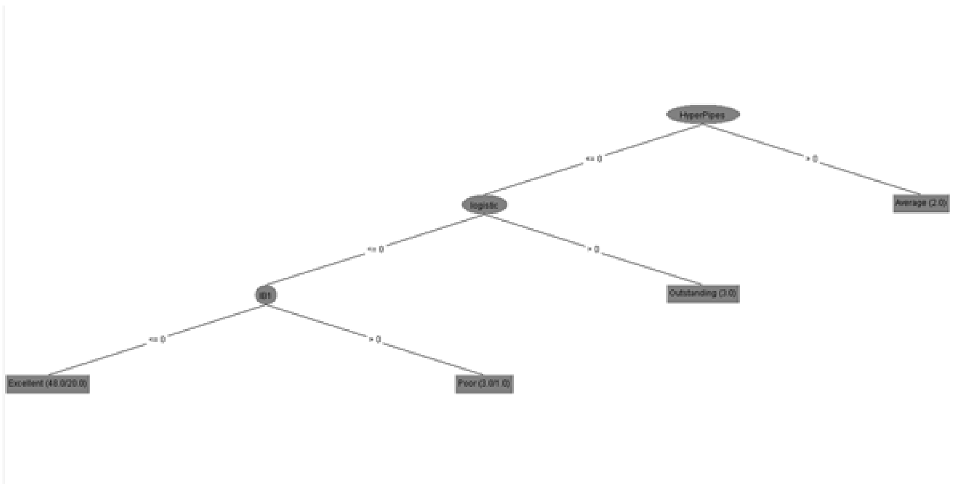

Figure 8: Decision Tree on Root Mean Squared Error on Voting meta layer of Credit Dataset

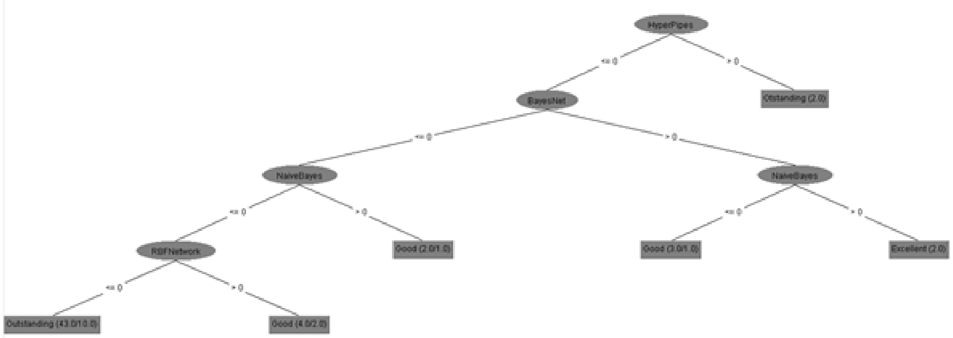

Figure 9: Decision Tree on Time Taken on Stacking meta layer of Credit Dataset

It has been observed from the analysis of the Credit dataset that it provide "Poor" results quite often in the presence of base classiifers- OneR and IB1. On the other hand, overall the performane measures get better in the presence of Bayes family.

5.3 Case Study 3: Medical dataset

This dataset represents the polynomial categorical class. In this project, the Medical dataset generates the following decision trees on the different performance measures respectively.

Decision Tree for Time Taken on Voting meta layer

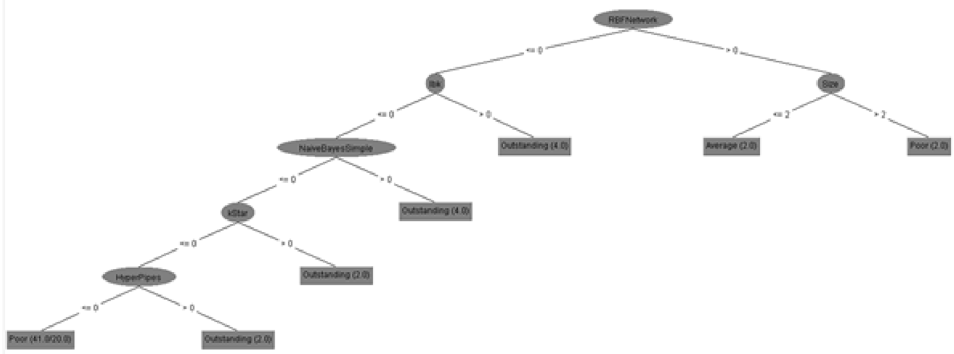

Figure 10: Decision Tree on Time Taken on Voting meta layer of Medical Dataset

Interpretation: As the result (Figure 10) represents that the absence of RBF Network, IBK, Naive Bayes Simple, Kstar, and HyperPipes respectively, probably gives the Poor result. However, it gives the Outstanding result in the presence of any of the these classifier individually. It also represents that the presence of RBF Network, probably generates the Poor result, if size greater than 2.

Decision Tree for Kappa Statistics on Voting meta layer

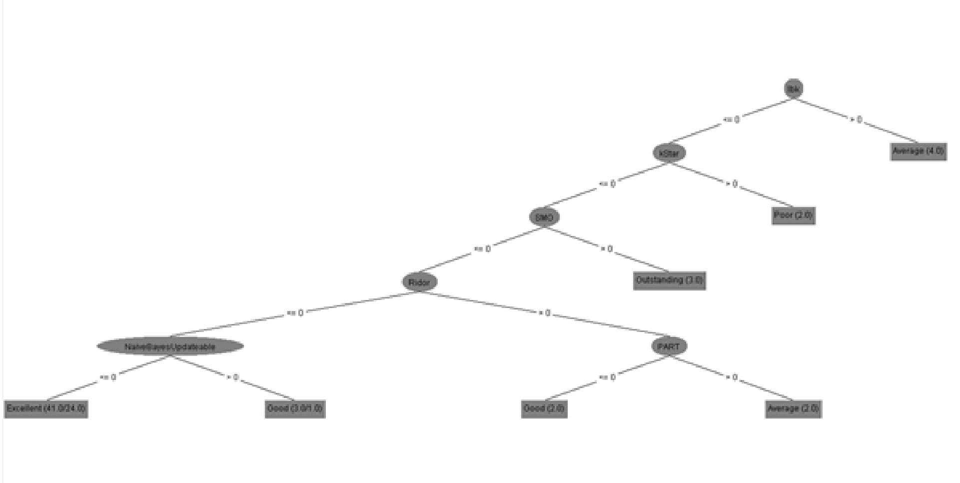

Figure 11: Decision Tree on Kappa Statistics on Voting meta layer of Medical Dataset

Interpretation: As the result (Figure 11) represents that the absence of IBK, Kstar, SMO, Ridor and Naive Bayes Updateable respectively, probably gives the Excellent result. However, it gives the Outstanding result in the presence of SMO individually. It also represents that the presence of Kstar generates the Poor result.

Decision Tree for Correctly Classified Instances on Voting meta layer

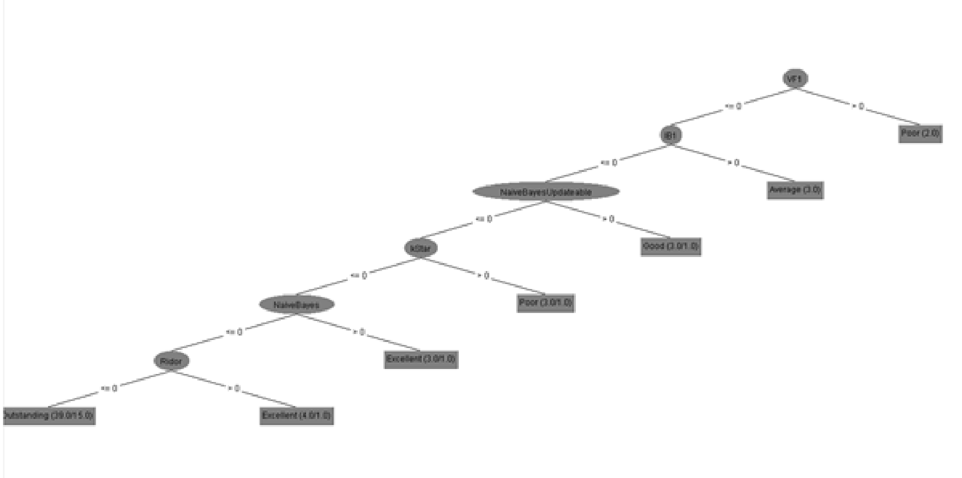

Figure 12: Decision Tree on Correctly Classified Instances on Voting meta layer of medical Dataset

Interpretation: As the result (Figure 12) represents that the absence of VF1, IB1, Naive Bayes Updateable, Kstar, Naive Bayes and Ridor respectively, probably gives the Outstanding result. However, it gives the Excellent result in the presence of Naive Bayes and Ridor individually. It also represents that the presence of Kstar and VF1 individually, probably generates the Poor result.

Decision Tree for Root Mean Squared Error on Voting meta layer

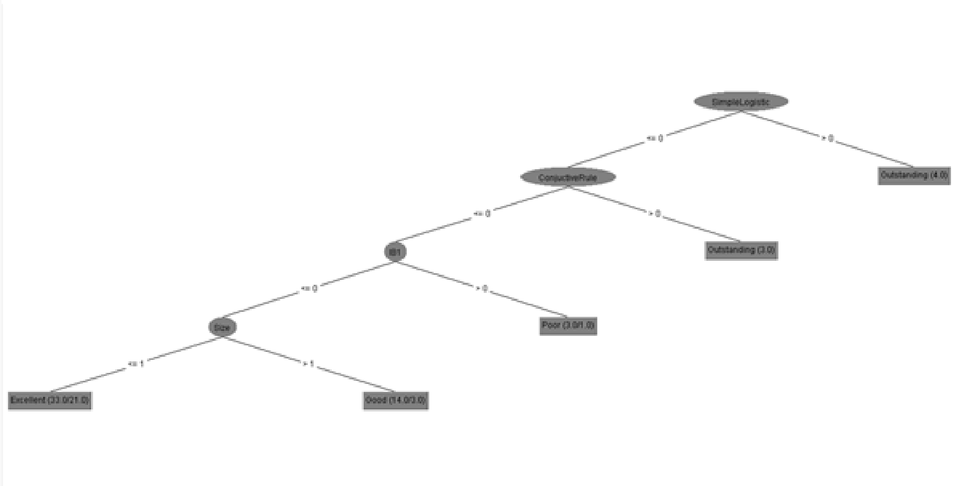

Figure 13: Decision Tree on Root Mean Squared Error on Voting meta layer of Medical Dataset

Interpretation: As the result (Figure 13) represents that the absence of Simple Logistic, Conjuctive Rule, IB1, respectively, probably gives the Excellent result. However, it gives the Outstanding result in the presence of Conjuctive Rule individually. It also represents that the presence of IB1 individually, probably generates the Poor result.

Decision Tree for Time Taken on Stacking meta layer

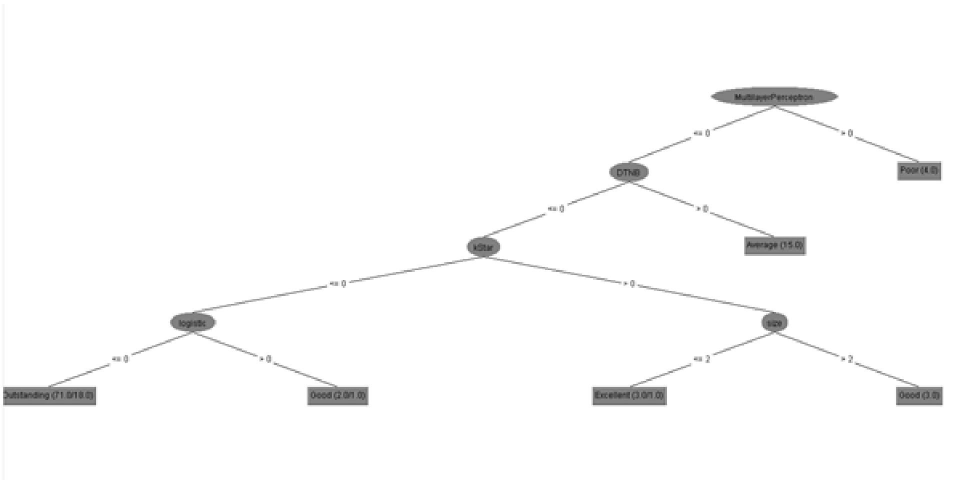

Figure 14: Decision Tree on Time Taken on Stacking meta layer of Medical Dataset

Interpretation: As the result (Figure 14) represents that the absence of BayesNet, Naive Bayes, Hyper Pipes respectively, probably gives the Poor result. However, it gives the Outstanding result in the presence of Hyper pipes individually and the Excellent result in the presence of Naive Bayes individually.

It has been observed from the analysis of the Medical dataset that it provide "Poor" results quite often in the presence of base classiifers from the "Lazy" family, specially OneR and IB1. However, the performane measures get better in the presence of Bayes family for this particular dataset.

6 Conclusion and Future Work

The growth of the data in every sector has been surpassing the previous record almost in every hour. Hence the pressure of competition gives birth to the demand of some additional knowledge every second. To meet this requirement in MCS, the procedure should be so proficient so that it can easily pick out which classifier in base classifiers is appropriate for which subset of dataset. This project finds which characteristics facilitate the selection method of base classifiers in a multiple classifier system with Meta learner. This task was pretty complex as there were so many intricate factors associated with this; like, there is a high usage of statistics right from selecting the features to interpreting the results through the important performance matrices. Also there are different kinds of tools used in this project like R Studio, Weka, Rapidminer, IBM SPSS etc. which require a significant amount of hands-on time. All these multifaceted tasks make this project intricate.

The related work has confirmed the different proportions of this project. The review starts from doing the comparison between MCS and Single classifiers and finds MCS more advantageous. Further, the literature review analyses the specific domain (Remote Sensing of Images), which has been using the MCS in an effective and efficient manner. The work done in this domain provides different ideas for applying novel methods. The literature review goes on to elaborate the dominance of stacked generalization over several other ensemble methods. The last part of the literature review outlines the growth of MCS and signifies how the different scientists sorted out the different issues of MCS and improved the expert system.

The core part of the methodology has been established with the help of the literature review, consisting of how to setup the frame work of this experimentation. Then the next phase of the methodology outlines how different experiments are going to be done in specific order. The last phase deals with the analysis of the results by building the J48 decision trees.

The evaluation advocates some core findings of this project; it is not necessary that multiple classifiers in the base layer always provide the better result every time, in comparison with a single classifier. As it has been observed in case study 3 of Medical Dataset Decision Tree on Root Mean Squared Error on Voting Meta layer. Moreover, there are some certain occasions where if the size of base classiifers is more than two on some particular datastes , it provide poor results. A single Base classifier, Simple Logistic performs outstandingly in that experiment. There are few more cases in case study that substantiate this fact. However in the entire analysis, there are very few cases in which single classifier perform outstandingly. In the absence of some specific classifiers for specific datasets, MCS work always Outstanding; such as, IB1 and Decision Stumps. There are some specific classifier families for the selected datasets, which always provide the poor results; such as, "Lazy". Furthermore, Vote Meta layer provides a better response in terms of diversity in performance measures on applying different datasets.

Now, this project is able to provide the reasons behind selecting or not including some stringent base classifier on a particular dataset in the MCS, as these reasons optimally intensify the performance of the MCS. Also, this project verify that in both of the Meta learners- Voting and Stacking, Voting is more trustworthy in terms of response from performance measures. This work can be extended and more productive in future by comprising more dataset of different variety and experimenting with more Meta learners in the Meta layer. Increment in the size of base classifiers can be another opportunity for intensify the pefrformance of the MCS in the future. The natural progression of this work attempts to provide a fruitful contribution in the advancement of Data Mining.

Acknowledgements

The accomplishment of this thesis could only be possible because of my supervisor Dr. Simon Caton. He always help me in unravelling the confusions and guide me to approach the final result of the thesis. He motivated me at very phase of my thesis. Without his guidance and backing, I could not finish this final work in my master study. Thank you very much, Professor!

Furthermore, I would like to show my grateful feeling to my parents Mrs. and Mr. B.P. Singh Tomer, with whose encouragement I accomplish my master study on time. Thank you very much, Mummy and Papa!

References

Arruti, A and Mendialdua, I and Sierra, B and Lazkano, E and Jauregi, E ((2014)). Expert Systems with Applications One method : NOV @, **41**: 6251–6260.

Ayodele, Taiwo Oladipupo ((2010)). Types of machine learning algorithms, *New Advances in Machine Learning* pp. 19–49.

Basu, Tanmay ((2012)). Effective Text Classification by a Supervised Feature Selection Approach.

Beitia, Mendialdua ((2015)). Contributions on Distance-Based algorithms , Multi-Classifier Construction and Pairwise Classification, **2015**(April).

Biggio, Battista and Fumera, Giorgio and Roli, Fabio, d. . . ((2009)). Multiple Classifier Systems for Adversarial Classification Tasks, *Multiple Classifier Systems* **5519**: 132–141.
URL: *http://dx.doi.org/10.1007/978-3-642-02326-2_14*

Cavalin, Paulo R. and Sabourin, Robert and Suen, Ching Y. ((2013)). Dynamic selection approaches for multiple classifier systems, *Neural Computing and Applications* **22**(3-4): 673–688.

Chen, Minjun and Shi, Leming and Kelly, Reagan and Perkins, Roger and Fang, Hong and Tong, Weida ((2011)). Selecting a single model or combining multiple models for microarray-based classifier development?–a comparative analysis based on large and diverse datasets generated from the MAQC-II project., *BMC bioinformatics* **12 Suppl 1**(Suppl 10): S3.
URL: *http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3236846&tool=pmcentrez&render*

Chen, Y., Wong, M. L. and Li, H. ((2014)). Applying Ant Colony Optimization to configuring stacking ensembles for data mining, *Expert Systems with Applications* **41**(6): 2688–2702.
URL: *http://dx.doi.org/10.1016/j.eswa.2013.10.063*

Demšar, Janez ((2006)). Statistical Comparisons of Classifiers over Multiple Data Sets, *The Journal of Machine Learning Research* **7**: 1–30.
URL: *http://dl.acm.org/citation.cfm?id=1248547.1248548*

Domingos, P. (2012). A few useful things to know about machine learning, *Communications of the ACM* **55**(10): 78.

Fernández-Delgado, Manuel and Cernadas, Eva and Barro, Senén and Amorim, Dinani and Amorim Fernández-Delgado, Dinani ((2014)). Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?, *Journal of Machine Learning Research* **15**: 3133–3181.

Giorgio Valentini , Francesco Masulli ((2002)).

Ko, Albert Hung Ren and Sabourin, Robert ((2013)). Single Classifier-based Multiple Classification Scheme for weak classifiers: An experimental comparison, *Expert Systems with Applications* **40**(9): 3606–3622.
URL: *http://dx.doi.org/10.1016/j.eswa.2012.12.067*

Marqués, A. I. and García, V. and Sánchez, J. S. ((2012)). Exploring the behaviour of base classifiers in credit scoring ensembles, *Expert Systems with Applications* **39**(11): 10244–10250.

Mendialdua, Iñigo and Arruti, Andoni and Jauregi, Ekaitz and Lazkano, Elena and Sierra, Basilio ((2015)). Neurocomputing Classi fi er Subset Selection to construct multi-classi fi ers by means of estimation of distribution algorithms, **157**: 46–60.

Mukhopadhyay, Anirban and Maulik, Ujjwal and Bandyopadhyay, Sanghamitra and Coello, Carlos Artemio Coello ((2014)). A survey of multiobjective evolutionary algorithms for data mining: Part i, *IEEE Transactions on Evolutionary Computation* **18**(1): 4–19.

Seewald, Alexander K. ((2002)). How to Make Stacking Better and Faster While Also Taking Care of an Unknown Weakness, *Icml* pp. 554–561.

Son, Hyojoo and Kim, Changmin and Hwang, Nahyae and Kim, Changwan and Kang, Youngcheol ((2014)). Classification of major construction materials in construction environments using ensemble classifiers, *Advanced Engineering Informatics* **28**(1): 1–10.
URL: *http://dx.doi.org/10.1016/j.aei.2013.10.001*

Ting, Kai Ming and Witten, Ian H. ((1999)). Issues in stacked generalization, *Journal of Artificial Intelligence Research* **10**: 271–289.

Tulyakov, Sergey and Jaeger, Stefan and Govindaraju, Venu and Doermann, David ((2008)). Review of classifier combination methods, *Studies in Computational Intelligence* **90**(Figure 1): 361–386.

Yang, Bin and Cao, Chunxiang and Xing, Ying and Li, Xiaowen ((2014)). Automatic Classification of Remote Sensing Images Using Multiple Classifier Systems, *Mathematical Problems in Engineering* **2015**.

Yang, Haibo and Zhao, Hongling and Wang, Zongmin ((2010)). Remote sensing classification based on hybrid multi-classifier combination algorithm, *ICALIP 2010 - 2010 International Conference on Audio, Language and Image Processing, Proceedings* (X): 1688–1692.