Advances and Applications in Mathematical Sciences
Volume 19, Issue 10, August 2020, Pages 1017-1028

MACHINE LEARNING CLASSIFIERS, META CLASSIFIERS COMPARISON AND ANALYSIS ON BREAST CANCER AND DIABETES DATASETS

VIDUSHI and MANISHA AGARWAL

Research Scholar, Banasthali Vidyapith
Rajasthan-304022, India
E-mail: vidushi.mtech@gmail.com

Associate Professor, Banasthali Vidyapith
Rajasthan-304022, India
E-mail: manishaagarwal18@yahoo.co.in

Abstract

Health sector is the most important and sensible area and machine learning is currently touching every sphere of technology. This paper is using different classification algorithms of this learning on well known datasets diabetes and breast cancer. Classification is the process of classifying data in correct class. Various classification algorithms are present in weka tool to classify the data e.g. naïve bayes, neural network, decision table, *j* 48, decision tree, etc. All these algorithms have their own benefits and limitations. This paper presents the comparison of different classifiers, meta-classifiers or multiple classifiers and shows that how efficient or accurate they are. In our analysis the best result showed by the combination of different classifiers like multilayer perceptron and Naïve bayes algorithm for Diabetes data set and the combination of Naïvebayes and J48 in Breast Cancer data set.

1. Introduction

In the today's era machine learning [14], [15] is the vast area to study and useful in extracting or classifying the meaningful information from huge database. In data mining classification is based upon machine learning algorithm. Classification is technique of classifying unknown dataset into a class based upon its relevance features and it is one of the important features of data mining.

2010 Mathematics Subject Classification: 68T05.
Keywords: meta classifier, Naïve bayes, machine learning, decision tree, weka, J48, multi layer perceptron.
Received November 9, 2019; Accepted August 2, 2020

In this paper we addressed the issue associated with individual classifier. The combination of more than one classifier called Meta Classifier [4], [11]. Through Meta classification the usage of combination of multiple classifiers is indicated [9]. We have different multiple Classifiers. These can be divided into 3 major categories. 1. Ensembles (Bagging or Boosting), 2. Voting, and 3. Stacking. In this paper we applied Meta classifiers VOTING technique on 2 known datasets Cancer and Diabetes. We used three parameters like correctly classified, incorrectly classified instance and ROC Area. There result is on the base of different classifiers and combination of different classifiers.

For the experimental work DIABETES and BREAST-CANCER datasets have taken from UCI repository available on web.

Bagging [19] and Boosting [20] are homogenous classifier and on the other hand voting or Rule fixed aggregation, and Stacking are heterogeneous classifier. Bagging build multiple model of same type using different datasets while boosting build different model which is use to learn fixing error in model.

Data mining is the process of extracting useful information from data. It is not a single step process; a number of steps take place in this process. Every step itself a complete process and takes a lot of time to complete. Different tools are used to perform this task; Weka [18] is one of them. Matlab can also be used to complete this task. The steps taken by every tool is almost same. The accuracy of data mining depends on the following factors:

Tool used,

Technique used,

Language or Algorithm used,

System configuration, etc.

The best result is judge by the Time complexity and space complexities of various algorithms. Time complexity is defined as the time taken by the algorithm and space complexity is defined by space taken by the program algorithm. The following figure 1 shows the different steps for evaluating the model for classification.

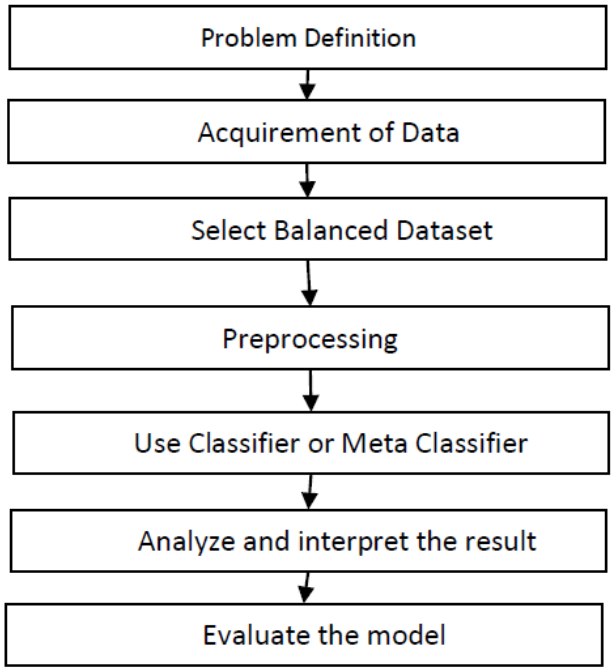

Figure 1. Steps to evaluate the model for classification.

The steps mentioned in above figure 1 are take place in the process of classification. Good algorithm or program takes less time and space to perform the required task. A new algorithm also can be implemented in performing the task. In Weka new classifier can be added to get the desired result. Java language is used to perform any task in Weka tool. For example KNN classifier present in Weka tool used for classification as well as clustering process. This KNN classifier used EUCLIDIAN DISTANCE formula to classify the data. We can use a different formula to classify the task and get the better result.

2. Literature

The most active and attractive research area from the last many years is the one involving different methodologies and systems for the combination of multiple predictive models or Meta Classifiers [1]. Weka tool in Machine learning have different classifier and Meta classifier. These Meta classifiers are: First type is Ensembles (Bagging or Boosting), second type is Voting, and

third type is Stacking. Models that have been derived from different executions of the same learning algorithm are often called Homogeneous and Models that have been derived from running different learning algorithms on the same data set are often called Heterogeneous[1].In this paper we have used Voting technique that is Heterogeneous. Voting build different model and calculate mean for the prediction of result. Bootstrap aggregation is also called Bagging in machine learning ensemble Meta algorithm [7], [13].

Bagging is suitable for Classification improvement and which is combination of classification of randomly generated training set. Its best suited for small size dataset. Various approaches of Meta learning are based on the dataset characteristics which automatically rank the classifier [3]. Different classification method decorate, bagging, multiclass classifier and multi boost *AB* are compared [5]. An overview of boosting algorithms is presented to build ensembles of classifiers. Variant of basic boosting technique and its comparison for supervised learning [2]. In stacking several classifiers are combined using the stacking method [10].

3. Weka Tool

Weka [16], [17] is a tool used in Machine Learning for performing various tasks in Data Mining. It is a tool used for compilation of various machine learning algorithms. The algorithms can be used in following ways:

Data sets can be uploaded directly from different repositories present, Or by writing own Java code. This tool is used for performing various tasks:

Data pre-processing,

Classification,

Regression,

Clustering,

Association,

Rules, and Visualization.

New machine learning schemes can be evolved. This tool made up of following keys [8]:

Explorer [8]: from where data can be explored.

Experimenter [8]: experiments are executed from here.

Knowledge flow [8]: it has explorer function with drag and drop interface.

Simple CLI [8]: provides command line interface for implementation of Weka commands for operating systems [8].

Single Classifier Approaches [6].

To classify the data following approaches are used:

Multi Layer Perceptron (Rule approaches),

Naïve bayes,

Decision Tree,

Neural Network,

K-nearest neighbor classifiers,

Artificial neural network,

Genetic classifier,

Logistic regression,

Support Vector Machine,

Discriminant analysis,

Logical statements (ILP),

Meta Classifier [6], [13].

It is also known as Multiple Classifier. It is a combination of single classifiers. Ensemble, Bagging or Boosting, and Voting are different techniques for combining more than one classifier to get better result. Result of the Meta classifiers are depends on selection of different single classifiers. Single classifiers have their own merits and demerits. How to choose different single classifiers so that better result can be obtained are the main and the most important task. The figure 2 below shows that Meta or multiple classifiers is a combination of more than one classifier. It also shows that Meta classifier working is depends on the merits or demerits of individual classifiers whose combination makes any multiple classifiers. This study combines the various classifiers like MultiLayer Perceptron (MLP), Naïve

bayes (NB), Decision Tree (DT), and J48and compare the result of all these Meta or multiple classifiers. The result totally depends on combination of different classifiers because all single classifiers have their own merits and demerits. The Time complexity and space complexity depend on the each single classifier which is used in the combination. For example Meta classifier (MLP+NB) complexity depends on the complexities of MLP and NB. Their combination can show better result as well as worst result also.

The basic concept of Meta Classifier is shown in below figure 2.

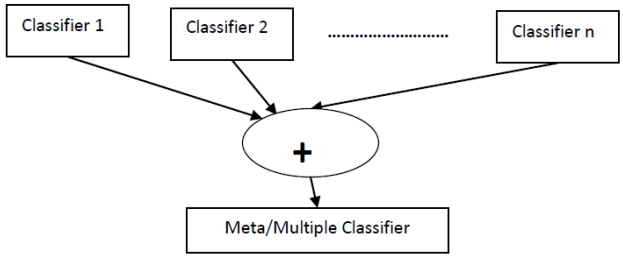

Figure 2. Meta or Multiple Classifier.

So it is very important and difficult task to choose the correct classifiers to get the desired result and to utilize the merits of individual classifiers. It might be possible that single classifier gives better result but when we perform Meta classification it gives comparatively low result. In the same way low result showing single classifier can give the better result in combination.

4. Result and Discussion

Different approaches can be used for single/individual classifiers to classify the data. The result varies by using different classifiers. Result depends on:

Type of classifier used, and

Data set.

The below table 1 shows the comparative result of combination of various classifiers for DIABETES and BREAST-CANCER datasets. Further histograms are plotted using the above table results.

Diabetes data set result shows-

(i) Multi Layer Perceptron (MLP), Naïve Base (NB) are comparatively more correctly classify (76.8229).

(ii) Multi Layer Perceptron (MLP), Naïve Base (NB) have highest ROC (0.826).

BREAST-CANCER dataset result shows-

(i) J48, Decision Tree are comparatively more correctly classify (75.5245).

(ii) Naïve Base, J48 have highest ROC (0.703).

The below table 1 shows comparison between the different classifiers result.

Table I. Different Classifier Result and Comparison:

DATA SETS		Breast Cancer			Diabetes	
MATA CLASSIFIER (VOTE)	Correctly Classify (in %)	Incorrectly Classify (in %)	ROC (in %)	Correctly Classify (in %)	Incorrectly Classify (in %)	ROC (in %)
MLP, DT, NB, J48	69.2308	30.7692	0.69	76.1719	23.8281	0.82
DT, NB	73.0769	26.9231	0.7	76.3021	23.6979	0.81
MLP, J48	67.4825	32.5175	0.62	73.4375	26.5625	0.81
MLP, NB	67.1329	32.8671	0.68	76.8229	23.1771	0.8
MLP, DT	67.4825	32.5175	0.65	75.1302	24.8698	0.81
J48, DT	75.5245	24.4755	0.67	73.3073	26.6927	0.79
NB, J48	73.7762	26.2238	0.7	75	25	0.82
SINGLE CLASSIFIERS						
MLP	64.6853	35.3147	0.62	75.3906	24.6094	0.79
DT	73.4266	26.5734	0.66	71.225	28.776	0.77
NB	71.6783	28.3217	0.7	76.3021	23.6976	0.82
J48	75.5245	24.4755	0.58	73.8281	26.1719	0.75

The above table 1 is the detail analysis of both breast cancer and diabetes datasets using different machine learning classification algorithms and also the different combinations of these classification algorithms. The research is the experimental analysis of all different classifiers on the well known datasets and the below figure 3, figure 4, figure 5, and figure 6 are the graphical representation of this research analysis work. The below figure 3 represents the different Meta classifiers [12] work on breast cancer datasets.

The below figure 3 shows pictorially the different meta classifiers result applied on breast cancer dataset.

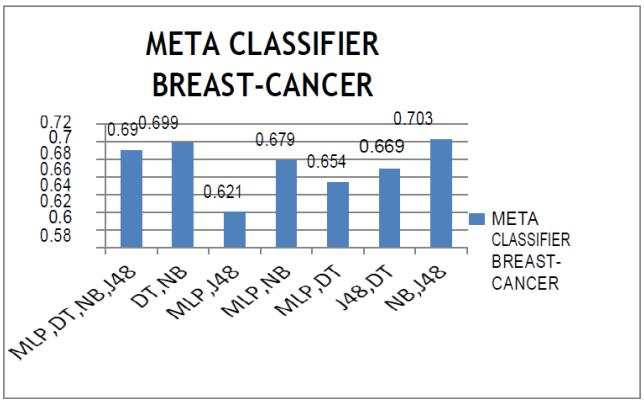

Figure 3. Different Meta classifiers result applied on breast cancer dataset.

The analysis of this study shown in below figure 4 the result of individual classifiers applies on breast cancer dataset.

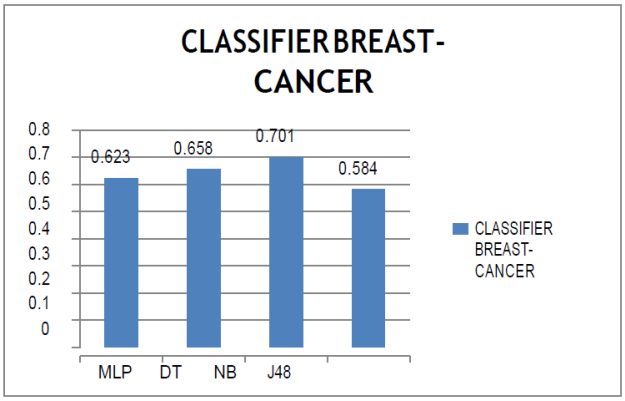

Figure 4. Individual Classifiers result applied on breast cancer dataset.

In the above two figures 3 and 4, the analysis of classifiers and Meta classifiers are shown using one of the well known dataset breast cancer in health sector. This dataset is taken from the UCI repository. Scientists or researchers are currently emphasis on the critical health care sector. Various techniques are already available to work in this field. This research uses the machine learning techniques and gets the desired classification results. Self learning and improvement with experience like human is the main motive of machine learning and it is also getting success in analysis work. This is one of the prominent technologies coming in the future world. It can also be said that it is the future of analysis world and can provide the solution to number of real time problems. In the above Figure best result shows with Meta-Classifier combination of Naïve base BREAST-CANCER dataset.

In the below two figure 5 and figure 6, the analysis of classifiers and Meta classifiers are shown using one of the well known and famous datasets diabetes in crucial and critical health sector.

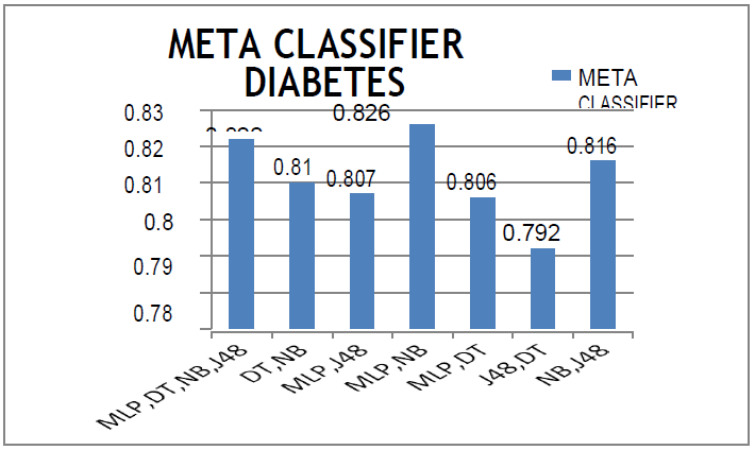

Figure 5. Different Meta classifiers result applied on diabetes dataset.

The analysis of this study shown in below figure 6 the result of individual classifiers applies on diabetes dataset.

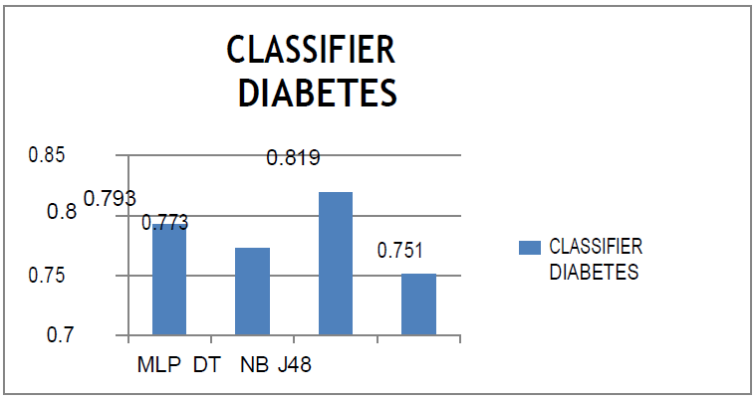

Figure 6. Individual Classifiers result applied on breast cancer dataset.

In the above Figure best result shows with Meta-Classifier Figure 6. Individual Classifiers result applied on breast cancer combination of Naïve Base (NB) for DIABETES dataset.

Result shows Meta Classifier can give much better result than single classifiers but the combination of different classifiers should be correct. The result of Meta Classifiers totally depends on its combination because all classifiers have its own benefits and limitations. The result matters how we use the merits of all classifiers in a single classifier called Meta Classifier.

5. Conclusion and Future Work

In the paper classifiers and Meta classifiers are discussed. How Meta classifier can be used to classify the data and also compared them with the individual classifiers. It is also discussed that Meta classifiers positives and negative points are depends on individual classifiers whose combinations make them. To work with Meta classifiers it is important to know about the classifiers because they directly affect the Meta classifiers. The drawbacks of Meta classifiers are the drawback of individual classifier. So it is vital to select the right classifier and work on it. In the future for coming research this study will be important to work on classification task and in the coming time researchers can treat this work as a base for developing new more better classifier or meta classifier to get more accurate result as per features of different datasets. This research used the Weka tool to get the experimental analysis result and to get more improved or accurate result Java language

API can be used. In the coming research more better tools like Matlab, R-tool or languages like python can be used.

References

[1] G. Tsoumakas, I. Katakis and I. Vlahavas, Effective voting of heterogeneous classifiers, In European Conference on Machine Learning, pp. 465-476, Springer, Berlin, Heidelberg, September 2004.

[2] A. Ferreira, Survey on boosting algorithms for supervised and semi-supervised learning, Institute of Telecommunications (2007).

[3] N. Bhatt, A. Thakkar and A. Ganatra, A survey and current research challenges in meta learning approaches based on dataset characteristics, International Journal of Soft Computing and Engineering 2(10) (2012), 234-247.

[4] Q. Sun and B. Pfahringer, Pairwise meta-rules for better meta-learning-based algorithm ranking, Machine Learning 93(1) (2013), 141-161.

[5] P. Kalaiselvi and C. Nalini, A comparative study of meta classifier algorithms on multiple datasets, International Journal of Advanced Research in Computer Science and Software Engineering 3(3) (2013), 654-659.

[6] G. Krempl, I. Žliobaite, D. Brzeziński, E. Hüllermeier, M. Last, V. Lemaire and J. Stefanowski, Open challenges for data stream mining research, ACM SIGKDD Explorations Newsletter 16(1) (2014), 1-10.

[7] G. Michael, A. Kumaravel and A. Chandrasekar, Detection of malicious attacks by Meta classification algorithms, International Journal of Advanced Networking and Applications 6(5) (2015), 2455.

[8] D. S. C. Nascimento, A. M. P. Canuto and A. L. V. Coelho, Multi-label meta-learning approach for the automatic configuration of classifier ensembles, Electronics Letters 52(20) (2016), 1688-1690.

[9] T. S. Devi and K. M. Sundaram, A comparative analysis of meta and tree classification algorithms using WEKA, Int. Res. J. Eng. Technol. (IRJET) 3(11) (2016), 77-83.

[10] G. Ayyappan, C. Nalini and A. Kumaravel, Construction of Meta Classifiers for Academic Research Data from Social Networks, International Journal of Civil Engineering and Technology 8(3) (2017).

[11] C. H. Lin, C. D. Kan, J. N. Wang, W. L. Chen and P. Y. Chen, Cardiac arrhythmias automated screening using discrete fractional-order integration process and meta learning based intelligent classifier, IEEE Access 6 (2018), 52652-52667.

[12] M. Chen and J. Shang, Recursive Spectral Meta-Learner for Online Combining Different Fault Classifiers, IEEE Transactions on Automatic Control 63(2) (2017), 586-593.

[13] F. D. Acosta-Escalante, E. Beltrán-Naturi, M. C. Boll, J. A. Hernández-Nolasco and P. P. García, Meta-classifiers in Huntington's disease patients classification, using iPhone's movement sensors placed at the ankles, IEEE Access 6 (2018), 30942-30957.

[14] S. Ledesma, M. A. Ibarra-Manzano, E. Cabal-Yepez, D. L. Almanza-Ojeda and J. G. Avina-Cervantes, Analysis of data sets with learning conflicts for machine learning, IEEE Access 6 (2018), 45062-45070.

[15] D. Côté, Using machine learning in communication networks, Journal of Optical Communications and Networking 10(10) (2018), D100-D109.

[16] R. Meenal, A. I. Selvakumar, K. Brighta, S. C. J. Joice and C. P. Richerd, Solar radiation resource assessment using WEKA. In 2018 2nd International Conference on Inventive Systems and Control (ICISC), January, 2018, pp. 1038-1042, IEEE.

[17] J. Khalfallah and J. B. H. Slama, A Comparative Study of the Various Clustering Algorithms in *E*-Learning Systems Using Weka Tools, In 2018 JCCO Joint International Conference on ICT in Education and Training, International Conference on Computing in Arabic, and International Conference on Geocomputing (JCCO: TICET-ICCA-GECO) (2018, November), (pp. 1-7). IEEE.

[18] J. M. Alonso and A. Bugarín, ExpliClas automatic generation of explanations in natural language for WEKA classifiers, In 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (2019, June), (pp. 1-6). IEEE.

[19] D. C. Sujatha and J. G. Jayanthi, MetaLASH Tree Bagging at Meta Level Using LASSO Regression Hoeffding Tree for Streaming Data, In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (2019, April), (pp. 1047-1051). IEEE.

[20] M. Rafati, S. R. Qasemi, A. Nejati and P. Amiri, Agm-boosting 3-5 GHz noise-cancelling LNA, In 2019 27th Iranian Conference on Electrical Engineering (ICEE) (pp. 376-379). IEEE.