Dominated Consensus - A new technique to combine classifiers for continuous data

Nishtha Sharma[1], Dr. Amit Sharma[2]

[1] *M.Tech Scholar, Computer Science & Engineering Department,*
Vedant College of Engineering and Technology, Kota, Rajasthan, India

[2] *Professor, Computer Science & Engineering Department,*
Vedant College of Engineering and Technology, Kota, Rajasthan, India

Abstract- *Popular combined classification methods are either voting based or mathematical combination based. This paper proposes a new scheme, named Dominated Consensus, which though similar to voting, takes inspiration from effects of mathematical combination. A user-defined parameter called window_size is introduced to widen the scope of consensus. In case, a consensus between participating classifiers is not achieved, decision of one of the dominating classifiers is accepted. The combination is implemented over Gaussian Naive Bayes and Logistic regression classifiers. Comparison with other individual and combined classifiers is presented to prove the effectiveness of proposed scheme. Also, applications where the proposal is more useful are indicated.*

Keywords: *Classification of continuous data, Gaussian naive Bayes, Logistic Regression, combining classifiers*

I. INTRODUCTION

Among the data mining techniques, classification and clustering hold great importance. Classification is an unsupervised data mining task aimed at assigning objects to one of the several predefined classes/categories. The task of classification encompasses applicability in various fields of science, few of them being spam filtering/ document classification, text classification, MRI scan categorization, identification of galaxies and more. Over the years, much work has been done in designing classifiers suitable according to the requirements of the classification process and applications. However, much work done uses discrete data. Handling continuous data is primarily done through clustering, the unsupervised approach towards learning data. Cases requiring only supervised learning approach for continuous data call for classification approaches able to deal with this aspect of data.

Data analysis applications majorly deal with continuous data. This is due to two reasons: the data collected is through analog techniques or are real values in actuality; and secondly continuous data contains more information than discrete data. Hence, classification of continuous data is an important field of research. All kinds of classification like Naïve Bayes, k-Nearest Neighbor, Support Vector Machines, etc have been proved good for tasks like document classification which involve discrete data only. Few classifiers have been demonstrated for continuous data. Continuous data is more precise, more informative, more time consuming and removes the need to estimate or round-off the measurement. Classification of continuous data is therefore a harder task than discrete data classification. Applications where continuous data classification is found to give promising results are character recognition, handwriting recognition, image segmentation and more [1,2,3].

In spite of the effectiveness claimed for classification algorithms throughout the literature, there still is not one single algorithm which can produce best performance for all the applications and according to all the requirements of the classification method. Also, each classifier has its own shortcomings and advantages making it unique from the other classifiers. Researchers choose classifiers for any classification problem depending upon the application, capabilities and limitations of each individual classifier. As a more general solution to the problem and in order to delimit the scope of classification to only a few applications, researchers suggest combination of various individual classifiers [4,5,6]. Combination of classifiers helps overcome the individual challenges of each classifier, selects the best features from the classifiers combined in order to yield better results for broader variety of applications.

Bennett et al [7] used probabilistic combination procedures to build meta-classifiers by considering context-sensitive reliabilities of the classifiers contributing in the combination. For this, reliability indicators and classifier outputs are used to judge the performance of each classifier in different situations. Xue and Converse[8] proposed combination of Maximum Entropy and Error- driven transformation-based learning model for classification of Chinese text with Maximum Entropy. Fragos et al [5] proposed combining learned values of probabilities by Naïve Bayes and Maximum Entropy classifiers using simple arithmetic functions of Sum, Harmonic Mean and Max. It was for text classification only. Kashyap and Buksh [9] proposed improvements to this by modifying maximum entropy before classification, adding one more merging operator along with max and harmonic mean, and the splitting criteria of to be evaluated dataset. Yet, very few researches have tried combining classifiers for Continuous data.

A very closely related work to ours is by Sehgal et al [10]. They proposed combination of Naïve Bayes and Logistic Regression Classifier for a single application to qualitative breast sonography and classify them as malignant or benign. The research pointed out that Naive Bayes and Logistic Regression if combined linearly do not produce any improvement over individual results; but a consensus approach is better.

This paper proposes a combination classifier for continuous data. A combination scheme different from existing ones is developed to overcome the drawbacks. Two very different classifiers - Logistic regression and Gaussian Naïve Bayes – are combined using Dominated consensus scheme. Studying performance over large variety of applications it is suggested when is it advisable to use the proposed combination of classifier.

II. BACKGROUND

Classification of continuous data involves an assumption that the distribution of continuous values of each class follows Gaussian distribution. For training data containing a continuous attribute a of a set of attributes A, the data is segmented class-wise followed by computation of mean and variance of a in each class. Mean of the values is represented through the symbol and variance by. Supposing an observation value v of class c, its class conditional probability can be computed using the following equation of normal distribution involving parameters and

$$P(c) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{x-\mu}{2\sigma}\right)^2} \qquad (1)$$

For an object $= (x_1, x_2, \dots, x_m)$, the posterior probability of it belonging to a class c can then be calculated as the product of all conditional probabilities of the values of individual attributes x_i.

$$P(X|c) = P(x_1|c).P(x_2|c) \dots\dots P(x_m|c) = \prod_{i=1}^{m} P(x_i|c) \quad (2)$$

Such posterior probabilities are computed for any test object towards all possible classes. The class to which the object in query belongs is predicted by selecting maximum value of posterior probability.

Logistic Regression can be considered a variant of linear regression. A linear regression is referred to as logistic regression when all the independent variables are continuous whereas the dependent variable Y is discrete. In the training phase of the classifier, all the continuous values of an object $X = (x_1, x_2, \dots, x_m)$ are taken as independent variables and the class c is taken as a dependent variable, say Y, where Y can be deduced using the following linear equation.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_m x_m \qquad (3)$$

The linear function tries to fit the best values of β into the classification process.

Logistic regression is normally used for predicting binary dependent variables (i.e. involving only two possible outcomes) but is extended to multinomial logistic regression to predict classes having more than two outcomes. Therefore, β values for each class are to be learnt.

The posterior probability of each attribute of object X towards class c can be calculated as

$$P_i(X|c) = \frac{1}{1 + e^{-\beta_i^c x_i}} \qquad (4)$$

The class to which the object in query belongs is predicted by selecting maximum value of this posterior probability.

some popular methods which are used to combine results from different classifiers on same set of training and testing examples are:

- ***Sum:*** The individual item class probabilities from both classifiers are added.
- ***Product:*** The individual item class probabilities from both classifiers are multiplied.
- ***Max:*** Larger of probabilities from both the classifiers is selected as the final probability of a test object towards a class.
- ***Harmonic Mean:*** Harmonic mean of both output probability is taken as final probability of a test object.

III. PROPOSED COMBINATION METHOD

Any direct combination as described above gives a result which can be summarized as best of the two. Sometimes the results may be better than both. But a major challenge of these methods is selecting the best when the posterior probabilities for more than one class are approximately same. In such situations, the performance of the combined classifier may get degraded. In this section, a method is proposed which is not a direct combination scheme. It rather prefers prediction of one classifier over others in case of disagreement. The case of agreement is decided through a user input called window_size. Suppose the predicted class output of a classifier is a list of classes in order of decreasing probabilities. Let c_{1i} be the predicted class by classifier 1 at i^{th} priority and c_{2j} be the predicted class by classifier 2 at j^{th} priority. Then, agreement is decided as following cases:

- ***Case I:*** If $c_{1i} = c_{2j} = c$, predict class c
- ***Case II:*** If $c_{11} = c_{2j}$, for any $j > w$, predict class c_{11} else predict c_{21}
- ***Case III:*** If $c_{21} = c_{1i}$, for any $i > w$, predict class c_{21} else predict c_{11}

The cases II and III are mutually exclusive. That is, for Classifier 1 – domination, Cases I and III are referred. For Classifier 2-domination, Cases I and II are referred. Hence, this method is named as Dominated Consensus. It is illustrated in Fig. 1 for two classifiers.

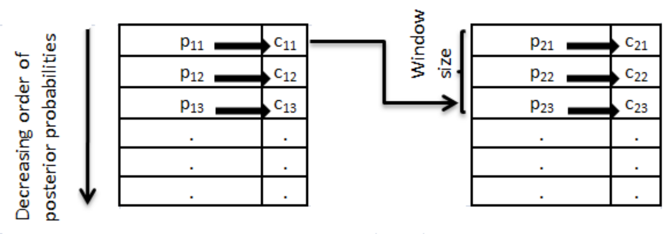

Fig 1 Illustration of agreement for two classifiers with window size =3

Thus, instead of combining the probabilities per class to yield some resultant probabilities, the actual predictions of the classifiers are tallied and a final prediction is made. This particular method can be easily extended to more than two classifiers. The only point of caution is that window_size should be at most one less than the total number of categories possible. The constraint can be expressed as

$$1 \leq window\_size \leq C$$

The proposed scheme of classification can now be summarized into broad steps for training and testing phases.

Training Phase

1. Train Gaussian Naïve Bayes classifier over training samples to learn values of $\square_{ik}$ and $\square_{ik}$ for every class k and attribute i.
2. Train Logistic Regression Classifier over training samples to learn $\square^{c}$ for every class c.

Test phase

1. For query document X, compute posterior probabilities towards each class using (2) and (4). Let these be ρ_1.
2. For query document X, compute likelihood towards every class using (). Let these be $\square_2$.
3. Convert $\square_1$ and $\square_2$ to equivalent ratio which sums upto 1 by

$$P_{1i} = \frac{\rho_{1i}}{\sum \rho_{1i}}, \forall i$$

And,

$$P_{2i} = \frac{\rho_{2i}}{\sum \rho_{2i}}, \forall i$$

4. Combine P_{1i} and P_{2i} through proposed combination method to predict class of query document.

In the test phase, during the combining step, Logistic Regression Classifier is treated as Classifier 1 and Gaussian Naïve Bayes Classifier is treated as Classifier 2. The domination of any of these in the proposed Dominated Consensus method is opted according to the nature of dataset. The conversion of probabilities as in step 3 is necessary (this is never attempted in any other research, and might be the root of degraded performance) so that the probabilities of both classifiers are at same level.

IV. Experiments And Results

A combination classifier with each of the Sum, Product, Harmonic Mean and Max Strategy is implemented as MATLAB program. The proposed classifier is implemented in two variants. First one has domination of Gaussian Naïve Bayes (referred as DC GNB) and in the second one, Logistic Regression dominates (referred as DC LR). Results are recorded here for a variety of datasets so that it can be deduced which classifier is better in a given situation. The experiment sets are formed as per number of categories possible. The performance of classifiers are measured in terms of classification accuracy, measured as

$$Accuracy = \frac{number\ of\ correctly\ classified\ test\ objects}{total\ number\ of\ test\ objects}$$

Wherever the number of categories is 3 or less, there is no role of window_size parameter. Its value is fixed at 2. For other datasets, value of window_size is varied to observe the effect. The datasets used for the experiments are listed in Table 1 with other details. All have been taken from the UCI machine learning repository[11].

Table 1
Details Of The Datasets Used For Experiments

Dataset	Number of Classes	Number of Instances	Number of Attributes	Ratio of Training and Test samples
Iris	3	150	4	80:20
Abalone-1	3	4177	8	80:20
Abalone-2	3	4177	8	75:25
Dermatology	6	366	33	80:20
Mice-1	8	1080	77	80:20
Mice-2	8	1080	77	90:10
Pendigits	10	10992	16	80:20

Fig 2 shows accuracy obtained by different classifiers over Iris dataset. Since, Gaussian Naive Bayes already gives 100% accuracy; a better performance cannot be obtained. Yet, all combinations are tried to observe the effect of combination function. Mathematical combinations of sum, Product, Harmonic mean and Average give 100% accuracy, implying that the conditional probability values of Gaussian Naïve Bayes are much higher than those of Logistic Regression wherever there is disagreement in predictions of both classifiers. Only the DC LR (Dominated Consensus method with domination of Logistic Regression) gives the accuracy as seen in individual Logistic Regression classifier. This is an expected behavior that whenever there is disagreement between the two classifiers, LR dominates the prediction.

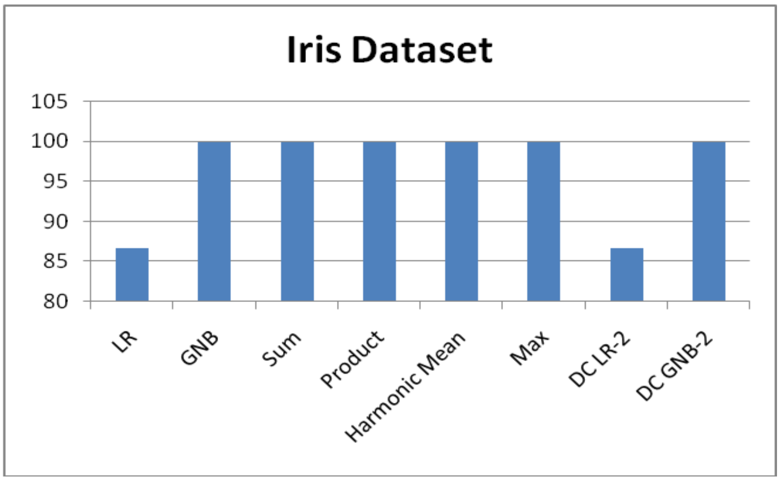

Fig 2 Comparison of accuracy of different classifiers over Iris Dataset

For both Abalone-1 and Abalone-2, there is slight difference in accuracies of individual LR and GNB classifiers which means that the values of probabilities in both will play a major role in deciding the accuracy of a combination classifier. All mathematical combinations, Sum, Product, Harmonic mean and Max classifiers have more accuracy than GNB, but lesser than LR. This indicates that the difference in values of probabilities generated by LR and GNB individually is very low. It also indicates that a consensus can be formed within low values of window_size. Since the value here is fixed at 2, DC-LR behaves as LR and DC-GNB as GNB. Fig 3 and Fig 4 show the accuracy values of all classifiers experimented over abalone datasets.

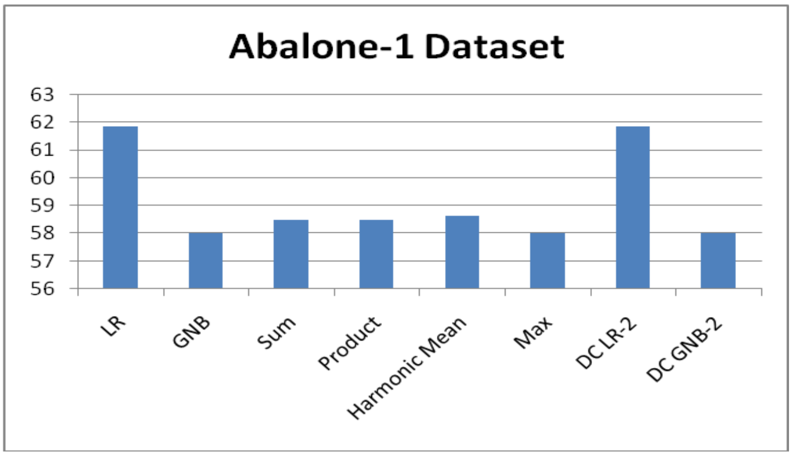

Fig 3 Comparison of accuracy of different classifiers over Abalone-1

Dermatology dataset shows a curious behavior that accuracy of the classifiers do not show any improvement through any of the combinations. Even if window_size in increased from 2 till 4, there is no improvement in the accuracy than the individual LR and GNB classifiers. The values are recorded in Fig 5.

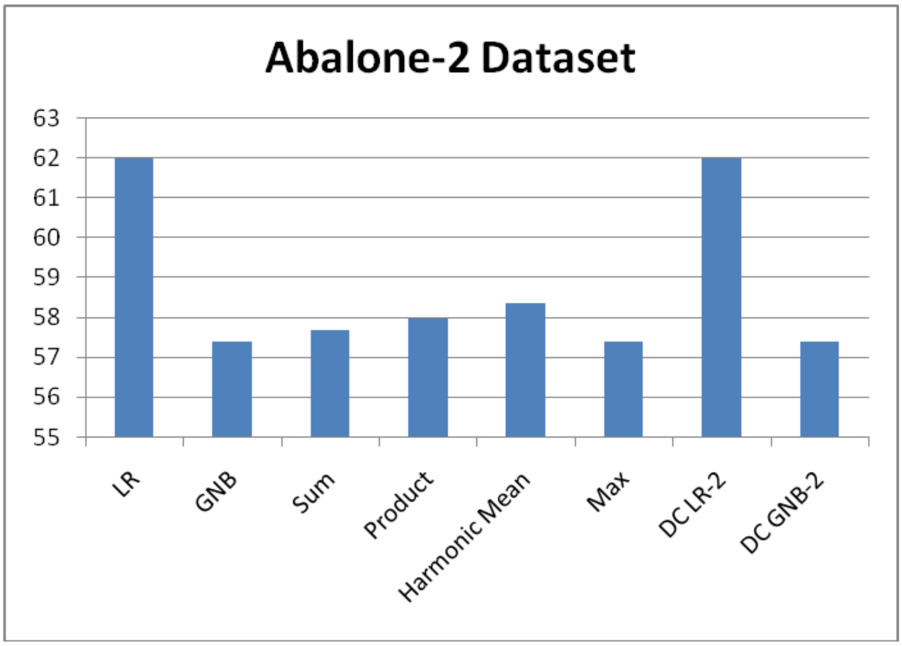

Fig 4 Comparison of accuracy of different classifiers over Abalone-2 Dataset

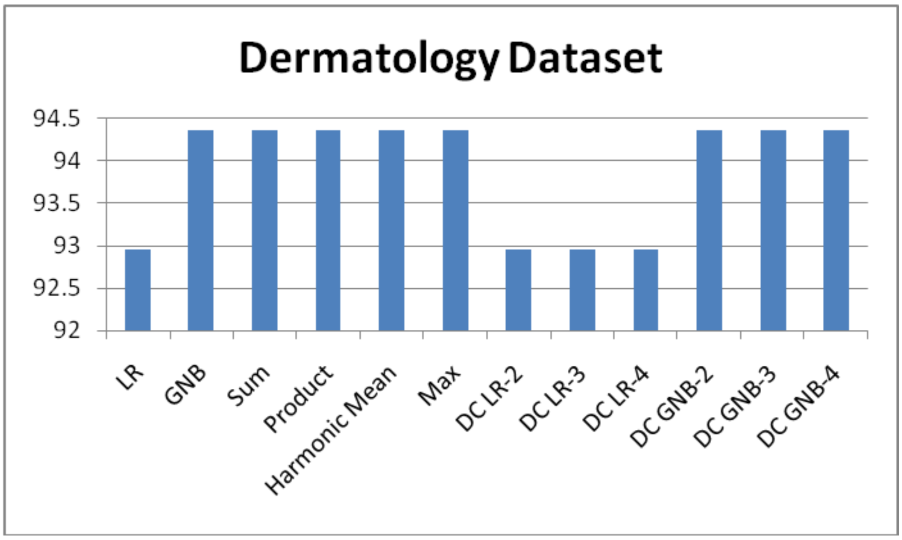

Fig 5 Comparison of accuracy of different classifiers over Dermatology Dataset

For ecoli dataset, the accuracy of LR classifier is much lower than the GNB classifier. Also, the accuracy achieved through mathematical combination classifiers is same as that of LR. This indicates that the probability values produced by LR classifier are much higher than GNB, though they do not give good result, yet are favoured due to mathematical computations. Whereas, the proposed classifier DC-LR achieves much better results than LR itself. For window_size from 2 till 4, accuracy is much higher than LR. The proposed classifier DC-GNB is also better than individual LR, but not better than GNB. As window_size is increased, accuracy falls in DC-GNB, because it tries to accommodate opinion of LR which has not produced good results individually. The recorded values are shown in Fig 6.

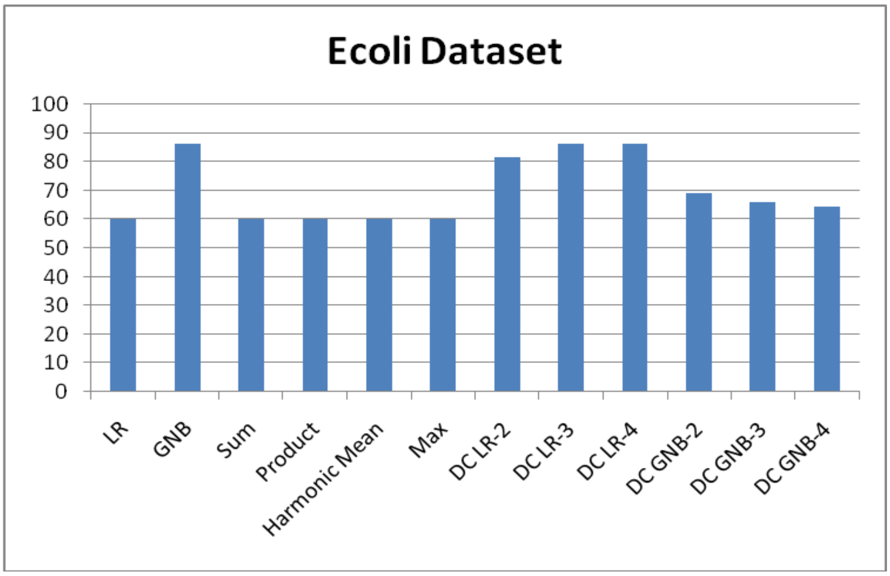

Fig 6 Comparison of accuracy of different classifiers over E-Coli Dataset

The behavior of mathematical combination classifiers in this set of experiments is same as in previous (ecoli). The accuracy of GNB is much lower than that of LR in individual case. The mathematical combinations predict with lower accuracy out of two. The proposed classifiers DC-LR with window_size varying from 2 till 4 show higher accuracy than all the rest. Proposed classifier DC-GNB at different window_size report accuracy higher than individual GNB.

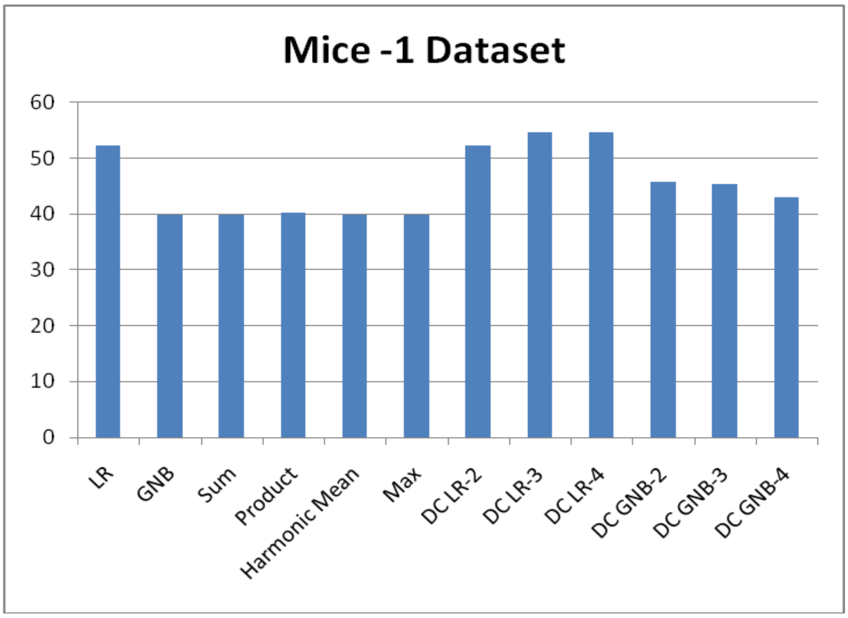

Fig 7 Comparison of accuracy of different classifiers over Mice-1 Dataset

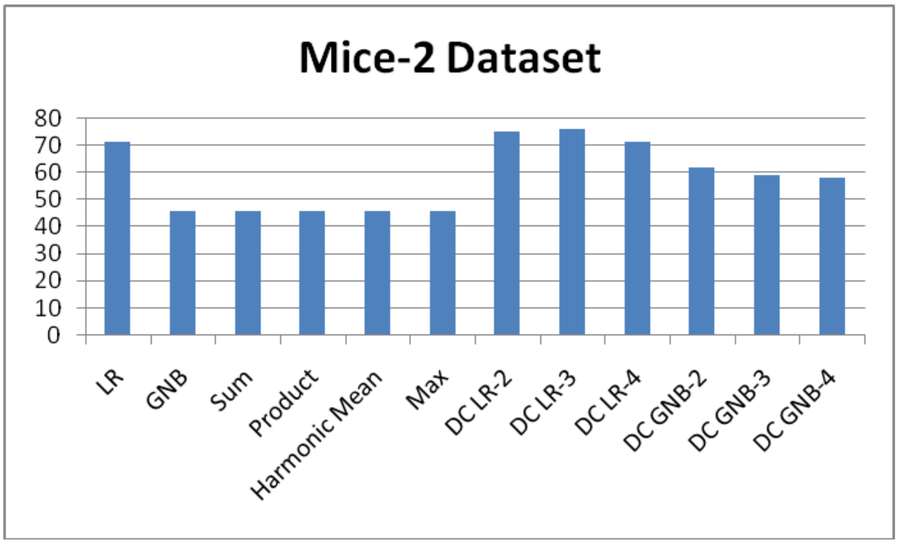

Fig 8 Comparison of accuracy of different classifiers over Mice-2 Dataset

The accuracies of individual classifiers are approximately equal here (see Fig 9). Mathematical combination classifiers show a slight improvement in the accuracy. The proposed classifier DC-LR gives better results than all at every value of window_size from 2 till 4. Proposed classifier DC-GNB is better than individual GNB in performance.

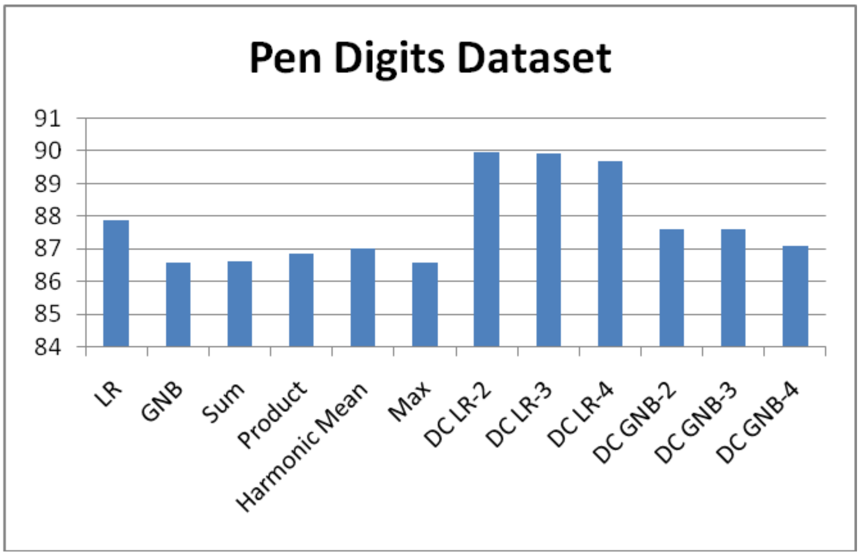

Fig 9 Comparison of accuracy of different classifiers over Pen Digits Dataset

V. Summary Of Results

The results of various experiments can be summarized into following observations and outcomes:

- Mathematical combination methods of Sum, Product, Harmonic Mean and Max are not sufficient to obtain an improvement in performance of single classifiers. Their accuracy depends on actual value of probabilities computed by single classifiers. In case a classifier produces high values of probabilities for some classes, and negligible for others, the output of any mathematical combination classifier will be inclined towards the prediction of that individual classifier. Thus, the purpose of combining results of more than one classifier gets nullified.

- Proposed classifiers perform better than mathematical combinations for all datasets.
- The effect of the proposed classifiers is more evident when number of instances is very large, as in Mice dataset, or number of categories is high as in Pendigits dataset. Hence, it can be claimed that the proposed classifiers are better combination methods when data is large and many classes are possible.

VIII. CONCLUSION

Classifying continuous data is continuing to be a leading area of research in pattern recognition. However, in spite of the classic and the new proposals, not one single algorithm is deduced to be the best for every application studied for classification. Each classifier has a set of advantages and drawbacks. Combination of classifiers therefore is expected to return the better results in a broader variety of applications too. The bases for any combined classification task are voting or any mathematical combination. The paper proposes a new scheme namely Dominated Consensus. The concept is similar to voting but acquires inspiration from the mathematical combinations. The Gaussian Naïve Bayes and Logistic Regression classifiers are combined in the proposal. The scope of the consensus is widened or reduced according to a user-defined parameter window_size. The classifiers participating in the combination achieve a desired consensus, failing which decision of the classification task depends on the decision of the dominating classifier.

Performance evaluation of the proposed algorithm is done by testing on a variety of real life datasets. An increased accuracy is achieved by the proposed classification task with increasing categories. Comparison of performance of the proposed combination is done with performances of the individual classifiers and the mathematical combination classifiers like Sum, Product, Harmonic Mean and Max. It can be claimed that the proposed combination method for classifying continuous data is suitable for a wide range of applications.

Extending the proposal to combine more than two classifiers is a future area to work upon. In addition, modifications in the proposed classification approach can be done to combine slower learning algorithms like Artificial Neural Networks, Support Vector Machines and more.

REFERENCES

[1] M. Bounhas and K. Mellouli, "A Possibilistic Classification Approach To Handle Continuous Data", *Proceedings of the 2010 IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)*, 2010.

[2] T. Le, F. Stahl, J. B. Gomes, M. M. Gaber and G. Di Fatta, "Computationally Efficient Rule-Based Classification for Continuous Streaming Data", M. Bramer and M. Petridis (eds.), *Research and Development in Intelligent Systems XXXI*, pp. 21-34, 2014.

[3] M. Volkovs, F. Chiang, J. Szlichta and R. J. Miller, "Continuous data cleaning", *Proceedings of the 2014 IEEE 30th International Conference on Data Engineering (ICDE)*, 2014.

[4] L. S. Larkey and W. B. Croft, "Combining Classifiers in Text Categorization", *Proceedings of SIGIR-96, 19th ACM International Conference on Research and Development in Information Retrieval (Zurich, CH, 1996)*, pp. 289–297 1996.

[5] Kostas Fragos, Petros Belsis and Christos Skourlas, "Combining Probabilistic Classifiers for Text Classification*", Procedia - Social and Behavioral Sciences, Volume 147, pp. 307–312, 3rd International Conference on Integrated Information (IC-ININFO),* 2014.

[6] Bruno Grilheres, Stephan Brunessaux and Philippe Leray, "Combining classifiers for harmful document filtering", RIAO '04, Coupling approaches, coupling media and coupling languages for information retrieval, *Proceedings of RIAO*, pp.173-185, 2004.

[7] P. N. Bennett, S. T. Dumais and E. Horvitz, "Probabilistic Combination of Text Classifiers Using Reliability Indicators: Models and Results", *Proceedings of SIGIR'02*, 2002.

[8] N. Xue and S. P. Converse, "Combining Classifiers for Chinese Word Segmentation", *Proceedings of the first SIGHAN workshop on Chinese language processing (SIGHAN '02),* Vol.18, pp. 1-7, 2002.

[9] H. Kashyap and B. Buksh, "Combining Naïve Bayes and Modified Maximum Entropy Classifiers for Text Classification", *International .Journal of Information Technology and Computer Science,* Vol. 9, No. 5, pp.32-38 , 2016.

[10] C. M. Sehgal, T. W. Cary, A. Cwanger, B. J Levenback, S. S. Venkatesh, "Combined Naïve Bayes and logistic regression for quantitative breast sonography", *Proceedings of the 2012 IEEE International Ultrasonics Symposium,* pp. 1686-1689, 2012.

[11] UCI Machine Learning Repository, http://ics.uci.edu/ mlearn/MLRepository.html