International Journal For Applied Research in Science and Engineering
Volume 5, Issue 3, October-2022 ISSN (Online): 2456-124X

HYBRID ENSEMBLE MODEL: A NOVEL PRECISION AWARE LEARNING APPROACH

[1]Pappu Kumar, [2]Dr. Mukesh Kumar
Research Schalor, Professor
Rabindranath Tagore University

Abstract: There are different methods and techniques for creating the ensembles. Bagging and boosting are widely used ensemble learning (EL) approaches used in literature to improve the performance of models. Bagging is a parallel approach whereas boosting is a sequential approach. Bagging trains the weak classifiers by obtaining the subsamples with replacement from the sample space and then combine the results using some approximation technique. Boosting on the other hand increases the weight of misclassified samples during each successive trials. Stacking is another known approach where multilevel stacking is used for classification of data. Each successive level of stack is trained on predictions of previous level of stack.
Keyword-MLP, KNN, SVM, PCA, LDA

INTRODUCTION

Ensembles could be created by combining methods of similar or dissimilar types which is widely known as homogeneous and heterogeneous ensembles respectively. There are ensemble approaches which have investigated its performance in data space e.g. adaboost. It is widely used in application such as human detection, traffic pre diction, image retrieval an so on. There are ensemble approaches that work on feature spaces. Some random independent features are selected for training the models. This is called random subspace technique. In both scenario, results could be combined by using majority voting or by applying any of the summation, maximum, minimum, product or median rule.

This part of work tries to propose an approach which combines the results ob-tained from bagging, boosting and stacking to make its final predictions. Strategy for combining the results for binary class data is different from multiclass data.

RELATED WORK

Due to its robustness and performance, in recent years researchers diverted their interest in EL approaches for learning. There are many new algorithms developed like bagging, classifier ensemble of neural network [1], boosting, heterogeneous ensemble of classifiers [2].

Many researchers have shown interest in developing ensemble of classifiers. [3] Designed graph based semi-supervised ensemble model by performing repeated iteration of feature selection. [3] developed an enhancement to random forest ensemble such that each subspace have sufficient good features. used the concept of rough set theory in the ensemble framework for dimensionality reduction. [4] used rotational space technique for feature selection to induce diversity in classifier in random forest algorithm. On the other side, some researchers have shown interest in the properties of ensem-ble. investigated kappa-error. focuses on generalizing ability and fuzziness of ensemble.

Researchers also focused on heterogeneous combination of classifiers to optimize the results in ensemble. [5] investigated and improved the efficacy of adaBoost by reducing dimensions using random subspace technique. They have analyzed the effect by reducing the correlation between features, reducing the impact of outliers in ad-aBoost training and proposed a novel idea for identifying weak learners. [5]combined rotational forest and AdaBoost for their ensemble.

Authors of [6] developed an algorithm which combines the properties of gradient boosting and random forest (Bagging) named as InfiniteBoost. implements another combination of boosting and random forest by muting trees and features called DART.applied boosting by sampling both rows and features without replacement from training data are proposed a model named as BagBoo.

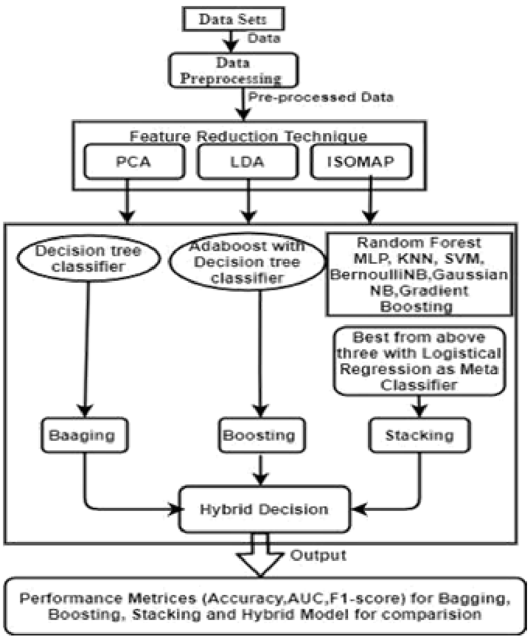
To the best of our knowledge none of the authors have tried to combine the all three EL technique, bagging, boosting and stacking for classification of data based on number of classes using feature reduction technique.

PROPOSED METHODOLOGY

Figure 1 shows methodology used for classification, by combining three EL techniques bagging, boosting and stacking. Basic steps involved are data pre-processing, applying feature reduction and using proposed “hybrid ensemble model” (HEM) for classification.

In the figure 1, individual datasets are pre-processed for missing values, NULL values and out of range values. On each dataset all three (PCA, LDA, ISOMAP) feature reduction techniques are applied separately. The reduced features obtained are used to perform further analytics using ML algorithms.

To evaluate the efficiency of the model and reduce the over fitting and under fitting problems, K-fold cross-validation technique is used. In K-fold cross-validation, there is a bias-variance trade-off correlated with the decision of K [10]. Generally, despite these criteria, one performs K-fold cross-validation with K=5 or K=10, since values have been experimentally shown to provide test error rate estimates that do not suffer from extreme bias or extremely high variance. This study has applied 5-fold cross-validation technique. All three ensemble approaches are applied namely bagging with decision tree as base classifier, adaboost with decision tree as base classifier and stacking to each data set



Representative eight algorithms out of which three best is selected for stacking are taken from several families of algorithm. Random forest model from tree family [12]
Multilayer layer perceptron (MLP) from neural network family , gradient boosting from ensemble family [13], bernoulli and gaussian from bayesion family, K-nearest neighbour classifier (KNN) and SVM from instance-based family and logistic re-gression from regression family .
Stacking optimize their predictions by using multiple levels. Subsequent levels use predictions of previous levels as a training data and apply meta model for predic-tions/classification. Three best performing algorithms are selected out of eight algorithms based on its accuracy for building the level-1 stack. Since complex transfor-mation is applied on datasets, it would be not recommended to use a complex algorithm as meta model. Logistic regression is a good choice. It is also used in literature as a meta model.

Datasets
This research work used 10 binary and 10 multiclass IoT datasets from UCI ML repos-itory [34] and kaggle. Dataset used are of both high and low dimensions, varies from small to large size to reduce any favorable impact on performance on proposed model.
Tables 3.1 contains details about datasets.

Sr No.	Datasets	Total features	Total classes	Instances
	Binary class			
1	Electric grid	14	2	10000
2	Extra sensory - B	277	2	2686
3	Football sensor	9	2	945
4	Pulsar star	9	2	9652
5	EEG signal	15	2	8123
6	Power system - A	129	2	5161
7	Hand gesture recognition	65	2	5811
8	Watch sensor	13	2	7386
9	Power system - B	99	2	4966
10	Machine sensor	75	2	10616
	Multiclass			
11	Cardiotocography sensor	41	10	2126
12	Extra sensory - A	277	6	2686
13	Mode detection	33	5	5894
14	Sky server	18	3	10000
15	Movement recognition	563	6	2948
16	Air quality sensor	16	5	9358
17	Energy prediction	29	8	19736
18	Big sensors	561	6	6373
19	Transport detection	38	5	5894
20	Direction sensor	25	4	5455

Table 3.1: Datasets

PROPOSED "HYBRID ENSEMBLE MODEL" (HEM)

The proposed hybrid model combines the predictions made by bagging [16], boosting, stacking to obtain its final results. Decision tree is used as base classifier for bagging and boosting. Adaboost is used as boosting variant. 30% of each dataset is used as test data for comparing the performance of HEM with its component models.
In the proposed "hybrid ensemble model" (HEM), for binary class data, class of an instance is predicted using majority votes. For example, if bagging predicted an instance as 0, boosting predicts it as 1 and Stacking predicts it as 1 then the final predicted class of an instance will be 1 as it got 2 votes (majority votes) out of 3. For multiclass data, HEM used predicted values as well as the probability values (proba-bility of predicted values made on the training dataset) of bagging, boosting, stacking for making final prediction of an instance. In case of tie between the classes during majority votes using predicted value, a class having highest predicted probability value amongst constituent model is a final class label for that instance. Finally, the study compares predicted value on test dataset (30% data) to measure the performance. The "HEM" aims to improve the stability of the EL model. Even if training data is slightly modified, the prediction will not change. Table 3.2 shows the pseudo code for HEM.

RESULTS AND COMPARISON

Experimentation is performed using google colab notebook which is an online cloud-based platform. Scikit learn library for ML [18] is used which offers a number of supervised and unsupervised learning algorithms via a simple python framework. Table 4.3 shows the number of features obtained after applying feature reduction techniques. The number of features is reduced such that 95% of the variance of data is covered.

Sr No.	Datasets	Total features	Reduced dimension PCA	LDA	IsoMap
1	Electric grid	14	10	1	11
2	Extra sensory - B	277	18	1	21
3	Football sensor	9	7	1	8
4	Pulsar star	9	4	1	3
5	EEG signal	15	5	1	9
6	Power system - A	129	22	1	27
7	Hand gesture recognition	65	42	1	23
8	Watch sensor	13	6	1	4
9	Power system - B	99	18	1	22
10	Machine sensor	75	13	1	16
11	Cardiotocography sensor	41	12	1	15
12	Extra sensory - A	277	11	4	23
13	Mode detection	33	5	3	11
14	Sky server	18	8	2	10
15	Movement recognition	563	110	3	195
16	Air quality sensor	16	4	4	5
17	Energy prediction	29	13	1	14
18	Big sensors	561	190	3	209
19	Transport detection	38	7	3	13
20	Direction sensor	25	18	3	18

Table 1: Dimensionality reduction using PCA, LDA and IsoMap.

Experimentation is conducted using new reduced features to obtain accuracy metric for eight algorithms on individual data sets separately. For selecting best three algorithms, accuracy values of ten binary class datasets and ten multiclass data are averaged across all reduction techniques and is shown is table. K-NN, SVM and GBM are the three best performers for both binary and multiclass datasets. Highest performer is written in bold.

Sr No.	Dataset type	RF	BNB	GNB	MLP	KNN	SVM	GBM	LR
1	Binary class	83.89	75.08	81.69	85.67	88.31	85.88	87.01	77.44
2	Multi class	79.98	72.07	77.91	81.99	86.31	85.66	86.61	77.80

Table 2: Average accuracy for eight algorithms.

After selecting three best models, level-1 train dataset is created using 5FCV, where model is fitted on k-1 folds and made predictions on a remaining fold (let us denote these predictions as A). Level-1 test dataset is created (let us denote is as Z) by using top 3 models on complete original train dataset (70% data of original data) and test dataset (30% of original data) . Finally, we train LR model as meta-classifier on level-1 train data (A) and predicted on the level-1 test dataset (Z). Results are obtained by averaging values on each dataset for binary and multiclass data set for accuracy, AUC and F1-score using PCA, LDA and IsoMap

Sr No.	Class types	Dataset	Feature reduction technique	Accuracy	AUC	F1- score
11	Multi class	Cardiotocography sensor	PCA	100	1	1
			LDA	100	1	1
			IsoMap	100	1	1
12		Extra sensory-A	PCA	75.232	0.44	0.752
			LDA	78.584	0.527	0.785
			IsoMap	71.322	0.522	0.713
13		Mode detection	PCA	89.228	0.959	0.892
			LDA	71.586	0.925	0.715
			IsoMap	85.581	0.948	0.855
14		Sky server	PCA	89.4	0.895	0.894
			LDA	94.35	0.948	0.943
			IsoMap	84.65	0.854	0.846
15		Movement recognition	PCA	93.22	0.989	0.932
			LDA	84.406	0.95	0.844
			IsoMap	92.542	0.985	0.925
16		Air quality sensor	PCA	90.17	1	0.901
			LDA	95.673	1	0.957
			IsoMap	90.384	1	0.903
17		Energy prediction	PCA	98.378	0.998	0.983
			LDA	100	1	1
			IsoMap	75.95	0.929	0.759
18		Big Sensors	PCA	89.725	0.996	0.897
			LDA	81.803	1	0.818
			IsoMap	83.764	0.993	0.837
19		Transport detection	PCA	91.687	0.962	0.916
			LDA	78.71	0.936	0.787
			IsoMap	89.737	0.959	0.897
20		Direction sensor	PCA	89.459	0.88	0.894

Table 3: Accuracy, AUC and F1-score with PCA, LDA and IsoMap using HEM.

To perform the comparative analysis between bagging, boosting, stacking and pro-posed HEM, output values are averaged across binary class and multiclass data set for accuracy, AUC and F1- score using PCA, LDA and IsoMap. These output values are obtained during the execution of HEM for its component models.

Table 3 shows the average accuracy for bagging, boosting, stacking and HEM for binary class and multiclass data using PCA, LDA and IsoMap. Highest score among the EL models using each reduction technique is marked in bold.

Sr. No	Dataset Type	Feature reduction technique	Bagging model	Boosting model	Stacking model	Hybrid model
1	Binary class	PCA	93.879	91.185	93.816	94.648
		LDA	79.117	77.029	81.685	78.536
		IsoMap	93.320	90.542	93.555	94.215
2	Multi class	PCA	88.963	84.523	91.712	90.649
		LDA	86.276	82.970	85.865	86.256
		IsoMap	85.586	80.269	86.090	86.219

Table 4: Average accuracy for ensemble models using PCA, LDA and IsoMap.

Table 4 shows the average AUC score for bagging, boosting, stacking and HEM for binary and multiclass data using PCA, LDA and IsoMap.

Sr.No	Dataset Type	Feature reduction technique	Bagging model	Boosting model	Stacking model	Hybrid model
1	Binary class	PCA	0.915	0.895	0.914	0.927
		LDA	0.757	0.741	0.768	0.751
		IsoMap	0.909	0.891	0.912	0.922
2	Multi class	PCA	0.903	0.897	0.924	0.911
		LDA	0.907	0.904	0.95	0.91
		IsoMap	0.871	0.855	0.916	0.906

Table 5: Average AUC for ensemble models using PCA, LDA and IsoMap.

Sr No.	Dataset Type	Feature reduction technique	Bagging model	Boosting model	Stacking model	Hybrid model
1	Binary class	PCA	0.902	0.871	0.896	0.917
		LDA	0.701	0.683	0.698	0.693
		IsoMap	0.896	0.867	0.891	0.911
2	Multi class	PCA	0.889	0.844	0.916	0.906
		LDA	0.861	0.829	0.858	0.862
		IsoMap	0.860	0.802	0.860	0.861

Table 6: Average F1-score for ensemble models using PCA, LDA and IsoMap.

To have better understanding values shown in table 1, table 2 and table 3 are visualized by plotting bar graph in figure 4 to figure 5 separately for binary class and multi class.

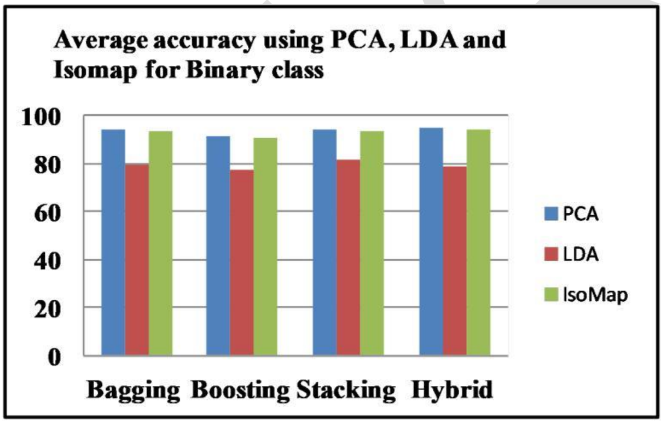


Figure 2: Average accuracy using PCA, LDA and IsoMap for binary class

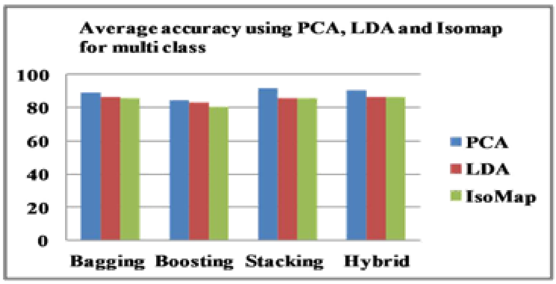


Figure 3: Average accuracy using PCA, LDA and IsoMap for multi class.

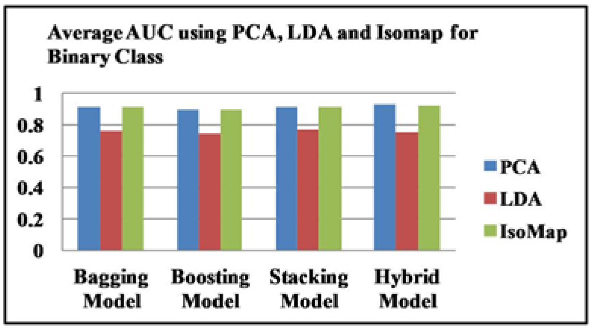


Figure 4: Average AUC using PCA, LDA and IsoMap for binary class.

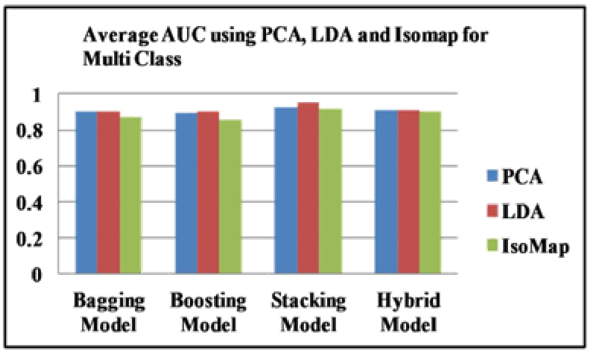


Figure 5: Average AUC using PCA, LDA and IsoMap for multi class.

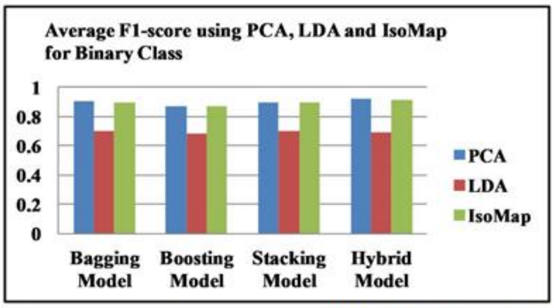


Figure 6: Average F1-score using PCA, LDA and IsoMap for binary class

CONCLUSIONS

In this research developed a novel approach called "hybrid ensemble model" (HEM) which does not require any tuning of parameters. On most of the performance parameters (accuracy, AUC, F1-score) it outperforms other state of the art EL techniques. This research recommends using the proposed HEM model as a generalized learner for binary classification with PCA and

IsoMap and for multi-class classification with IsoMap as a feature reduction technique in the absence of any previous knowledge on a problem from a IoT domain.

Limitations and Future work

Following are the limitations of the proposed hybrid model.

1. It can be further reduce the time complexity by using a big data framework for implementing the algorithm.
2. Impact on performance in specific classification tasks from particular applications.
3. Examine the robustness to hybrid approach by applying it to benchmark data sets representing different problems.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” Computer networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[2] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth, “Machine learning for internet of things data analysis: A survey,” Digital Communications and Networks, vol. 4, no. 3, pp. 161–175, 2018.

[3] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of things (iot): A vision, architectural elements, and future directions,” Future generation computer systems, vol. 29, no. 7, pp. 1645–1660, 2013.

[4] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiqa, and I. Yaqoob, “Big iot data analytics: architecture, opportunities, and open research challenges,” ieee access, vol. 5, pp. 5247–5261, 2017.

[5] A. Janusz, M. Grzegorowski, M. Michalak, Ł. Wróbel, M. Sikora, and D. Ślęzak, “Predicting seismic events in coal mines based on underground sensor measure-ments,” Engineering Applications of Artificial Intelligence, vol. 64, pp. 83–94, 2017.

[6] Z. Geng, H. Wang, M. Fan, Y. Lu, Z. Nie, Y. Ding, and M. Chen, “Predicting seismic-based risk of lost circulation using machine learning,” Journal of Petroleum Science and Engineering, vol. 176, pp. 679–688, 2019.

[7] M. J. Ismail, R. Ibrahim, and I. Ismail, “Adaptive neural network prediction model for energy consumption,” in 2011 3rd International Conference on Computer Re-search and Development, vol. 4. IEEE, 2011, pp. 109–113.

[8] Y. K. Penya, C. E. Borges, D. Agote, and I. Fernández, “Short-term load fore-casting in air-conditioned non-residential buildings,” in 2011 IEEE International Symposium on Industrial Electronics. IEEE, 2011, pp. 1359–1364.

[9] T. G. Dietterich, “Approximate statistical tests for comparing supervised classi-fication learning algorithms,” Neural computation, vol. 10, no. 7, pp. 1895–1923, 1998.

