Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

PM2.5 Prediction using Heterogeneous Ensemble Learning

Shrabani Medhi[1], Pallav Kashyap[2], Akansha Das[3], Jitjyoti Sarma[4]

[1]Computer Science and Engineering, Girijananda Chowdhury University, Guwahati, India

[2,3,4]Computer Science and Engineering, Girijananda Chowdhury Institute of Management and Technology, Assam Science and Technology University, Guwahati, India

Email: [1]shrabani_cse@gcuniversity.ac.in, [2]pallavkashyap2@gmail.com, [3]akanshadastez19@gmail.com, [4]jitjyotisarma2000@gmail.com

Abstract

Air pollution is a great concern to mankind and is causing too many adverse effects on every living organism on earth by increasing lung diseases, skin diseases, and many other problems caused by it. This research presents a comprehensive study on the application of heterogenous ensemble learning techniques for PM2.5 concentration prediction, aiming to enhance prediction accuracy and provide insights into the driving factors behind pollution levels. The primary objective is to conduct a comparative analysis of heterogenous ensemble method, namely, blending and stacking in conjunction with individual base models, such as multiple linear regression (LR), decision trees (DT), support vector regression (SVR) and artificial neural networks (ANN). In total 28 models were created using blending and 28 models were created using stacking. Hyperparameter tuning is done to optimize the models.

Keywords: PM2.5 Prediction, Support Vector Machine, Decision Tree, Multiple Linear Regression, Artificial Neural Network, Ensemble Learning, Stacking, Blending.

1. Introduction

The growing world of today faces a serious threat from air pollution. The concentration of harmful chemicals in the atmosphere and growing industrialization are to blame for the increasing toxicity of the air. An estimated 1.6 million deaths in India were attributed to air

DOI: https://doi.org/10.36548/jaicn.2023.4.003
Received: 17.09.2023, received in revised form: 18.10.2023, accepted: 02.11.2023, published: 13.11.2023

pollution in 2019[1]. In recent years due to increased growth, urbanization and improved lifestyle in Guwahati city air pollution have increased tremendously. Guwahati has one of the highest black carbon pollution levels in the world [2].

The concentration of PM2.5 in Guwahati is much higher than the permissible limit. PM2.5 are particulate matter with width having diameter of less than 2.5 microns, or even lesser than that. The particles are suspended in the air in solid and liquid form. Example, ash, soot, dust, etc. The incredibly small size of the particulate matter allows it to easily enter the respiratory system and travel all the way to the lungs. There are many short-term and long-term health effects associated with PM2.5 exposure. Short-term health issues including irritation in throat, lung, and nose, as well as cough, sneezing, etc. Long-term effects may be serious health issues related to lung function, asthma and heart diseases[3]–[5]. The monitoring stations monitor PM2.5 and determine the AQI (air quality index) according to it. Government uses AQI numbers to show the quality of air to the public. It is shown in Table 1[6]. Increase in AQI means increase in air pollution and vice versa. According to Central Pollution Control Board (CPCB), air quality is categorised as six groups, namely, good, satisfactory, moderate, poor, very poor and severe.

Majority of these deaths were caused by particulate matter 2.5 (PM2.5) pollution. Monitoring air pollution is an important task and to solve the task machine learning (ML) models can be used. By applying different types of machine learning models air pollution analysis and prediction and forecasting of pollutants can be performed. The techniques used in machine learning can been effective in developing prediction models for forecasting air pollution. Several computational models that are based on machine learning paradigm and soft computing have been used to perform PM2.5 prediction and analysis. Support Vector Machine (SVM) [7], Neural Network, and other supervised machine learning approaches [8], [9] have found to be perform better for air pollution prediction than traditional arithmetic methods like Ridge Regression, Logistic Regression with respect to accuracy and error metrics. However, ensembled learning methods are based on a learning paradigm which combines various machine learning techniques. Ensembled learning can be homogenous or heterogenous [10]. Stacking is a well-known ensemble approach that is used to predict multiple models or learning algorithms via a meta model to produce an optimal predictive model. Stacking is primarily used to improve the model performance. The main rule of stacking is that it takes the output of

sub-models as inputs and learned about how to best combine the inputs predictions to make a better output prediction.

Table 1. AQI Values of PM2.5

Category of AQI	**AQI**	
	Index Value	Breakpoints for PM2.5 (μ/m3, average observed in 24 hours)
Good	0 to 50	0.0 to 30.0
Satisfactory	51 to100	31 to 60
Moderate	101 to 200	61 to 90
Poor	201 to 300	91 to 120
Very poor	301 to 400	121 to 250
Severe	401 to 500	250+

By stacking, the current accuracy can be increased and a model that is superior to all individual intermediate models can be produced. Blending is an ensemble machine learning algorithm. It follows the same approach as stacking, but the difference between stacking and blending is that stacking uses out-of-fold predictions for the train set of the next layer (known as the meta-model), and blending uses a validation set to train the next layer. Blending helps to improve performance and increase accuracy. Blending determines how to optimally integrate the predictions from several contributing ensemble member models through the use of a machine-learning model. Use of ensembled learning outperforms single classifiers and regressors[11], [12].

Using heterogenous ensemble, the model is designed to predict the PM2.5 concentration. [13]Ensemble learning techniques has shown lots of potential in the field of air quality prediction due to their ability to combine multiple models for improved accuracy and prediction [14]. Inclusion of several base models provide a basis for comparison. These simple

models include various linear models, SVM, DT, ANN and LR. Each base model is trained using the training data and tuning is done using appropriate hyperparameters.

The models is evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Squared Error (MSE) and Root Mean Squared Logarithmic Error (RMSLE) [15]. These measurements help to gather information about the model's accuracy, precision, and fitness quality. The literature review, methodology and results and discussion are explained in the sections below.

2. Related Work

The best method for AQI prediction to support climate control is found in this research[16]. Support vector regression (SVR), random forest regression (RFR), and multivariate analysis (MAA) are the three different approaches that are suggested. It was discovered that Cat Boost regression produced the best accuracies for New Delhi and Bangalore, whereas SMOTE produced the lowest RMSE values for Kolkata and Hyderabad. The research's innovation is that SMOTE is used to balance the dataset and the best regression models were chosen after careful investigation. The study in [17] reviews several modelling strategies and data processing techniques to increase the efficiency of any model. Deterministic models are used to forecast air quality; WRF models are used to make predictions about the atmosphere; statistical approaches are used to evaluate relationships between air quality and air pollution elements; and regression models are used to forecast concentration levels of pollutants. Recent advances in machine learning and artificial intelligence have enabled more intelligent predictors.

The study [18]reviews various modelling strategies and data processing techniques to increase the efficiency of any model. Deterministic models are used to forecast air quality; WRF models are used to make predictions about the atmosphere; statistical approaches are used to evaluate relationships between air quality and air pollution elements; and regression models are used to forecast concentration levels of pollutants. The [6] thesis forecasts the Delhi Air Quality Index for a few future time periods using a variety of time series forecasting techniques. The levels of particulate matter (PM2.5 and PM10), sulphur dioxide (SO2), carbon monoxide (CO), and nitrogen dioxide (NO2) have been forecasted for a particular selected

region in Delhi. The study's conclusions also cite a number of secondary sources that shed light on the basic issues surrounding air pollution. One model employed a gated recurrent unit, while the other used a combination of decision trees, linear regression, long short-term memory, and gated recurrent units. The mean square error, root mean square error, and mean absolute error are performance indicators that are used to determine error rates. Certain variables exhibit improved overall performance when two models are integrated. A method for predicting air pollution using a long-short-term memory (LSTM) recurrent neural network is presented in the research carried out [7]. Based on its concentration over the previous hours and average traffic statistics, this approach is used to forecast the concentration of a certain air pollutant for the following day. The proposed method was experimentally evaluated by contrasting the suggested approach with the ARIMA model, multilayer perceptron, standard recurrent neural networks, and LSTM. To predict the level of air pollution the following day, the recommended strategy should be put into reality using medical diagnostic instruments and air pollution monitoring equipment.

3. Proposed Work

The workflow diagram of the research is given in Figure.1.

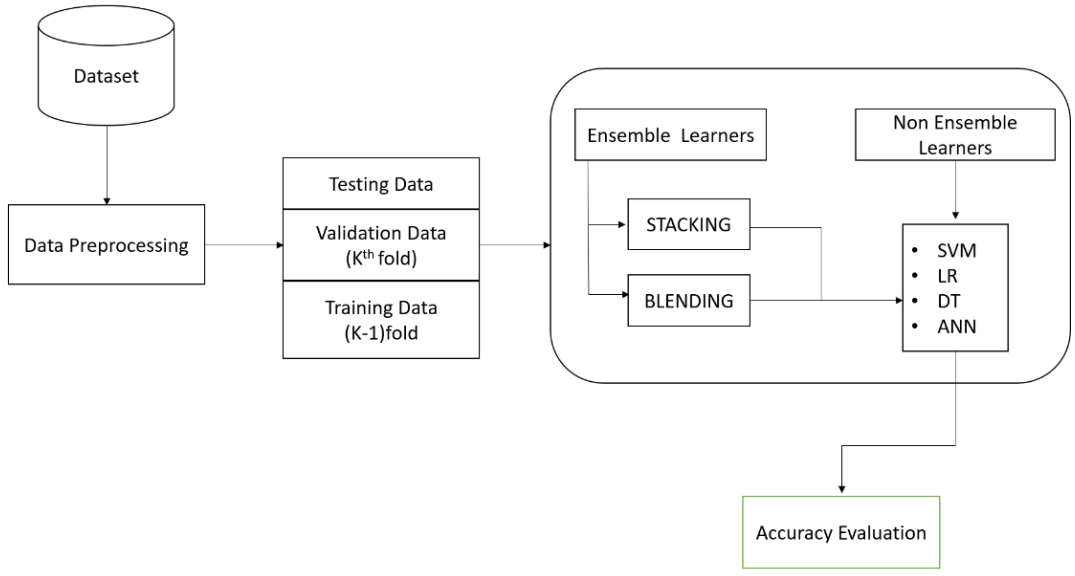

Figure 1. Workflow Diagram of the System

3.1 Data Source

Data from Guwahati's Continuous Ambient Air Quality Monitoring Station (CAAQMS) was gathered for this study from the Pollution Control Board, Assam, which is situated at Bamunimaidam. Data from time series are used. The data set includes Guwahati city's CAAQMS data for the three-year period ending in December 2022. There are 33067 data used in total. The parameters used in the study are given in Table. 2. The descriptive statistics indicates that there is no high value of skewness in data. It indicates that there is no sharp increase in the data. The high value of kurtosis in PM2.5 indicates the presence of data discontinuities. The proposed ensemble model aims at predicting the 1 hour ahead PM2.5 concentration for classification and regression.

3.2 Data Preprocessing

In preprocessing the data is cleansed through processes such as filling in missing values or resolving the inconsistencies present in the data. Outlier classification is applied in the preprocessing to detect the maximum and minimum outliers. It is found that RF has highest number of missing values and RH, SR and BP have least missing values. Interpolation is performed with the help of an imputer function. The strategy that is used here is mean value. Outliers are detected using Inter Quantile Range (IQR). Quantile based flooring and capping is used to deal with the outliers. The boxplot for outliers is shown in Figure 2. The data contains multiple inputs having different units. It is important that all the data are scaled into a particular range so that all attributes get equal weightage. Normalization is done so that an attribute having lesser significance with a large scale doesn't suppress another attribute of greater significance. After outlier classification, the last stage in data preprocessing is normalization[19]. In the normalization process, minmax normalization is done.

$$\text{Xscaled} = \frac{\text{X}-\text{Xmin}}{Xmax-Xmin} * (D - C) + C \qquad (1)$$

Table 2. Summary of Measurement Site and Observed Attributes

Measurement Site	Type	Attributes
Guwahati City	Meteorological conditions	Rainfall, temperature, pressure, wind direction, wind speed,

		and relative humidity
	Criteria gases	NO_2, SO_2
	Particulates	PM10, PM2.5

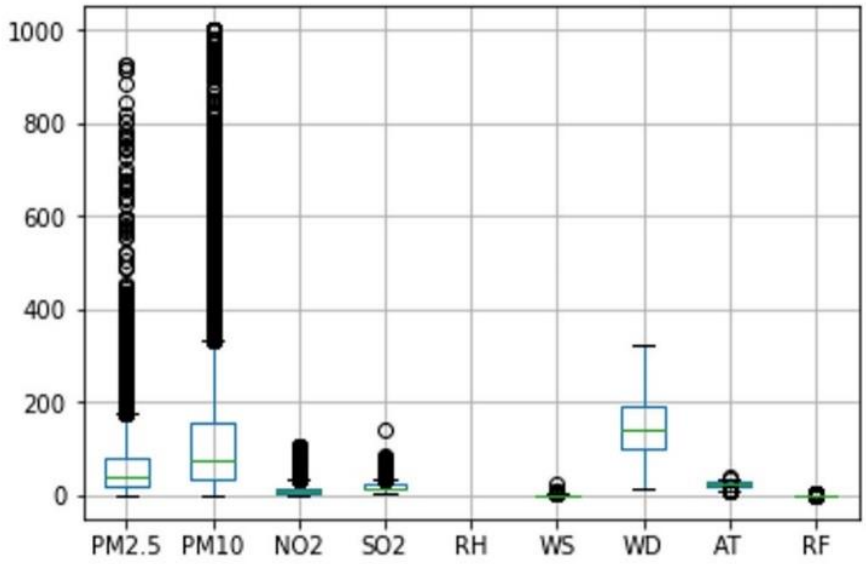

Figure 2. Detection of Outliers

3.3 Feature Selection

Feature selection is the process of selecting a subset of attributes from the dataset that contains the most relevant information for performing the prediction. Researchers suggest that reducing the number of input variables helps to lower the computational cost of modelling and hence improves the prediction capacity. Correlation matrix is used to check for correlation between features and hence determine the optimal number of input variables. It can be observed from Figure 3 that the attributes are not highly correlated. Correlation is very less. So, all the attributes present in the dataset is considered. Feature extraction is performed if there is redundant data. It involves selecting the optimum attributes. Many machine learning models have been found to perform better when their distribution is normal. When skewness is detected, they perform worse. Therefore, it's critical to determine whether there is skewness in the data and to apply mappings and transformations to change the skewed distribution into a normal distribution. Figure 3 shows the skewness values of different features. It is observed

that RF has the highest skewness and AT has the lowest skewness. Logarithmic transformation is used to reduce the skewness.

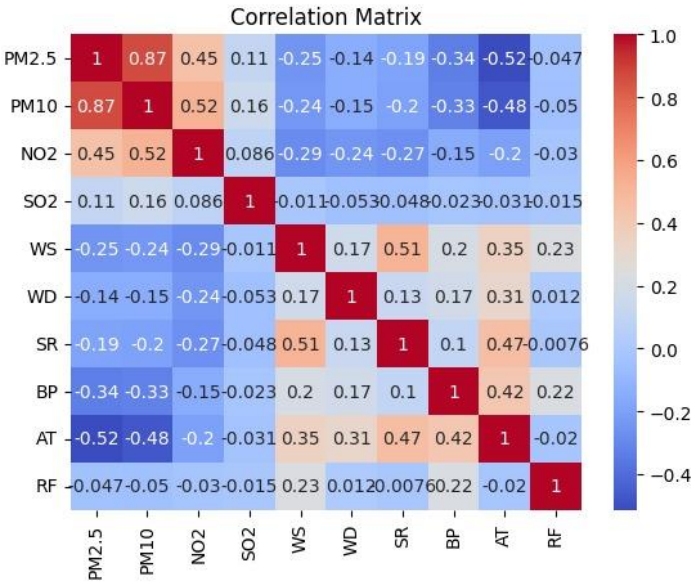

Figure 3. Correlation Matrix

3.4 Model Training

To assess the model's performance, the dataset is split into training and test sets. Thirty percent of the data is used for testing, while seventy percent is used for training. The model is tested and its generalizability assessed using the cross-validation approach, specifically the k-fold cross validation. Scikit-Learn library and Python library is used to build 56 models:(28) using stacking and (28) using blending to predict PM2.5 concentration. In order to reduce the error rate hyperparameter tuning is fitted into the models. RandomSearchCV and GridSearchCV is used to perform hyperparameter tuning. 10-fold cross validation (10-CV) was done to obtain an enhanced evaluation of training accuracy. Using this method, the training dataset is divided into 10 subsets. Out of the 10 subsets, 9 are used for training each model and 1 subset is used as testing dataset. This process is repeated 10 times representing ten folds in 10-CV. Stacking and blending as the ensembled techniques is used. Here future predictions are enhanced and improved by learning the mistakes of the past predictions. The base learner parameters were as follows: ANN [20] uses Multi -Layer Perceptron with three hidden layers: 5, 5 and 10 nodes respectively. Rectified Linear Unit (ReLU) is used as the activation function

and Adam optimizeris used in adjusting the learning rate . Maximum iteration was set to 2000. For SVM, RBF kernel is used and regularization of 100 is used. For DT, criterion=MSE is used. 1-100 estimators is used for homogenous ensembling. To evaluate the models MAE, MSE, R2 score and RMSE are used[21], [22].

Table 3. Error Metrics of Non-Ensembled Models

Non-ensembled Model	MAE	MSE	RMSE	RMSLE	Accuracy Score
SVM	16.391	1046.676	32.352	0.44	0.752
LR	15.334	990.775	31.476	0.465	0.753
DT	15.457	1040.084	32.25	0.462	0.743
ANN	13.151	809.677	28.454	0.38	0.808

4. Results and Discussion

In Table 3 the performance of non-ensembled models, namely MLR, DT, SVM and ANN is shown based on error metrics. SVM is a model commonly used for both regression and classification tasks. In this case, it performs reasonably well, with an accuracy score of 0.752. However, its MAE, MSE, and RMSE values are relatively high, suggesting that there might be room for improvement in terms of prediction accuracy and precision. LR is similar to SVM in terms of performance, with an accuracy score of 0.753. The MAE, MSE, and RMSE values are slightly better than SVM, indicating slightly better prediction accuracy. The accuracy score of DT is slightly lower than SVM and LR. The MAE, MSE, and RMSE values are also similar to those of SVM. The Artificial Neural Network outperforms the other models in all aspects. It has the lowest MAE, MSE, RMSE, and RMSLE values, indicating the highest prediction accuracy and precision. Additionally, it has the highest accuracy score among the models, making it the best performer in this analysis. Out of all the models ANN performs best with accuracy of 80.8%. In Table 4 the performance of 28 ensembled stacking models is given. When two base learner models used with one estimator: [LR+ANN(ANN), LR+ANN(LR), LR+DT(DT), DT+ANN(ANN), DT+ANN(DT)], all the stacking models performed better

compared to the individual base models. Only LR+DT(LR) did not perform significantly better than the individual base models. When three models are used as base models and one model as meta estimator, high accuracy scores were observed in many cases, but the regression metrics vary. Only for SVM+LR+DT(LR) a slightly lower accuracy score and slightly higher regression metrics is observed. Different combinations of base models, and the choice of the base model significantly influences the results. Careful consideration of the trade-off between accuracy and regression metrics is essential. High accuracy score and reasonable regression metrics is obtained for the models: SVM+LR+DT(SVM), SVM+LR+ANN(SVM), SVM+LR+ANN(ANN), SVM+DT+ANN(SVM). Very high accuracy is achieved when ANN is used as the meta estimator. When four base models and one meta estimator used, SVM+LR+DT+ANN(DT), the model with SVM, LR, DT and ANN as base learners and DT as the meta estimator outperforms all the other models with an accuracy of 100%. However, in some cases, it may lead to overfitting, as suggested by the high MSE, RMSE, and RMSLE values.

In Table 5 the performance of 28 ensembled blending models is given. It is observed that blending models performs worst if compared with stacking and non-ensembled models. When two models are used as base models and one as meta estimator, the blended models show mixed results, with many of them exhibiting negative accuracy scores and high RMSLE values. This suggests that these blended models may not be effective at improving prediction accuracy. The blended models with three base models show mixed results, with some exhibiting negative accuracy scores and mixed regression metrics. It appears that finding an effective combination of three models for blending is challenging in this scenario. Further exploration and fine-tuning of the blending approach, including the choice of base models and their weighting, may be necessary to achieve better results. When four models are used as base models, these blended models appear to have improved results compared to the previous combinations with three base models. In particular, the SVM+LR+DT+ANN model stands out with positive accuracy and low error metrics, suggesting that it might be a suitable choice for making predictions.

The accuracy of non-ensembled models, ensembled stacking and ensembled blending models are given in Figure 4, 5, 6.

Table 4. Error Metrics of Ensembled Stacking Models

Ensembled Stacking Model	**MAE**	**MSE**	**RMSE**	**RMSLE**	**Accuracy Score**
SVM+ANN(ANN)	0.058	0.023	0.154	0.152	0.553
SVM+ANN(SVM)	0.039	0.01	0.103	0.095	0.787
SVM+DT(DT)	0.053	0.025	0.158	0.142	0.454
SVM+DT(SVM)	0.066	0.026	0.162	0.143	0.42
SVM+LR(LR)	0.066	0.028	0.169	0.175	0.561
SVM+LR(SVM)	0.063	0.027	0.166	0.156	0.415
LR+ANN(ANN)	0.047	0.016	0.129	0.127	0.75
LR+ANN(LR)	0.058	0.022	0.149	0.149	0.674
LR+DT(DT)	0.049	0.017	0.132	0.137	0.679
LR+DT(LR)	0.066	0.032	0.18	0.163	0.375
DT+ANN(ANN)	0.052	0.018	0.137	0.142	0.676
DT+ANN(DT)	0.057	0.023	0.153	0.162	0.721
SVM+LR+DT(SVM)	0.148	0.0492	0.221	0.265	0.922
SVM+LR+DT(LR)	0.148	0.049	0.221	0.271	0.893
SVM+LR+DT(DT)	0.148	0.049	0.221	0.277	0.831
SVM+LR+ANN(SVM)	0.148	0.049	0.221	0.239	0.95
SVM+LR+ANN(LR)	0.148	0.049	0.221	0.27	0.901
SVM+LR+ANN(ANN)	0.148	0.049	0.221	0.249	0.95
SVM+DT+ANN(SVM)	0.148	0.049	0.221	0.237	0.971
SVM+DT+ANN(DT)	0.148	0.049	0.221	0.262	0.867

SVM+DT+ANN(ANN)	0.148	0.049	0.221	0.236	0.976
LR+DT+ANN(LR)	0.32	0.145	0.381	0.248	0.966
LR+DT+ANN(DT)	0.32	0.145	0.381	0.232	0.955
LR+DT+ANN(ANN)	0.32	0.145	0.381	0.218	0.988
SVM+LR+DT+ANN(SVM)	3.167	253.976	15.936	0.197	0.828
SVM+LR+DT+ANN(LR)	2.867	139.535	11.812	0.397	0.905
SVM+LR+DT+ANN(DT)	2.819	2.577	5.076	4.879	1
SVM+LR+DT+ANN(ANN)	2.856	140.441	11.85	0.378	0.905

Table 5. Error Metrics of Ensembled Blending Models

Ensembled Blending Model	**MAE**	**MSE**	**RMSE**	**RMSLE**	**Accuracy Score**
SVM+LR(SVM)	0.229	0.07	0.266	0.186	0.059
SVM+LR(LR)	0.257	0.087	0.296	0.208	0.074
SVM+DT(SVM)	0.233	0.102	0.32	0.225	-0.402
SVM+DT(DT)	0.257	0.098	0.314	0.22	-0.169
SVM+ANN(SVM)	0.348	0.167	0.409	0.312	-1.177
SVM+ANN(ANN)	0.257	0.086	0.293	0.195	-0.196
LR+DT(LR)	0.3	0.13	0.361	0.244	-0.432
LR+DT(DT)	0.31	0.153	0.391	0.273	-1.325
LR+ANN(LR)	0.188	0.056	0.238	0.171	-0.074

LR+ANN(ANN)	0.22	0.071	0.267	0.181	-0.01
DT+ANN(DT)	0.287	0.125	0.353	0.241	-0.684
DT+ANN(ANN)	0.358	0.148	0.384	0.262	-0.45
SVM+LR+DT(SVM)	0.208	0.067	0.258	0.172	0.269
SVM+LR+DT(LR)	0.31	0.124	0.353	0.245	-0.538
SVM+LR+DT(DT)	0.328	0.137	0.371	0.262	-0.348
SVM+LR+ANN(SVM)	0.219	0.067	0.26	0.179	-0.157
SVM+LR+ANN(LR)	0.287	0.113	0.336	0.229	-0.104
SVM+LR+ANN(ANN)	0.278	0.127	0.356	0.356	-0.965
SVM+DT+ANN(SVM)	0.237	0.082	0.287	0.183	-0.174
SVM+DT+ANN(DT)	0.273	0.103	0.322	0.214	-0.466
SVM+DT+ANN(ANN)	0.26	0.085	0.291	0.194	0.051
LR+DT+ANN(LR)	0.329	0.133	0.365	0.258	-0.486
LR+DT+ANN(DT)	0.288	0.104	0.322	0.223	-0.678
LR+DT+ANN(ANN)	0.308	0.136	0.369	0.257	-0.108
SVM+LR+DT+ANN(SVM)	0.223	0.072	0.268	0.182	0.132
SVM+LR+DT+ANN(DT)	0.18	0.057	0.239	0.171	0.204
SVM+LR+DT+ANN(LR)	0.229	0.071	0.267	0.186	0.026

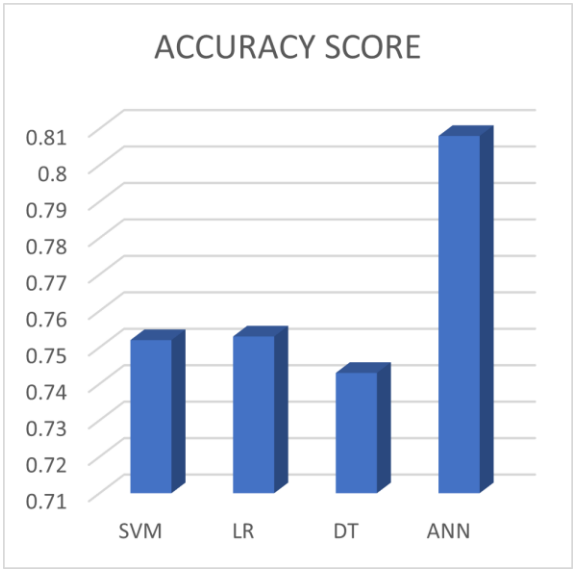

Figure 4. Accuracy of Non-Ensembled Models

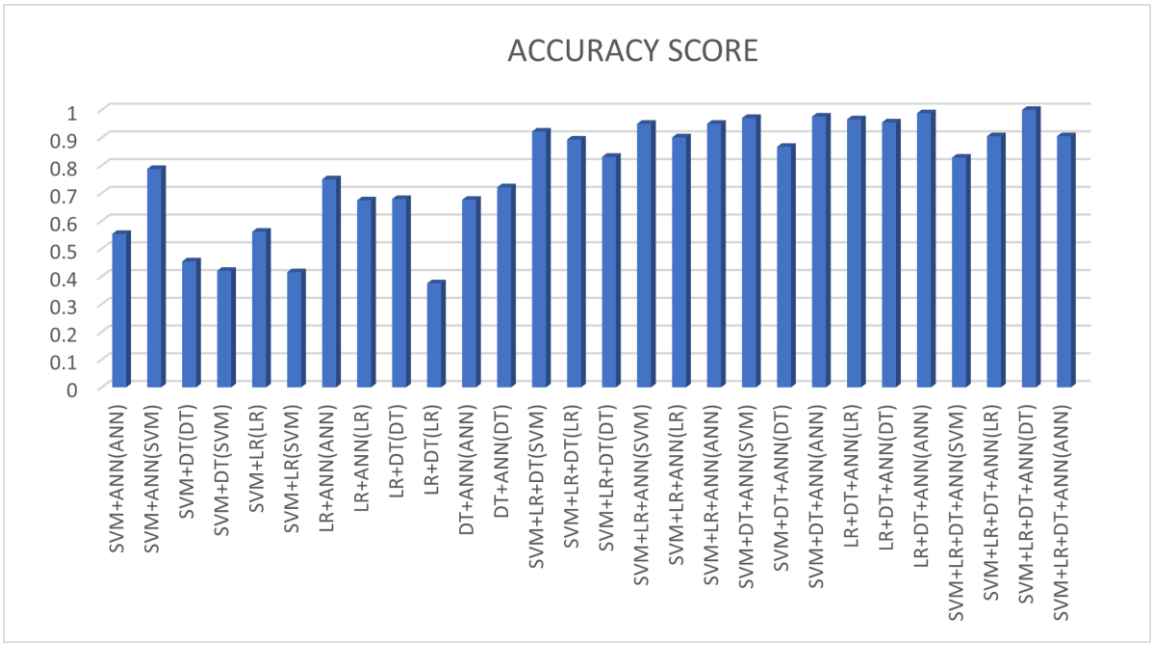

Figure 5. Accuracy of Ensembled Stacking Models

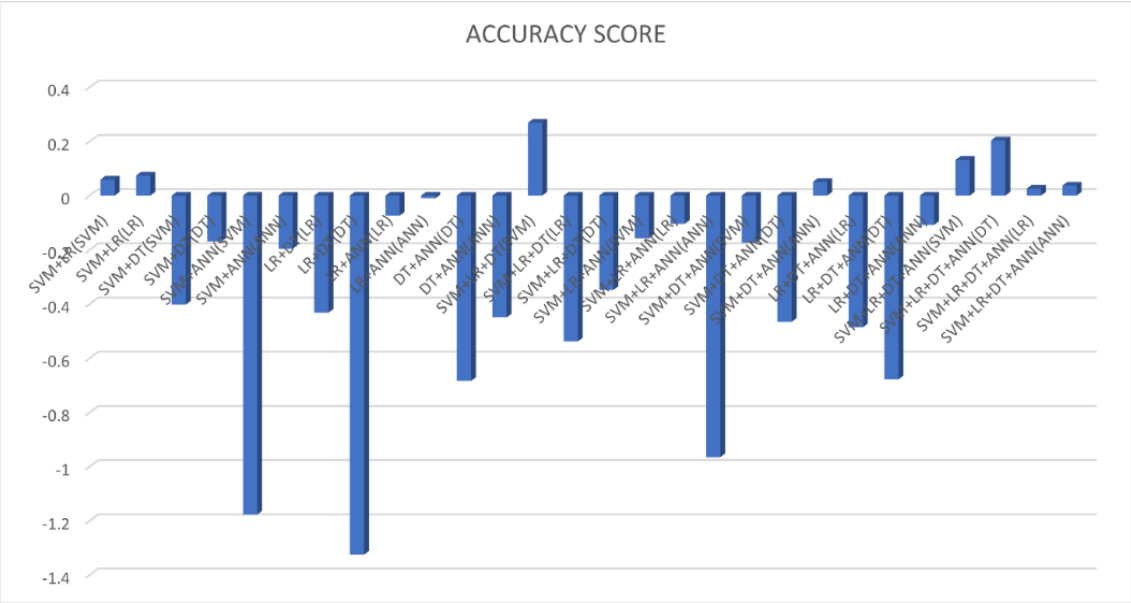

Figure 6. Accuracy of Ensembled Blending Models

5. Conclusion

A rigorous comparison is performed between heterogenous ensemble learning models using stacking and blending with decision tree, artificial neural network, multiple linear regression and support vector machine as the base learners and non-ensemble models to predict the PM2.5 concentration in Guwahati city. Total of 4 non-ensembled models, 28 ensembled stacking models and 28 ensembled blending models were tested. From the results it is observed that ensemble stacking models outperformed ensemble-blending and non-ensemble models. Among all the non-ensembled models, the Artificial Neural Network outperforms the other non-ensemble models: DT, SVM and LR. It has the lowest MAE, MSE, RMSE, and RMSLE values, indicating the highest prediction accuracy and precision. Among ensemble stacking models, when two base models are used all the stacked models performed better than the individual models except DT+LR (LR). When three models are used as base models and one model as meta estimator, high accuracy scores are observed in many cases, but the regression metrics vary. Very high accuracy is achieved when ANN is used as the meta estimator. When four base models and one meta estimator used, SVM+LR+DT+ANN(DT), outperforms all the other models with an accuracy of 100%. However, in some cases, it may lead to overfitting, as suggested by the high MSE, RMSE, and RMSLE values. Blending models performs worst if compared with stacking and non-ensembled models. When two models are used as base models

and one as meta estimator, the blended models show mixed results, with many of them exhibiting negative accuracy scores and high RMSLE values When four models are used as base models, these blended models appear to have improved results compared to the previous combinations with three base models. Overall ensemble stacking models performed best and blending models performed worst. The performance of non-ensembled models is average. In future prediction using homogenous ensemble models can be performed. For hourly PM2.5 concentration ensemble stacking models performed the best among blending and non-ensemble models. However more research can be done by including more features for PM2.5 concentration prediction.

ACKNOWLEDGEMENT

"First and foremost, we would like thank almighty GOD for everything that happens to us and makes us patient, dedicated and courageous to our work. We are very grateful to Pollution Control Board, Assam located in Bamunimaidan for providing us the necessary dataset without which it was not possible for us to carry out our research. Finally, we wish to warmly thank our university "Girijananda Chowdhury University-Guwahati" for encouraging us to do research work in our field".

References

[1] S. Medhi and M. Gogoi, "Visualization and Analysis of COVID-19 Impact on PM2. 5 Concentration in Guwahati city," in 2021 International Conference on Computational Performance Evaluation (ComPE), IEEE, 2021, pp. 012–016.

[2] N. Barman and S. Gokhale, "Urban black carbon-source apportionment, emissions and long-range transport over the Brahmaputra River Valley," Science of the Total Environment, vol. 693, p. 133577, 2019.

[3] S. Gonzalez-Gorman, S.-W. Kwon, and D. Patterson, "Municipal efforts to reduce greenhouse gas emissions: Evidence from US cities on the US-Mexico border," Sustainability, vol. 11, no. 17, p. 4763, 2019.

[4] N. A. H. Janssen, P. Fischer, M. Marra, C. Ameling, and F. R. Cassee, "Short-term effects of PM2. 5, PM10 and PM2. 5–10 on daily mortality in the Netherlands," Science of the total environment, vol. 463, pp. 20–26, 2013.

[5] K.-H. Kim, E. Kabir, and S. Kabir, “A review on the human health impact of airborne particulate matter,” Environment international, vol. 74, pp. 136–143, 2015.

[6] C. A. Pope III and D. W. Dockery, “Health effects of fine particulate air pollution: lines that connect,” Journal of the air & waste management association, vol. 56, no. 6, pp. 709–742, 2006.

[7] W. C. Leong, R. O. Kelani, and Z. Ahmad, “Prediction of air pollution index (API) using support vector machine (SVM),” Journal of Environmental Chemical Engineering, vol. 8, no. 3, p. 103208, Jun. 2020, doi: 10.1016/j.jece.2019.103208.

[8] S. M. Cabaneros, J. K. Calautit, and B. R. Hughes, “A review of artificial neural network models for ambient air pollution prediction,” Environmental Modelling & Software, vol. 119, pp. 285–304, Sep. 2019, doi: 10.1016/j.envsoft.2019.06.014.

[9] H. Maleki, A. Sorooshian, G. Goudarzi, Z. Baboli, Y. Tahmasebi Birgani, and M. Rahmati, “Air pollution prediction by using an artificial neural network model,” Clean Techn Environ Policy, vol. 21, no. 6, pp. 1341–1352, Aug. 2019, doi: 10.1007/s10098-019-01709-w.

[10] B. Weng, “Application of machine learning techniques for stock market prediction,” Apr. 2017, Accessed: Feb. 04, 2023. [Online]. Available: https://etd.auburn.edu//handle/10415/5652

[11] W. Sun and Z. Li, “Hourly PM2.5 concentration forecasting based on mode decomposition-recombination technique and ensemble learning approach in severe haze episodes of China,” Journal of Cleaner Production, vol. 263, p. 121442, Aug. 2020, doi: 10.1016/j.jclepro.2020.121442.

[12] L. Xu and Y. Zhang, “Quality Prediction Model Based on Novel Elman Neural Network Ensemble,” Complexity, vol. 2019, p. e9852134, May 2019, doi: 10.1155/2019/9852134.

[13] A. Jain, F. Smarra, and R. Mangharam, “Data predictive control using regression trees and ensemble learning,” in 2017 IEEE 56th annual conference on decision and control (CDC), IEEE, 2017, pp. 4446–4451.

[14] K. Nti, A. F. Adekoya, and B. A. Weyori, “A comprehensive evaluation of ensemble learning for stock-market prediction,” Journal of Big Data, vol. 7, no. 1, pp. 1–40, 2020.

[15] W. Wang and Y. Lu, “Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model,” in IOP conference series: materials science and engineering, IOP Publishing, 2018, p. 012049.

[16] N. S. Gupta, Y. Mohta, K. Heda, R. Armaan, B. Valarmathi, and G. Arulkumaran, "Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis," Journal of Environmental and Public Health, vol. 2023, 2023, Accessed: Oct. 03, 2023. [Online]. Available: https://www.hindawi.com/journals/jeph/2023/4916267/

[17] V. Kumar, S. Singh, Z. Ahmed, and N. Verma, "Air Pollution Prediction using Machine Learning Algorithms: A Systematic Review," International Journal of Engineering Research & Technology, vol. 11, no. 12, Dec. 2022, doi: 10.17577/IJERTV11IS120026.

[18] M. Vm, S. G. Gh, and S. Kamalapurkar, "Air Pollution Prediction Using Machine Learning Supervised Learning Approach," vol. 9, no. 04, 2020.

[19] "Feature Scaling: Normalization and Standardization - Quinn-Yann - 博客园." Accessed: Jan. 11, 2023. [Online]. Available: https://www.cnblogs.com/quinn-yann/p/9808247.html

[20] E. Koech, "Softmax Activation Function — How It Actually Works," Medium. Accessed: Feb. 17, 2023. [Online]. Available: https://towardsdatascience.com/softmax-activation-function-how-it-actually-works-d292d335bd78

[21] Y. Liu, "Error awareness by lower and upper bounds in ensemble learning," International Journal of Pattern Recognition and Artificial Intelligence, vol. 30, no. 09, p. 1660003, 2016.

[22] A. Mishra, "Metrics to evaluate your machine learning algorithm," Towards data science, pp. 1–8, 2018.