ENHANCING THE ACCURACY IN PREDICTION OF AIR POLLUTION DURING LOCKDOWN USING PREDICTIVE LINEAR REGRESSION COMPARED WITH LOGISTIC REGRESSION

C. Devi[1], S. Christy[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The purpose of this work is to identify the accuracy in prediction of air pollution during the lockdown. **Materials and Methods:** The research work contains two groups namely Predictive linear regression and Logistic regression. Each group consists of a sample size of 10 and the study parameters include an alpha value of 0.05, a beta value of 0.2, and a power value of 0.8. The performance analysis for maximum accuracy in the prediction of air pollution during lockdown using Predictive linear regression over Logistic regression which identifies and predicts the air pollution.
Results and Discussion: The accuracy using Predictive linear regression is 98% is more accurate than the logistic regression of 92% in predicting air pollution.
Conclusion: The Predictive linear regression model is significantly better than the logistic regression in predicting air pollution. It can also be considered a better option for predicting air pollution during the lockdown.

Keywords: Air Quality, Predictive Linear Regression, Logistic Regression, Machine Learning, Air Pollution, Accuracy.

[1]Research Scholar, Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.
[2*]Department of Information Technology, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India. Pincode: 602105.

1. Introduction

Air pollution is a mixture of particles and gases whose sources and composition vary spatially and temporally. Air pollution in the environment derives both from anthropogenic and natural sources. Air pollution is an important environmental risk factor in the propagation of diseases such as lung cancer, autism, asthma, low birth weight, etc. Air pollution differs from place to place and depends on multiple pollutant sources such as industrial emissions, heavy traffic congestion, temperature, pressure, wind, humidity, burning of fossil fuels, etc. (Wang et al. 2021). Regulation of air quality is an important task of the government in developing countries for ensuring people's health and welfare. Applications include the most dangerous air pollutants are Particulate Matter (PM), Nitrogen Oxide (NO_2), Ozone (O_3) and Sulphur Dioxide (SO_2) (Smiianov et al. 2020). The appropriate extraction feature is chosen and trained using Predictive linear regression. It performs a regression task. The kernel function of logistic regression is used to transform the original input set into a higher-dimensional feature space (Ryan, Silver, and Schofield 2021). By examining the relationship between one or more existing independent variables, the Predictive linear regression model forecasts a dependent variable. The application of this research is to provide suggestions for the Air Quality Index (AQI) for effective communication of the air quality status of an area to people in terms, which are easy to understand and is used for decision making in many countries (Baldasano 2020).

In the last 5 years, there have been 87 articles in IEEE explore and 157 in Google Scholar related to this study. Our strategy for examining the causal effect of lockdowns on air quality relies on a comparison of changes in air pollutant species, including particulate matter ($PM_{2.5}$) and (PM_{10}). The proposed framework indicated that predicting air pollution during lockdown may be accomplished with 98% accuracy (Silva, Ávila, and Gonçalves 2021). Two sets of difference-in-differences (DID) analyses are adapted in this study, which provides important advantages over other empirical strategies such as before and after lockdown comparisons and interrupted time series by introducing control groups. Then, the overall government response index from the OxCGRT dataset was added to their DID design to construct a Two-Way Fixed Effects (TWFE) event-study specification according to researchers to handle DID with multiple Periods (Cole, Elliott, and Liu 2020). Besides, this dataset also provides 8 different lockdown measures allowing us to compare the impact on air quality from different lockdown measures (Li et al. 2021). This study examined the evidence of heterogeneous environmental impacts from lockdowns among countries that differed in developmental level, industrial structure, and population. The new study extends past work by investigating if the addition of air pollution improves discrimination and whether a Predictive linear regression could improve it (Ravindra et al. 2021) Attempting to improve the model is the logical next step for air pollution; a novel predictive model in a population-based sample was recently pursued (Praveena and Aris 2021). To predict air pollution during the lockdown, various machine learning methods are used with minimum redundancy and maximum relevance feature selection algorithms to select the most important feature among all the features of the air pollution that is affecting people (Singh et al. 2020).

Our institution is keen on working on latest research trends and has extensive knowledge and research experience which resulted in quality publications (Rinesh et al. 2022; Sundararaman et al. 2022; Mohanavel et al. 2022; Ram et al. 2022; Dinesh Kumar et al. 2022; Vijayalakshmi et al. 2022; Sudhan et al. 2022; Kumar et al. 2022; Sathish et al. 2022; Mahesh et al. 2022; Yaashikaa et al. 2022). Some datasets are aimed at theoretical research rather than being processed as per their real-life application. Most of the existing standard feature processes are for short-term analysis, so researchers have created their features. This research focuses on improved accuracy in the prediction of air pollution during lockdown using predictive linear regression compared to logistic regression.

2. Materials and Methods

This work is carried out at the Machine learning lab in Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences. The study consists of two sample groups, Predictive linear regression, and Logistic regression. Each group consists of 10 samples with a pre-test power of 0.18. The sample was collected from the Prediction of air pollution during lockdown by keeping the threshold at 0.05, GPower at 80%, confidence interval at 95%, and enrolment ratios as 1. The pseudocode for Predictive Linear regression is shown in table 1 and pseudocode for Logistic Regression as shown in table 2. The dataset used for classification is taken from the Kaggle Database of Indian Air Quality, an open-source data respiratory for air pollution ("Kaggle: Your Machine Learning and Data Science Community"). Table 3 represents the preview of the dataset. It contains details about the attribute and its description.

The dataset contains 9 columns and 245 rows. In the $PM_{2.5}$ column, it represents the fine particulate matter that is 2.5 micrometers less in diameter. The dataset was split into training and ("Kaggle: Your Machine Learning and Data Science Community") testing parts accordingly using a test size of 0.2.

The proposed work is designed and implemented with the help of Python3 software. The platform for assessing deep learning was the Windows 10 OS. The hardware configuration was an Intel Core i7 processor with a RAM size of 8GB. The system sort was used for 64-bit. For the implementation of the code, the Python language was used. As for code execution, the datasets are worked behind to perform an output process for accuracy. For training of the Predictive linear regression, the test size is about 20% of the total dataset and the remaining 80% is used for the training set. The whole dataset is fitted for training the Predictive linear regression and Logistic regression model.

Predictive Linear Regression

The regression evaluation is regularly used for prediction and the goal of this version is to assemble a mathematical version that may be applied for expecting the based variable primarily based totally at the inputs of impartial variables or the predictors. LR version has been used to attain the big dating in addition to correlation among the based variable and the predictors or the impartial variable. Predictive linear regression is a representative model, even with its simplistic representation. Its methods are used for predictive analysis and show the relationship between the continuous variables. It shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis). The generic equation for Predictive linear regression is given in equation 1.

$$Y = a + bX + e \qquad \textbf{(1)}$$

Where
a is the intercept
b is the slope of line
e is the error team.

This equation can be used to predict the value of a target variable based on given predictor variables (s)

Logistic Regression

Logistic Regression is the algorithm employed to detect a user-defined sample is polluted or not. Logistic regression is the appropriate regression model to conduct analysis when the dependent variable is dichotomous. For example, here, the data set gets classified into two classes: polluted or not polluted. Like all regression analyses, logistic regression is a predictive analysis.

$$\mathbf{P/(1\text{-}P) = eY} \qquad (2)$$

Here, log(p/1-p) is the link function, this equation 2 is used in logistic regression.

Logistic regression is used to explain the relationship between one or more independent variables and one dependent binary variable. The logistic function is used to generate log odds of an attribute that signifies the probability of the attribute. Log odds are an alternate way of expressing probabilities, which simplifies the process of updating them with new evidence.

Based on the logistic function, the system classifies the training data to be either 0 or 1 and verifies its accuracy using the test data. The result of the user input is also 0/1 and not the PM2.5 level.

STATISTICAL ANALYSIS

For Statistical analysis, the statistical package for the social science version 26 software tool was used. For accuracy, an independent sample T-test was used. The Statistical package for social science (SPSS) software tool was also used to calculate standard deviation mean errors. The significance level of proposed and Existing algorithms are displayed. It contains statistical values of groups of proposed and existing algorithms. The independent variables are Particulate Matter (PM), T(Average Temperature), TM(Maximum Temperature), and Tm(Minimum Temperature), and the dependent variables are accuracy and loss, these are recognized to predict air pollution.

3. Results

The proposed Predictive Linear Regression and Logistic Regression were run at different times in Anaconda Navigator with a sample size of 0.2. These 10 data samples are used for each algorithm along with their loss values to calculate statistical values that can be used for comparison. From the results, it is observed that the mean accuracy of the Predictive linear regression algorithm was 87% and the Logistic regression algorithm was 75%. The mean value for Predictive linear regression is better when compared with the Logistic regression, with a standard deviation of 7.81430 and 11.18603 respectively as shown in table 6. The mean, standard deviation, and standard error mean for Predictive linear regression are 87.3730, 7.81, and 2.47 respectively. On the other hand, the loss values of Logistic regression for mean, standard deviation and standard error mean are 75.9910, 11.18, and 3.53 respectively as shown in table 7. The group statistics values along with mean, standard deviation, and standard error mean for the two algorithms are also specified. In the graphical representation of the comparative analysis, the mean of loss between the two algorithms of Predictive linear regression and Logistic regression are classified. This indicates that Predictive linear regression is significantly better with an accuracy of 87% when compared with Logistic regression accuracy of 75%. Accuracy for Prediction using Predictive Linear Regression with sample size 10 is shown in table 4 and Accuracy for Prediction using Logistic Regression with sample size 10 is shown in table 5. Fig. 1 denotes the comparison of Predictive linear regression and Logistic regression in terms of mean accuracy.

4. Discussion

From the result of this study, Predictive linear regression is proved to be having better accuracy than Logistic regression. Predictive linear regression has an accuracy of 98% whereas a Logistic regression has an accuracy of 89%. The statistical analysis of the two groups shows that Predictive linear regression has more accuracy than Logistic regression. The standard mean error including standard deviation is slightly less than Logistic regression.

On the public tableau, predicted AQI values for the next six months are visualised. This model has a high level of accuracy. Using Predictive linear regression, the predicted AQI has a 96% prediction. Future enhancements will include expanding the region's scope to include as many regions as possible. Currently, this project is aimed at predicting the AQI values of various Bangalore regions. Furthermore, the scope of this project can be expanded to predict AQI for other cities by combining data from different cities (Sethi and Mittal 2020) . Other algorithms, such as artificial neural networks, were able to predict how changes in the level of lockdown affected air quality in Sao Paulo City based on their research. The ANN results showed a Mean Absolute Percentage Error (MAPE) of around 30% even when using a limited data set of pollutant levels combined with meteorological information. According to the current study, Predictive Linear regression shows that the accuracy value is %. As a result, among the algorithms listed above, Predictive linear regression is the best (Shehzad, Sarfraz, and Shah 2020). Predictions based on the pre-lockdown period do not accurately predict the bias during the lockdown. As a result, we use Predictive Linear regression with Cross-Validation over the months for which we want accurate predictions. The model parameters are chosen to maximise the cross-validated root mean square error (RMSE) with a high accuracy range of 95% to 99%. The Po Valley's air quality improved in spring 2020, as it did in many other areas around the world where restrictions on human activities were imposed to limit the spread of the COVID-19 virus. However, air quality improvement appeared to be influenced solely by traffic-related pollutants for the gaseous pollutants studied in this study (NO_2, benzene, and NH_3). Lockdown rules in Northern Italy determined reductions of nitrogen oxide and benzene emissions from road traffic in the order of 35% of the regional average by using a Predictive Linear regression algorithm to prove the reduction of NO_2, Benzene, and NH_3 emissions from road traffic (Lonati and Riva 2021).

The limitations of the proposed work are due to inconsistent data and difficulty in getting the right datasets for analysis. Future work can be concentrated on effective data preprocessing techniques and the usage of ensemble machine learning algorithms can be focused.

5. Conclusion

Based on the experimental results, the Predictive linear regression has been proved to predict the air pollution during lockdown compared to Logistic regression. The quality of datasets formed with air gases and accuracy is improved. Finally, the Predictive Linear regression is significantly better with an accuracy of 87% when compared with Logistic regression accuracy of 75%.

Declarations

Conflicts of Interest

No conflicts of interest in this manuscript.

Author Contributions

Author DC was involved in data collection, data analytics, data extraction, manuscript writing. Author CS was involved in conceptualization, data validation and critical review of the manuscript.

Acknowledgments

The author would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.
1. Vee Eee Technologies Solution Pvt.Ltd., Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Baldasano, José M. 2020. "COVID-19 Lockdown Effects on Air Quality by NO in the Cities of Barcelona and Madrid (Spain)." The Science of the Total Environment 741 (November): 140353.

Cole, Matthew A., Robert J. R. Elliott, and Bowen Liu. 2020. "The Impact of the Wuhan Covid-19 Lockdown on Air Pollution and Health: A Machine Learning and Augmented Synthetic Control Approach." Environmental and Resource Economics. https://doi.org/10.1007/s10640-020-00483-4.

Dinesh Kumar, M., V. Godvin Sharmila, Gopalakrishnan Kumar, Jeong-Hoon Park, Siham Yousuf Al-Qaradawi, and J. Rajesh Banu. 2022. "Surfactant Induced Microwave Disintegration for Enhanced Biohydrogen Production from Macroalgae Biomass: Thermodynamics and Energetics." Bioresource Technology 350 (April): 126904.

Kumar, J. Aravind, J. Aravind Kumar, S. Sathish, T. Krithiga, T. R. Praveenkumar, S. Lokesh, D. Prabu, A. Annam Renita, P. Prakash, and M. Rajasimman. 2022. "A Comprehensive Review on Bio-Hydrogen Production from Brewery Industrial Wastewater and Its Treatment Methodologies." Fuel. https://doi.org/10.1016/j.fuel.2022.123594.

Li, Qiulun, Qingyang Zhu, Muwu Xu, Yu Zhao, K. M. Venkat Narayan, and Yang Liu. 2021. "Estimating the Impact of COVID-19 on the PM Levels in China with a Satellite-Driven Machine Learning Model." Remote Sensing 13 (7). https://doi.org/10.3390/rs13071351.

Lonati, Giovanni, and Federico Riva. 2021. "Regional Scale Impact of the COVID-19 Lockdown on Air Quality: Gaseous Pollutants in the Po Valley, Northern Italy." Atmosphere. https://doi.org/10.3390/atmos12020264.

Mahesh, Narayanan, Srinivasan Balakumar, Uthaman Danya, Shanmugasundaram Shyamalagowri, Palanisamy Suresh Babu, Jeyaseelan Aravind, Murugesan Kamaraj, and Muthusamy Govarthanan. 2022. "A Review on Mitigation of Emerging Contaminants in an Aqueous Environment Using Microbial Bio-Machines as Sustainable Tools: Progress and Limitations." Journal of Water Process Engineering. https://doi.org/10.1016/j.jwpe.2022.102712.

Mohanavel, Vinayagam, K. Ravi Kumar, T. Sathish, Palanivel Velmurugan, Alagar Karthick, M. Ravichandran, Saleh Alfarraj, Hesham S. Almoallim, Shanmugam Sureshkumar, and J. Isaac JoshuaRamesh Lalvani. 2022. "Investigation on Inorganic Salts K2TiF6 and KBF4 to Develop Nanoparticles Based TiB2 Reinforcement Aluminium Composites." Bioinorganic Chemistry and Applications 2022 (January): 8559402.

Praveena, Sarva Mangala, and Ahmad Zaharin Aris. 2021. "The Impacts of COVID-19 on the Environmental Sustainability: A Perspective from the Southeast Asian Region." Environmental Science and Pollution Research International, January. https://doi.org/10.1007/s11356-020-11774-0.

Ram, G. Dinesh, G. Dinesh Ram, S. Praveen Kumar, T. Yuvaraj, Thanikanti Sudhakar Babu, and Karthik Balasubramanian. 2022. "Simulation and Investigation of MEMS Bilayer Solar Energy Harvester for Smart Wireless Sensor Applications." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102102.

Ravindra, Khaiwal, Tanbir Singh, Akash Biswal, Vikas Singh, and Suman Mor. 2021. "Impact of COVID-19 Lockdown on Ambient Air Quality in Megacities of India and Implication for Air Pollution Control Strategies." Environmental Science and Pollution Research. https://doi.org/10.1007/s11356-020-11808-7.

Rinesh, S., K. Maheswari, B. Arthi, P. Sherubha, A. Vijay, S. Sridhar, T. Rajendran, and Yosef Asrat Waji. 2022. "Investigations on Brain Tumor Classification Using Hybrid Machine Learning Algorithms." Journal of Healthcare Engineering 2022 (February): 2761847.

Ryan, Robert G., Jeremy D. Silver, and Robyn Schofield. 2021. "Air Quality and Health Impact of 2019-20 Black Summer Megafires and COVID-19 Lockdown in Melbourne and Sydney, Australia." Environmental Pollution 274 (April): 116498.

Sathish, T., V. Mohanavel, M. Arunkumar, K. Rajan, Manzoore Elahi M. Soudagar, M. A. Mujtaba, Saleh H. Salmen, Sami Al Obaid, H. Fayaz, and S. Sivakumar. 2022. "Utilization of Azadirachta Indica Biodiesel, Ethanol and Diesel Blends for Diesel Engine

Applications with Engine Emission Profile.” Fuel. https://doi.org/10.1016/j.fuel.2022.123798.

Sethi, Jasleen Kaur, and Mamta Mittal. 2020. “Monitoring the Impact of Air Quality on the COVID-19 Fatalities in Delhi, India: Using Machine Learning Techniques.” Disaster Medicine and Public Health Preparedness, October, 1–8.

Shehzad, Khurram, Muddassar Sarfraz, and Syed Ghulam Meran Shah. 2020. “The Impact of COVID-19 as a Necessary Evil on Air Pollution in India during the Lockdown.” Environmental Pollution. https://doi.org/10.1016/j.envpol.2020.115080.

Silva, Pedro, Mariana Ávila, and Márcia Gonçalves. 2021. “Air and Water Quality Improvement during COVID-19 Lockdown.” Proceedings of the 7th International Conference on Geographical Information Systems Theory, Applications and Management. https://doi.org/10.5220/0010397201090115.

Singh, Vikas, Shweta Singh, Akash Biswal, Amit P. Kesarkar, Suman Mor, and Khaiwal Ravindra. 2020. “Diurnal and Temporal Changes in Air Pollution during COVID-19 Strict Lockdown over Different Regions of India.” Environmental Pollution. https://doi.org/10.1016/j.envpol.2020.115368.

Smiianov, Vladyslav A., Oleksii V. Lyulyov, Tetyana V. Pimonenko, Tetyana A. Andrushchenko, Serhii Sova, and Natalia V. Grechkovskaya. 2020. “THE IMPACT OF THE PANDEMIC LOCKDOWN ON AIR POLLUTION, HEALTH AND ECONOMIC GROWTH: SYSTEM DYNAMICS ANALYSIS.” Wiadomosci Lekarskie 73 (11): 2332–38.

Sudhan, M. B., M. Sinthuja, S. Pravinth Raja, J. Amutharaj, G. Charlyn Pushpa Latha, S. Sheeba Rachel, T. Anitha, T. Rajendran, and Yosef Asrat Waji. 2022. “Segmentation and Classification of Glaucoma Using U-Net with Deep Learning Model.” Journal of Healthcare Engineering 2022 (February): 1601354.

Sundararaman, Sathish, J. Aravind Kumar, Prabu Deivasigamani, and Yuvarajan Devarajan. 2022. “Emerging Pharma Residue Contaminants: Occurrence, Monitoring, Risk and Fate Assessment – A Challenge to Water Resource Management.” Science of The Total Environment. https://doi.org/10.1016/j.scitotenv.2022.153897.

Vijayalakshmi, V. J., Prakash Arumugam, A. Ananthi Christy, and R. Brindha. 2022. “Simultaneous Allocation of EV Charging Stations and Renewable Energy Sources: An Elite RERNN-m2MPA Approach.” International Journal of Energy Research. https://doi.org/10.1002/er.7780.

Wang, Nan, Jiawei Xu, Chenglei Pei, Rong Tang, Derong Zhou, Yanning Chen, Mei Li, et al. 2021. “Air Quality During COVID-19 Lockdown in the Yangtze River Delta and the Pearl River Delta: Two Different Responsive Mechanisms to Emission Reductions in China.” Environmental Science & Technology 55 (9): 5721–30.

Yaashikaa, P. R., P. Senthil Kumar, S. Jeevanantham, and R. Saravanan. 2022. “A Review on Bioremediation Approach for Heavy Metal Detoxification and Accumulation in Plants.” Environmental Pollution 301 (May): 119035.

Tables and Figures

Table 1. Pseudocode for Predictive linear regression

//I: Input dataset Records
1. Import the required packages
2. Convert the Data sets into Numerical values after the extraction feature
3. Assign the data to X_train, Y_train, and Y_test variables
4.Using train_test_split() function, pass the training and variables
5. Give test_size and the random_state as parameters for Splitting the data using the Linear Training

model
6. Compiling the model using metrics as accuracy
7. Calculate the Accuracy of the model
OUTPUT: Accuracy

Table 2. Pseudocode for Logistic regression

//I: Input dataset Records
1. Import the required packages
2. Convert the Data sets into Numerical values after the extraction feature
3. Assign the data to X_train, Y_train, and Y_test variables
4. Using train_test_split() function pass the training and variables
5. Give test_size and the random_state as parameters for splitting the data using a Logistic regression model
6. Compiling the model using metrics as accuracy
7. calculate the accuracy of the model
OUTPUT: Accuracy

Table 3. Air quality dataset collected from Kaggle Inc

ATTRIBUTES	**DESCRIPTION**
PM2.5	It defines fine particulate matter. Particle pollution is the term for a mixture of solid particles and liquid droplets found in the air.
T(Average Temperature)	The average temperature of the air is indicated by a properly exposed thermometer during a given period, usually a day, month, or year.
TM(Maximum Temperature)	The maximum temperature is the highest temperature at a place in a given period, and there is no limit.
Tm(Minimum Temperature)	Daily air pollution minima is recorded by the screen thermometer.
SLP(Atmospheric pressure at sea level)	Unit of measurement equal to air pressure at sea level, about 14.7 pounds per square inch.

VM(Maximum sustained wind speed)	It is associated with a common indicator of the intensity of the storm in tropical cyclones.

Table 4. Accuracy for prediction using Predictive Linear regression with sample size 10

Test size	**Accuracy**	**Loss**
Test 1	98.23	1.77
Test 2	96.45	3.55
Test 3	94.37	5.63
Test 4	92.89	7.11
Test 5	89.45	10.55
Test 6	83.76	16.24
Test 7	82.39	17.61
Test 8	80.23	19.77
Test 9	78.32	21.68
Test 10	77.64	22.36

Table 5. Accuracy for Prediction using Logistic regression with sample size 10.

Test size	**Accuracy**	**Loss**
Test 1	92.06	7.94
Test 2	89.32	10.68
Test 3	84.31	15.69
Test 4	82.46	17.54
Test 5	77.65	22.35
Test 6	73.52	26.48
Test 7	71.23	28.77
Test 8	68.28	31.72
Test 9	62.86	37.14
Test 10	58.22	41.78

Table 6. Group Static Analysis, Representing Predictive Linear regression(mean accuracy 87.3730, standard deviation 7.81430) and Logistic Regression(mean accuracy 75.9910, standard deviation 11.18603)

	Algorithm	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**
Accuracy	Predictive linear regression	10	87.3730	7.81430	2.47110
	Logistic regression	10	75.9910	11.18603	3.53733
Error	Predictive linear regression	10	12.6270	7.81430	2.47110
	Logistic regression	10	24.0090	11.18603	3.53733

Table 7. Independent Sample Tests results with a confidence interval at 95% and level of significance as 0.05 (Predictive Linear regression appears to perform significantly better than Support Vector Machine with the value of p=0.032).

Levene's Test for Equality of Variances				**T-test for Equality of Means**						
							Mean Difference	**Std.Error Difference**		
Accuracy	Equal variances assum							.04013		

	ed									
	Equal variances not assumed							.04013		
Error	Equal variances assumed							.04013		

	Equal variance not assumed							.04013		

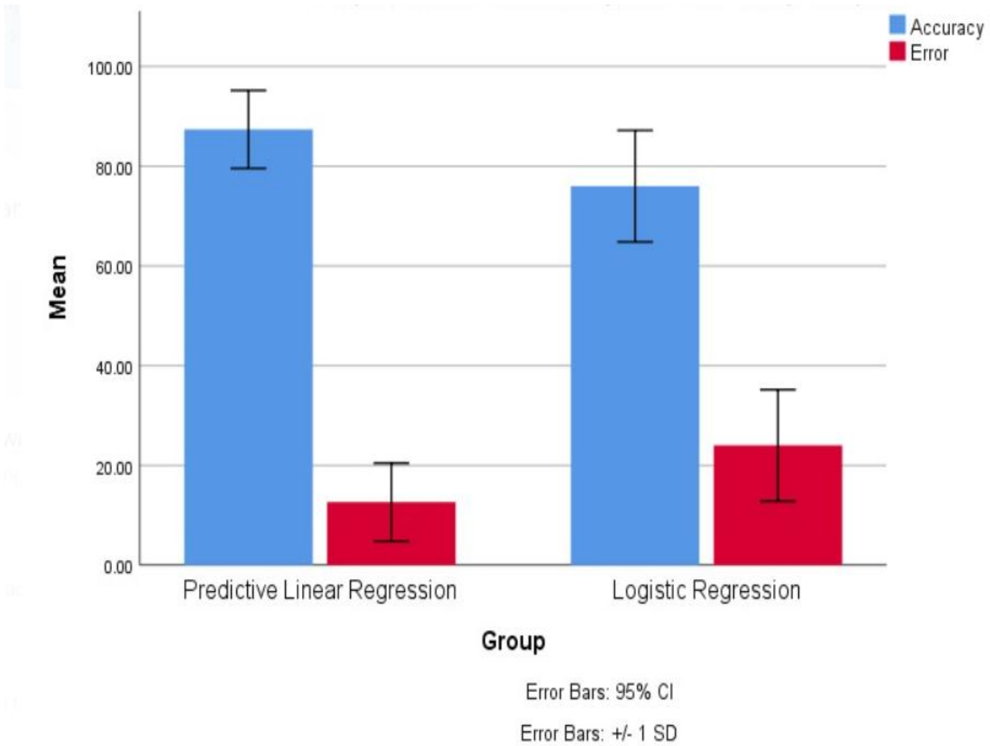

Fig. 1. Comparison of Predictive Linear regression and Logistic Regression in terms of accuracy. The mean accuracy of Predictive Linear regression is greater than Logistic Regression and the standard deviation is also slightly higher than Logistic Regression. X-axis: Predictive linear regression vs Logistic Regression. Y-axis: Mean accuracy and Error of prediction + 1 SD.