**A Major Project**

**On**

**REGRESSION TECHNIQUES BASED WHETHER FORECASTING PREDICTION**

**Submitted to**

**The Department of Information Technology**

In partial fulfilment of the academic requirements of

Jawaharlal Nehru Technological University for

The award of the degree of

**Bachelor of Technology**

**In**

**Information Technology**

(2020-2024)

SAI TRILOCHAN 20311A12G7

G. SANGAMESH 21315A1218

NITHIN GUNDA 20311A12H8

Under the Guidance of

Dr. Subhani Shaik

Associate Professor

**Sreenidhi Institute of Science and Technology**



(An Autonomous Institution)

Yamnampet, Ghatkesar, R.R. District, Hyderabad – 501301

2024

**Sreenidhi Institute of Science and Technology**



**Department of Information Technology**

**CERTIFICATE**

This is to certify that this project-II report on “**REGRESSION TECHNIQUES BASED WHETHER FORECASTING PREDICTION”,** submitted by

SAI TRILOCHAN (20311A12G7), G.SANGAMESH (21315A1218), NITHIN GUNDA (20311A12H8), in the year 2024 in partial fulfilment of the academic requirements of Jawaharlal Nehru Technological University for the award of the degree of Bachelor of Technology in **INFORMATION TECHNOLOY ,** is a bonafide work that has been carried out by them as part of their **project during fourth year second semester ,** under our guidance . This report has not been submitted to any other institute or university for the award of any degree.

**Internal guide Project Co-ordinator Head of Department**

Dr. Subhani Shaik Dr. Subhani Shaik Dr. Sunil Bhutada

Associate Professor Associate Professor Professor

Department of IT Department of IT Department of IT

**External Examiner**

**Date:**

**DECLARATION**

We **SAI TRILOCHAN (20311A12G7),** **G.SANGAMESH (21315A1218),**

**NITHIN GUNDA (20311A12H8), students of Sreenidhi Institute of Science and Technology, Yamnampet, Ghatkesar, studying IV-year II semester, INFORMATION TECHNOLOGY** **solemnly declare that the project-2 work, titled “Comparative Study of Ensemble Machine Learning Techniques for Airline Delay Prediction” is submitted to Sreenidhi Institute of Science and Technology for partial fulfilment for the award od degree of Bachelor of Technology in INFORMATION TECHNOLOGY.**

It is declared to the best of our knowledge that the work reported does not form part of any dissertation submitted to any other University or Institute for award of any degree.

**SAI TRILOCHAN ( 20311A12G7)**

**G. SANGAMESH (21315A1218)**

**NITHIN GUNDA (20311A12H8**

**ACKNOWLEDGEMENT**

We would like to express my gratitude to all the people behind the screen who helped us to transform an idea into a real application.

We would like to express my heart-felt gratitude to our parents without whom we would not have been privileged to achieve and fulfil my dreams.

We are grateful to our CEO-**Mr. K. Abhijit Rao,** Director- **Prof.Dr.C.V. Tomy,**

Principal- **Dr. T. Ch. Siva Reddy,** who most ably runs the institution and has had the major hand in enabling me to do my project.

We would like to thank our Coordinator & Internal guide Dr. **Subhani Shaik** for Project-II for their technical guidance, constant guidelines, encouragement, and support in carting out my project on time at my college.

The satisfaction and euphoria that accompany the successful completion of the task would be great but incomplete without the mention of the people who made it possible with their constant guidance and encouragement crowns all the efforts with success. In this context, I would like to thank all the other staff members, both teaching and non-teaching, who have extended their timely help and eased my task.

**SAI TRILOCHAN ( 20311A12G7)**

**G. SANGAMESH (21315A1218)**

**NITHIN GUNDA (20311A12H8)**

**ABSTRACT**

Whether forecasting is most important thing for observe the atmosphere observations by location and time. Regression techniques are a mainstay of information and have been chosen into numerical machine learning concepts. This may be perplexing because we can use regression to refer to the class of problem and algorithm. This paper provides the temperature predictions in weather forecasting time to time using regression-based techniques. We have to use four techniques like Ordinary Least Squares Regression, Logistic Regression, Multivariate Adaptive Regression Splines and Locally Estimated Scatterplot Smoothing. We focus on mean square error rate of each technique. Different types of metrics are test in this research. Low MSE is indicate for better weather forecasting results. The statistical results consider the less MSE rate for optimal.

**Keywords:** Regression techniques, weather forecasting, Prediction, MSE

**INDEX**

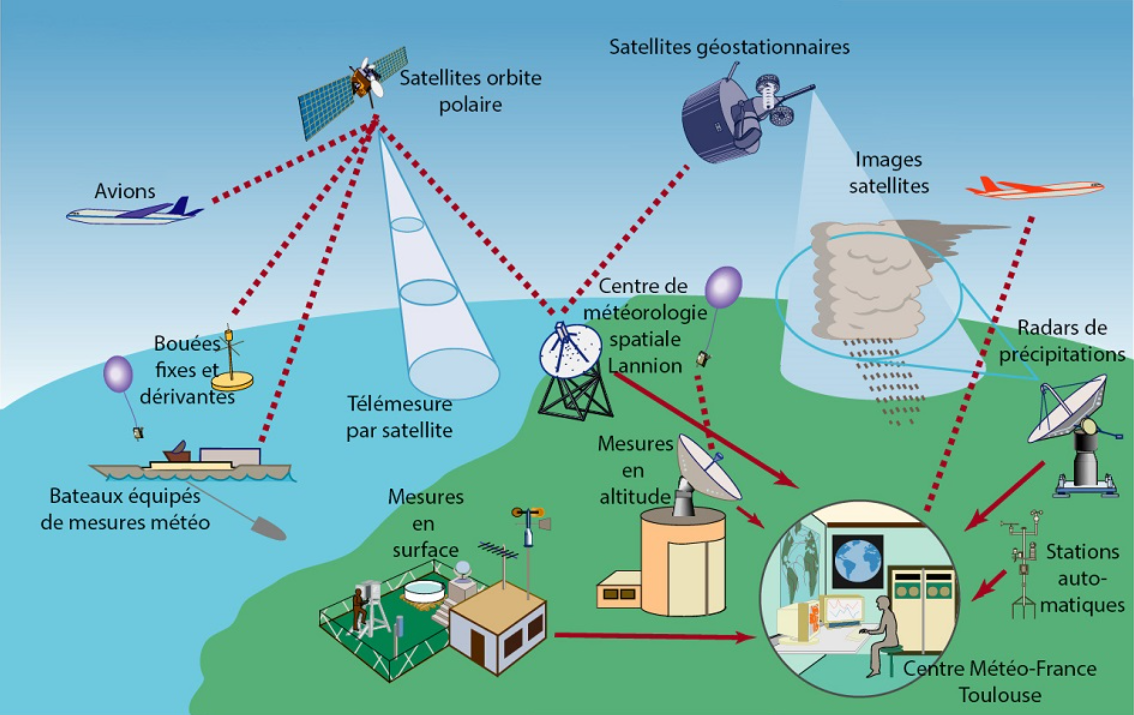
|  |  |
| --- | --- |
| **Contents** | **Page no** |
| 1. **Introduction** | 1-2 |
| 1. **Proposed Methods and Architecture** | 3 |
| 1. **Results and Analysis** | 4 |
| 3.1 Dataset Description | 4 |
| 3.2 Data Preprocessing | 4 |
| 3.3 Data Visualization | 5-8 |
| 3.4 Regression Analysis on Dataset | 9 |
| 3.5 Ordinary Least Squares Regression (OLSR) | 9 |
| 3.6 Logistic Regression | 10 |
| 3.7 Multivariate Adaptive Regression Splines (MARS) | 11 |
| 3.8 Locally Estimated Scatterplot Smoothing (LOESS) | 12 |
| Comparative Study | **13** |
| 4.1 Metrics for Error Rate | 13 |
| 4.1.1 Mean square error | 13 |
| 4.1.2 Mean Absolute error | 13 |
| 4.1.3 R Squared Value | 13 |
| 4.1.4 Mean Square Logarithmic Error (MSLE) | 13 |
| 4.1.5 Variance | 14 |
| 4.1.6 Max Error | 14 |
| 4.1.7 Median Absolute Error | 14 |
| 4.1.8 Huber Loss | 14 |
| Conclusion | 16 |
| References | 17-18 |

**CHAPTER 1**

**INT1RODUCTION**

Weather forecasting is most important scientific technique to predict the status of atmosphere at certain time and locations. In past days weather forecasting carried out by manually alters barometric pressure and current weather conditions. But now it measures on computer-based models that consider many factors for predict [1]. The researchers connect linear relationship between input attribute to corresponding target attribute. This procedure is different in non linear prediction of the weather with multiple attribute relations. Weather forecasting make by gathering quantitative data about the present status and past trend of the atmosphere by using technology to predict the conditions of atmosphere analysis. Weather conditions most important for farmers, business and normal public for protection of life and properties [2]. Actually, machine learning based weather forecasting started from 2018. Initially creating medium range weather forecasting for machine learning. For benchmark problems machine learning play key role in many research areas. Due to this accessibility researchers work on wide variety of backgrounds [3].

Machine learning based weather forecasting is an initial technology for predict the weather conditions in high quality [4]. These machine learning models take less time and resources in single attempt. Machine learning based techniques heavily used for weather forecasting throughout the world [6]. The procedure train and validation of the data continuously generate the accurate results for weather conditions. Quality forecasting is most important for daily life of public [7].



**Figure 1:** system in operational methodology [5]

The figure 1 shows the system for weather forecasting analysis of data from outside environment. Different fields connected with weather forecasting for successfully execute their work in proper time without any loss.

The following paper discuss the proposed method and architecture in section 2. Section 3 state the results and analysis. Final section concludes the paper.

**CHAPTER II**

**PROPOSED METHODS AND ARCHITECTURE**

Regression techniques are a mainstay of information and have been chosen into numerical machine learning concepts. This may be perplexing because we can use regression to refer to the class of problem and algorithm [8]. This paper provides the temperature predictions in weather forecasting time to time using regression-based techniques. We have to use four techniques like Ordinary Least Squares Regression, Logistic Regression, Multivariate Adaptive Regression Splines and Locally Estimated Scatterplot Smoothing. We focus on mean square error rate of each technique. The following figure 2 is a proposed architecture of our research. Different phases included in this architecture for weather forecasting.

1. Environment (outside world)
2. Weather sensors (perceive the information for environment)
3. Model Implementation (suitable model prepare based on algorithms)
4. Classification (classify the data based on certain conditions)
5. Weather conditions (predict the results)

Weather Conditions

Classification

Environment

Weather Sensors

Model Implementation

**Figure 2:** block diagram of proposed system [9]

**CHAPTER III**

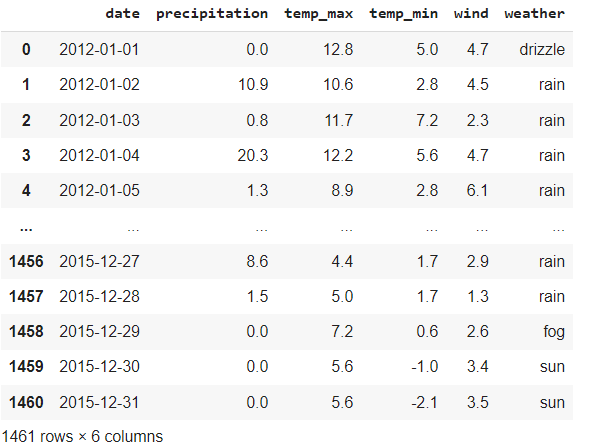
**RESULTS AND ANALYSIS**

The statistical results generated using python programming environment with Jupiter notebook tool. Different types of regression techniques used for predicting the result of weather conditions.

## 3.1 Dataset description

Upload the dataset into system for next preparation of data for classification. It consists of six attributes and 1461 records.

**Table 1:** Dataset [10]

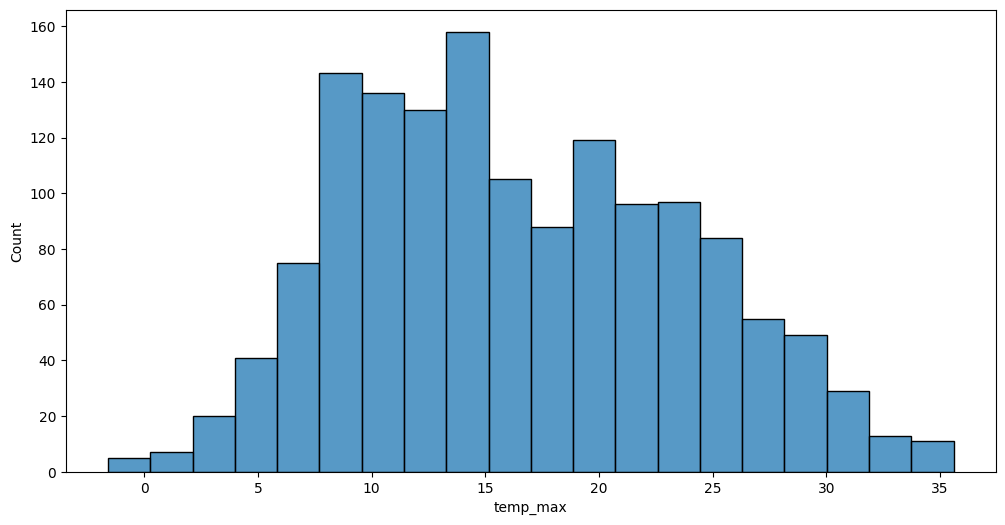


## 3.2 Data Preprocessing

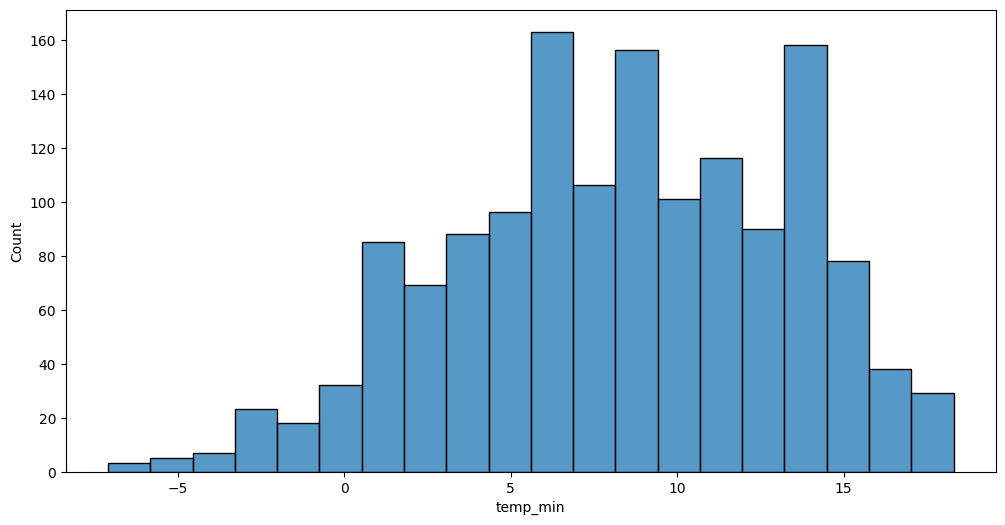
Data preprocessing is a procedure of making the pure data from given raw data. Prepare data for suitable machine learning models. It is the initial and vital step while generating a machine learning model [11].

## 3.3 Data Visualization

## The following figures 3 and 4 shows the data visualization of temperature in minimum and maximum. [12]



**Figure 3:** graph for data visualization in maximum temperature



**Figure 4:** graph for data visualization in minimum temperature

### The following figure 5 represents the maximum temperature in each month in each year from 2019 to 2023.

### 

**Figure 5:** data visualization max temperature in each month in each year

### The following figure 6 represents the minimum temperature in each month in each year from 2019 to 2023.

## 

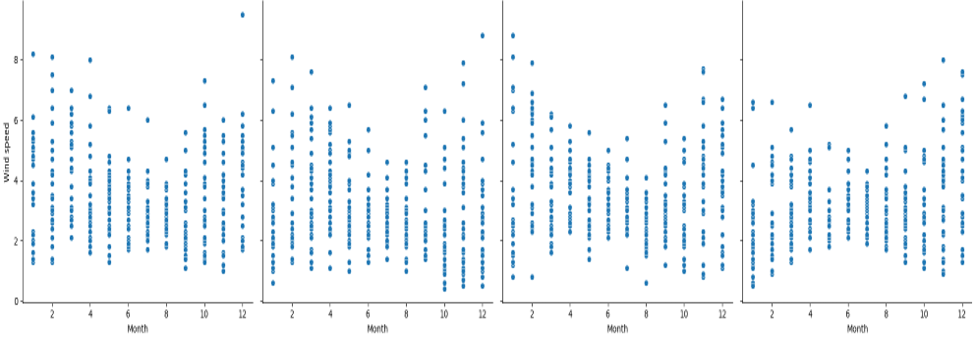
**Figure 6:** data visualization min temperature in each month in each year

## A form of water, such as rain, snow, or sleet, that condenses from the atmosphere, becomes too heavy to remain suspended, and falls to the Earth's surface. The following figure 5.6 shows the overall precipitation in each month in each year and 7 shows the Precipitation in wind speed in each month in each year [13].

## 

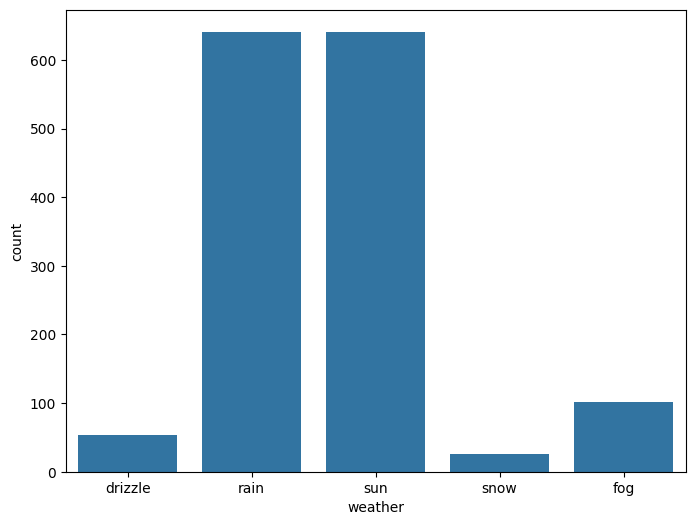
## Figure 7: Precipitation in each month in each year

## 

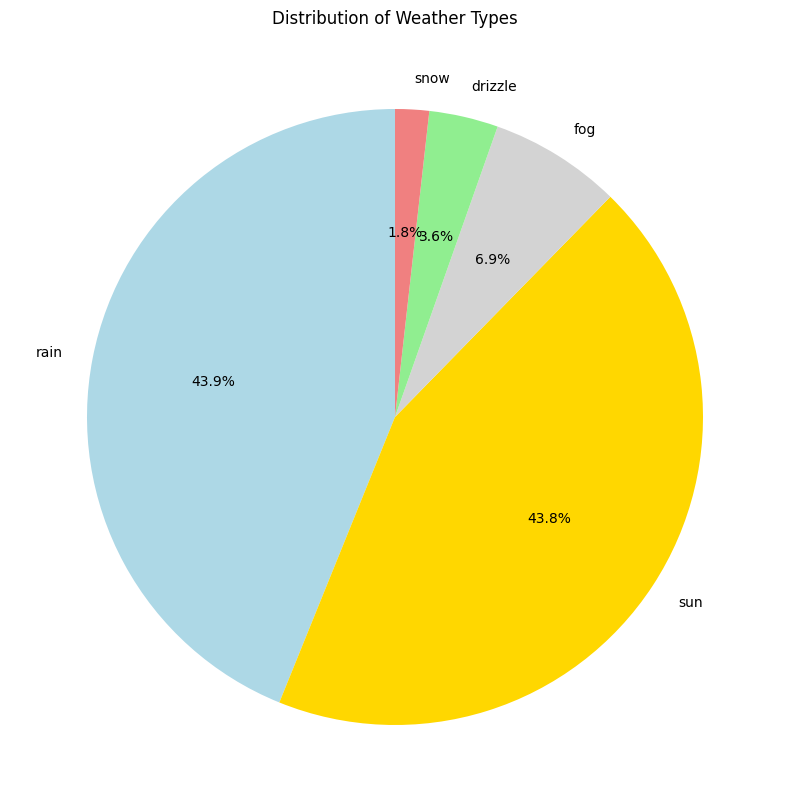


## Figure 8: Precipitation in each month in each year (wind speed)

The following figure 8shows thedata visualization of different weather conditions like fog, snow, rain, sun and dazzle. Figure 9shows thedata visualization for distribution of weather types in percentage.



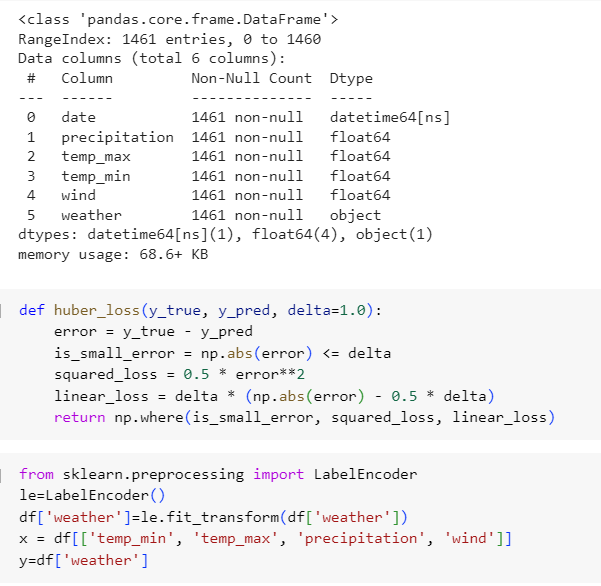
**Figure 9:** data visualization of different weather conditions



**Figure 10:** data visualization of different weather conditions

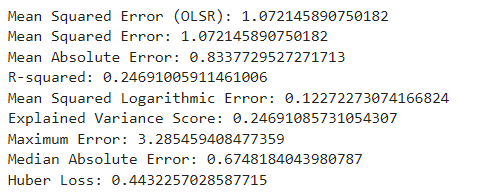
# 3.4 Regressoion Analysis of dataset

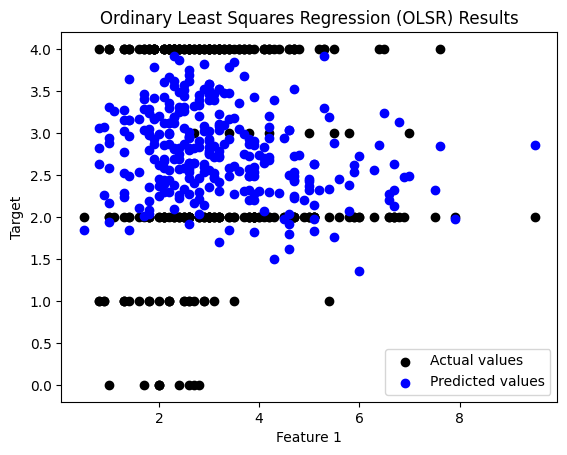
Regression analysis is a set of numerical methods used for the assessment of associations between a dependent variable and independentvariables.



## 3.5 Ordinary Least Squares Regression (OLSR)

## This method OLSR generate the MSE is 1.07, MAE is 0.83, MSLE is 0.24, maximum error is 3.28, median absolute error is 0.67 and huber loss is 0.44.

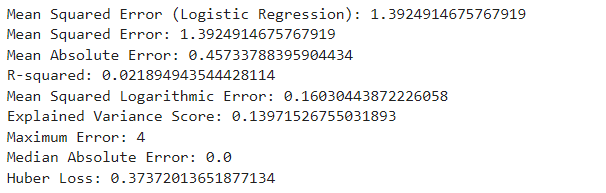


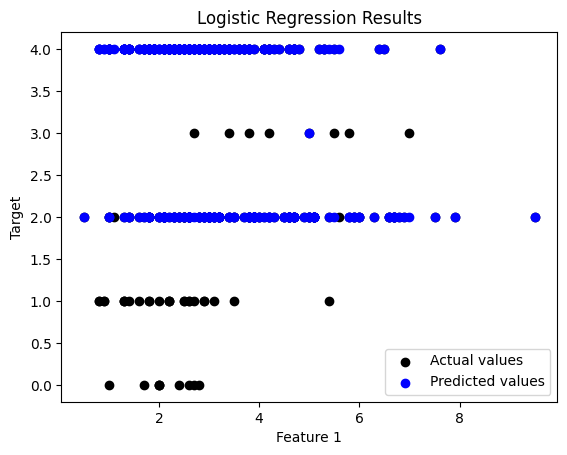


**Figure 11:** OLSR results for actual vs target values

## 3.6 Logistic Regression

## This method Logistic regression generate the MSE is 1.39, MAE is 0.45, MSLE is 0.16, maximum error is 4, median absolute error is 0.0 and huber loss is 0.37.





**Figure 12:** Logistic Regression results for actual vs target values

## 3.7 Multivariate Adaptive Regression Splines (MARS)

## This method MARS generate the MSE is 1.99, MAE is 0.75, MSLE is 0.12, maximum error is 3.27, median absolute error is 0.60 and huber loss is 0.39.

## 

## 

**Figure 13:** MARS results for actual vs target values

## 3.8 Locally Estimated Scatterplot Smoothing (LOESS)

## This method LOESS generates the MSE is 1.26, MAE is 0.97, MSLE is 0.13, R squared is 0.11, Variance 0.11, Maximum error is 3.01, Median absolute error is 0.94 and Huber loss is 0.54.

## 

## 

**Figure 14:** LOESS results for actual vs target values

**CHAPTER IV**

**COMPARATIVE STUDY**

**4.1 Metrics for Error Rate**

**4.1.1 Mean square error**

It is calculated by taking the average, specifically the mean, of errors squared from data as it relates to a function. Among four models MARS and OLSR get less MSE values. Less value prefers in MSE context for good accuracy.

**4.1.2 Mean Absolute error**

MAE takes the average of absolute errors for a group of predictions and observations as a measurement of the magnitude of errors for the entire group.

The MAE is higher than the MSE, which suggests that your model has some small errors that accumulate in the MAE. Less than 5% is most acceptable MAE in research.

Four models get less than 1 MAE. It shows considered all models. Among four models’ logistic regression get less amount of MAE. It prefers more than other models. Lower values prefer mostly than higher. In outlier point of view, we prefer MAE than MSE because MAE is more relevant to error rate prediction.

**4.1.3 R Squared Value**

R-squared is a statistical measure that represents the goodness of fit of a regression model. The value of R-square lies between 0 to 1. All models are accepted in the sense of R squared value.

**4.1.4 Mean Square Logarithmic Error (MSLE)**

It is similar to mean square error metric. It is called the Mean Squared Logarithmic Error. It can also be interpreted as a measure of the ratio between the actual and predicted values. Among four models get less MSLE values. Less value prefers in MSLE context for good accuracy.

**4.1.5 Variance**

Variance refers to the changes in the model when using different portions of the training data set. So, it is required to make a balance between bias and variance errors. For an accurate prediction of the model, algorithms need a low variance and low bias.

Among four models Logistic Regression and LOESS preferable in variance context. Because these models predict less values.

**4.1.6 Max Error**

The Max-Error metric is the worst-case error between the predicted value and the true value.

Among four models OLSR, MARS and LOESS get less error value in max error metric. LOESS more less error value among other two. Logistic regression gets highest error value.

**4.1.7 Median Absolute Error**

The Median Absolute Error is the median difference between the observations (true values) and model output (predictions).

Lower values are preferable to accurate result. Here Logistic regression model consider for better to generate perfect accuracy.

**4.1.8 Huber Loss**

Huber Loss is a popular loss function used in machine learning for regression tasks.

Huber Loss is a hybrid between MSE and MAE and is designed to provide the benefits of both loss functions. Huber Loss has the advantage of being less sensitive to outliers than MSE while still providing a more balanced approach to evaluating the performance of a regression model compared to MAE.

Huber loss metric also prefer low value for best model in machine learning. Four models are generating less than 1 value. Logistic regression generates 0.37 less value compares to others.

Table 1: Comparative analysis of four models

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Name of the Regressor** | **MSE** | **MAE** | **R squared**  **Value** | **MSLE** | **Variance** | **Max Error** | **Median AE** | **Huber Loss** |
| OLSR | 1.07 | 0.83 | 0.24 | 0.12 | 0.24 | 3.28 | 0.67 | 0.44 |
| Logistic Regression | 1.39 | 0.45 | 0.02 | 0.16 | 0.13 | 4 | 0 | 0.37 |
| MARS | 0.99 | 0.75 | 0.30 | 0.12 | 0.30 | 3.27 | 0.60 | 0.39 |
| LOESS | 1.26 | 0.97 | 0.11 | 0.13 | 0.11 | 3.01 | 0.94 | 0.54 |

In our research used four models in different metrics like MSE, MAE, R Squared value, MSLE, Variance, Median Absolute Error, Max Error and Huber loss.

Among all metrics most preferable metrics are MSE, MAE, and Huber loss. Among these three metrics MAE most preferable for outliers. Logistic regression get 0.45 for MAE value. The is less compare to other models.

But one thing is remembering all these metrics values based on dataset size, dataset type, training data, test data and pre-processing of dataset.

**CHAPTER V**

**CONCLUSION**

Regression techniques are a mainstay of information and have been chosen into numerical machine learning concepts. This may be perplexing because we can use regression to refer to the class of problem and algorithm. This paper provides the temperature predictions in weather forecasting time to time using regression-based techniques. We have to use four techniques like Ordinary Least Squares Regression, Logistic Regression, Multivariate Adaptive Regression Splines and Locally Estimated Scatterplot Smoothing. We focus on mean square error rate of each technique. Different types of metrics are tested in this research. Low MSE indicates better weather forecasting results. The statistical results consider the less MSE rate for optimal.

**CHAPTER VI**

**REFERENCES**

[1] Imran Maqsood, Muhammad Riaz Khan, and Ajith Abraham, “An ensemble of neural network for weather forecasting”, Neural Computing & Application (2004) 13: 112–122.

[2] Amanpreet Kaur, JK Sharma, and Sunil Agrawal, “Artificial neural networks in forecasting maximum and minimum relative humidity”. International Journal of Computer Science and Network Security, 11(5):197-199, May 2011.

[3] https://rp5.ru/aRussianwebsitewithweatherdata of airport weather station from all around the world

[4] https://colab.research.google.com/ Google’s free cloud service for AI developers for educational or research purposes.

[5] <https://www.encyclopedie-environnement.org/en/air-en/introduction-weather-forecasting/>

[6] Culclasure, Andrew, "Using Neural Networks to Provide Local Weather Forecasts" (2013). Electronic Theses and Dissertations.32. https://digitalcommons.georgiasouthern.edu/etd/32

[7] Trebing, K., Staǹczyk, T. and Mehrkanoon, S., 2021. Smaat-unet: Precipitation nowcasting using a small attention-unet architecture. Pattern Recognition Letters, 145, pp.178-186.

[8] Mr. Sujan Reddy, Ms. Renu Sri andSubhani Shaik,” Sentimental Analysis using Logistic Regression”, International Journal of Engineering Research and Applications (IJERA), Vol.11, Series-2, July-2021.

[9] Ms. Mamatha, Srinivasa Datta and Subhani Shaik,” Fake Profile Identification using Machine Learning Algorithms”, International Journal of Engineering Research and Applications (IJERA), Vol.11, Series-2, July-2021.

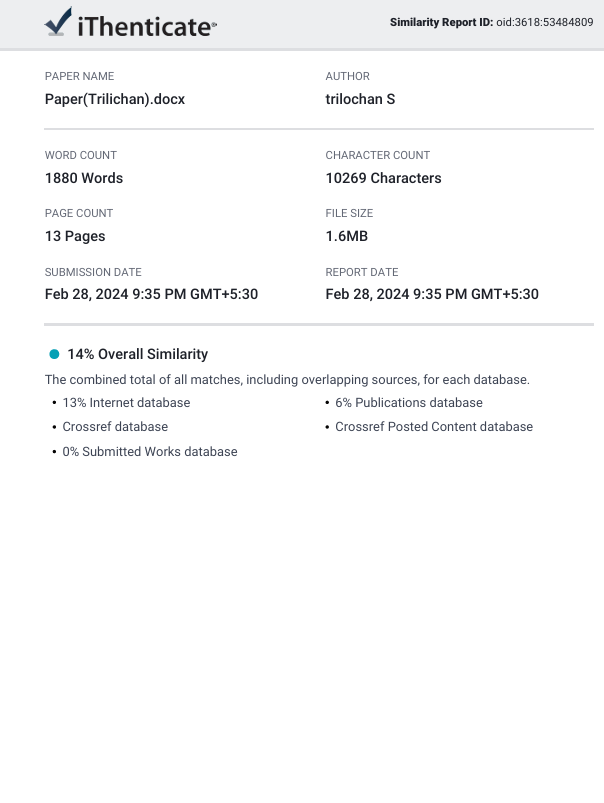
[10] R. Vijaya Kumar Reddy, Subhani Shaik,B. Srinivasa Rao, “Machine learning based outlier detection for medical data” Indonesian Journal of Electrical Engineering and Computer Science, Vol. 24, No. 1, October 2021.

[11] [M. Jagan Chowhaan](https://journalajrcos.com/index.php/AJRCOS/article/view/339),  [D. Nitish](https://journalajrcos.com/index.php/AJRCOS/article/view/339),  [G. Akash](https://journalajrcos.com/index.php/AJRCOS/article/view/339), [Nelli Sreevidya](https://journalajrcos.com/index.php/AJRCOS/article/view/339), Dr. Subhani Shaik, “Machine Learning Approach for House Price Prediction” Asian Journal of Research in Computer Science , Volume 16, Issue 2, Page 54-61, June- 2023.

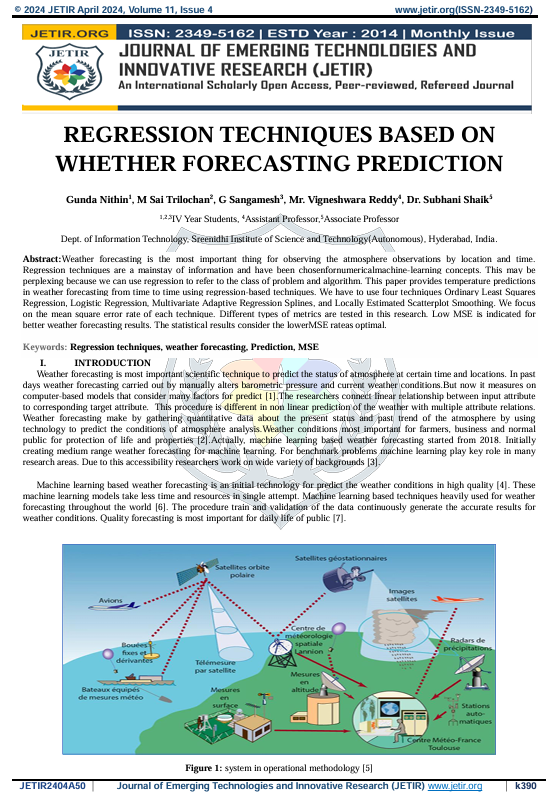
[12] Neeraja, Anupam, Sriram, Subhani Shaik and V. Kakulapati,” Fraud Detection of AD Clicks Using Machine Learning Techniques”, Journal of Scientific Research and Reports, Volume 29, Issue 7, Page 84-89, June-2023.

[13] P. Pranathi, V. Revathi, P. Varshitha, Subhani Shaik and Sunil Bhutada,” Logistic Regression Based Cyber Harassment Identification”, Journal of Advances in Mathematics and Computer Science, Volume 38, Issue 8, Page 76-85, June-2023.

**PLAGARISM REPORT**



**PAPER FIRST PAGE**



**APPENDIX C**

# ABSTRACT

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Sreenidhi Institute of Science and Technology Department of Information Technology**  **Major Project** | | | | | |  | |
| **BATCH NO:** | | | **04** | |  | |  | |
|  | | | | | | | | |
|  | **Roll No.** | | | **Student Name** | | **Title of the Project** | | |
| 20311A12C2 | | | N.SHYAMALIKA | | Enhanced Fashion Recommendation System Based on Size and Color Preferences Using Machine Learning | | |
| 20311A12C3 | | | V.NAVNEETH | |
| 20311A12F2 | | | J.NAVALIKA | |

Fashion recommendation systems play a crucial role in enhancing online shopping experiences by providing personalized suggestions tailored to individual preferences. However, existing systems often face limitations in accurately matching user preferences, particularly concerning size and color preferences for both men and women. This abstract presents an enhanced fashion recommendation system leveraging machine learning techniques to address these shortcomings. We created our own dataset for women and men with parameters based on gender and colour variations. Dataset is trained with multiple machine learning algorithms and accuracy of the models are compared and flask web application is developed to show predicted recommendation based on user sizes. The proposed fashion recommendation system represents a significant advancement over existing systems by addressing their limitations and offering personalized recommendations based on size and color preferences for both men and women. Through the integration of machine learning techniques, the system enhances accuracy, usability, and user satisfaction in the online fashion shopping domain.

# APPENDIX D

**CORRELATION BETWEEN THE GROUP PROJECT - I AND THE PROGRAM OUTCOMES (POS), PROGRAM SPECIFIC OUTCOMES (PSOS)**

|  |  |  |
| --- | --- | --- |
| **ROLL NO** | **NAME** | **TITLE** |
| 20311A12C2 | N.SHYAMALIKA | **REGRESSION TECHNIQUES BASED WHETHER FORECASTING PREDICTION** |
| 20311A12C3 | V.NAVNEETH |
| 20311A12F2 | J.NAVALIKA |

**Table 1: Project correlation with appropriate POs/PSOs**(Please specify the level of Correlation, H/M/L against POs/PSOs)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **H** | **High** | **M** | **Moderate** | **L** | **Low** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **SREENIDHI INSTITUTE OF SCIENCE AND TECHNOLOGY DEPARTMENT OF INFORMATION TECHNOLOGY**  **Projects Correlation with POs/PSOs** | | | | | | | | | | | | | | |
| **PO 1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** | **PSO1** | **PSO2** | **PSO3** |
| M | M | L | H | M | L | M | L | M | M | L | M | H | H | H |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **Students** | **Project Coordinator** | **HOD** |
| N.Shyamalika | Dr.Subhani Shaik | Dr.Sunil Bhutada |
| V.Navneeth | Asso.Prof & Asso.Head | Prof & HOD |
| J.Navalika | Dept of IT | Dept of IT |

# APPENDIX E

## Table 1: Nature of the Project (Please tick √ Appropriate for your project)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Title** | **Nature of Project** | | |
| **Product** | **Application** | **Research** |
|  | Enhanced Fashion Recommendation System Based  on Size and Color Preferences Using  Machine Learning |  |  | √ |

**Table 2: Domain of the Project (Please tick √ Appropriate for your project)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.No** | **Title** | **Domain of the Project** | | | | |
| **ARTIFICIAL INTELLIGEN CE, MACHINE LEARNING AND DEEP**  **LEARNING** | **COMPUTER NETWORKS, INFORMATION SECURITY, CYBER SECURITY** | **Data Warehousing Data Mining Big Data Analysis** | **CLOUD COMPU TING, INTERN ET OF THINGS** | **SOFTW**  **-ARE ENGINE ERING, IMAGE PROCE SSING** |
| **01** | **REGRESSION TECHNIQUES**  **BASED WHETHER FORECASTIG PREDICTION** | √ |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **Students** | **Project Coordinator** | **HOD** |
| M Sai Trilochan | Dr.Subhani Shaik | Dr.Sunil Bhutada |
| G .Sangamesh | Asso.Prof & Asso.Head | Prof & Head |
| Gunda Nithin | Dept of IT | Dept of IT |