African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 162-171)
www.abjournals.org

COMPARATIVE ANALYSIS OF WEATHER PREDICTION USING CLASSIFICATION ALGORITHM: RANDOM FOREST CLASSIFIER, DECISION TREE CLASSIFIER AND EXTRA TREE CLASSIFIER

Oni Oluwabunmi Ayankemi[1*], Iskilu Zainab Adesola[2] and Lawrence Adeolu[3]

[1-3]Department of Computer Studies, Faculty of Science, The Polytechnic Ibadan, Ibadan.

[*]Corresponding Author's Email: onioluwabunmia@gmail.com

Cite this article:

Oluwabunmi O. A., Zainab I. A., Adeolu L. (2024), Comparative Analysis of Weather Prediction Using Classification Algorithm: Random Forest Classifier, Decision Tree Classifier and Extra Tree Classifier. African Journal of Mathematics and Statistics Studies 7(2), 162-171. DOI: 10.52589/AJMSS-F6H03BNE

Manuscript History

Received: 18 Jan 2024

Accepted: 23 Apr 2024

Published: 17 May 2024

ABSTRACT: *Comparison of machine learning models is carried out in order to determine which models are best to deploy as a system. However, for the purpose of our research, we carried out a comparative analysis on Random Forest classifier, Decision Tree classifier and Extra Tree classifier for weather prediction systems as we focused on seeking the classifier with the highest performance metrics. Based on the metrics, accuracy score, the best model for the system was determined. We carried out training, testing and validation of the three different models on the same dataset from the Kaggle dataset. We were able to implement Random Forest Classifier, Decision Tree Classifier and Extra Tree Classifier from Scikit-Learn to make weather prediction and using matplotlib to visualize the accuracy score of the implemented models. The Random Forest Classifier was chosen as the best able to achieve the highest at 66% accuracy.*

KEYWORDS: Weather, Prediction, Classification, Decision trees, Random forest, Logistic regression, Support vector machine.

Article DOI: 10.52589/AJMSS-F6H03BNE
DOI URL: https://doi.org/10.52589/AJMSS-F6H03BNE

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 162-171)

www.abjournals.org

INTRODUCTION

Weather forecasting has been a standout amongst the most experimentally and technologically troublesome issues over the world in the most recent century. Environmental change has been looking for a great deal of consideration for a long time because of the sudden changes that happen. There are several limitations in better execution of weather forecasting thus it ends up hard predicting weather here and now with effectiveness.

Various Machine Learning Techniques are applied on weather data to predict climate parameters like temperature, wind speed, rainfall, meteorological pollution. A lot of human activities are dependent on the weather condition, in the past hazardous weather events have caused humans a lot of damage and losses. A timely and accurate prediction of the weather condition can help humans plan better and help reduce disastrous weather events. The aim of this study is to develop a weather prediction system using **Random Forest**, **Decision Tree** and **Extra Tree Classification Algorithm**. The model built from this model will be able to predict the weather condition from the data provided to it. The method of collecting and sourcing required information and data for this project is secondary data collection and the data set used for this project is obtained from Kaggle (www.kaggle.com/dataset).

RELATED WORKS

Pratyush Muthukumar et al. (2021) made an attempt to predict PM2.5 atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data. For the research, they proposed a two-stage model capable of learning spatiotemporal trends based on remote-sensing satellite imagery of air pollution and data of ground-based sensors monitoring air pollutants and meteorological features. Los Angeles was their focus during their research. In their conclusion, they said that the result of their research could match to explain various real-world events, chemical processes, and physical processes of ground-based PM2.5 in Los Angeles county.

Manepalli et al. (2020) carried out a research titled Weather Prediction Through Machine Learning. The objective of their work is to design an effective rainfall prediction agent model using support vector machines and multiple linear regressions. To evaluate the proposed model, it was implemented using MATLAB and compared with existing numerical models. The algorithm used in this research work was Decision Tree and Linear Regression. Two sets of experiments are conducted on the data, time series forecasting agent model given using SVR, secondly working on multivariate regression problem as visualizing agent model. This statistical learning method was evaluated as the best capability of forecast rainfall prediction with respect to temperature correlation existence. The capability of generalization of the agent model due to the input data structure feed is tested for RMSE with training and testing sets to validate the model competence index. The SVR resulted in a 15.9% improved accuracy rate.

Kashikar et al. (2019) made an attempt on weather prediction using Scikit-Learn. They focused their research on the agricultural benefit of the prediction system. The dataset used for their research was obtained from Kaggle. Jaipur city was their focus for the research. Based on the test result they had, they concluded that the whole system performed according to the designed aim and objectives of the project. The weather prediction model system was able to forecast weather for next 7 days to 1 month with accuracy and efficiency.

Article DOI: 10.52589/AJMSS-F6H03BNE
DOI URL: https://doi.org/10.52589/AJMSS-F6H03BNE

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 162-171)

Munmun et al. (2018) carried out a weather forecast prediction research (an integrated approach for analyzing and measuring weather data). In their research, they built a system that predicts the future weather conditions based on the current weather data. The data mining techniques namely Chi square test and Naïve Base statistics are applied on the dataset to extract the useful information from the dataset. The transformed dataset is stored in a database that is collected from the user.

Abrahamsen et al. (2018) investigated the Machine Learning in Python for Weather Forecast based on Freely Available Weather Data. Weather data from frost.met.no was collected using a newly developed Python API. The data consists of hourly temperature and precipitation measurements in the period 01.01.2016 T00:00 to 31.12.2017 T23:00 from weather station SN30255 at latitude: 59.091 and longitude: 9.66 in Porsgrunn, Norway. These data were used to train and tube several Auto-Regressive Artificial Neural Networks (AR-ANN) by using TensorFlow from Python. 48 consecutive hours is set to be predicted by four different models.

METHODOLOGY

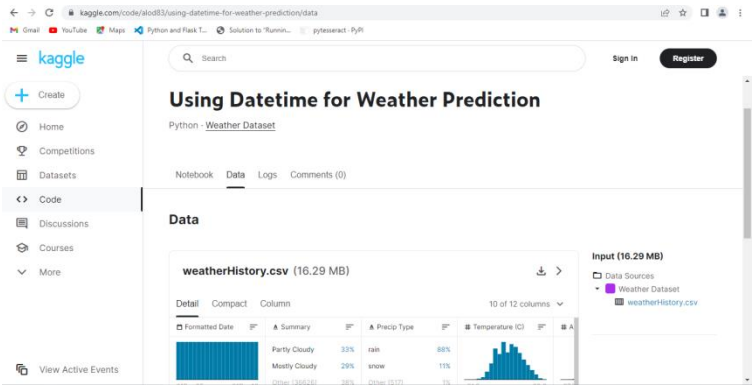
This is the process of acquiring the necessary tools for the development of the program. The project is developed with **Python** Programming Language on **jupyter notebook** in a **conda** environment which was dependent on some of the libraries such as **scikit- learn 0.22**, **pandas**, **matplotlib** which were used for training, testing, validation and evaluation of the machine learning models.

The data set of this research was obtained from Kaggle. This data is the weather history from 2006 to 2016. It contains a total amount of 96453 data entries and 12 columns. Some of the column headers are precip type, temperature, humidity, wind speed, pressure etc.

DATA COLLECTION

1. It consisted of various data points consisting of date of reading, precipitation, pressure, humidity, and many more.

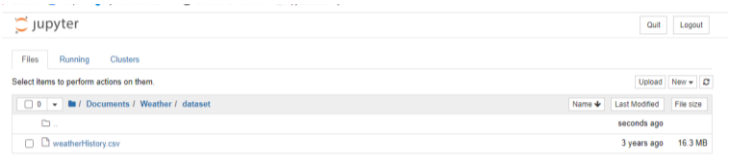
2. The dataset was imported into the notebook.



Dataset on kaggle

Article DOI: 10.52589/AJMSS-F6H03BNE
DOI URL: https://doi.org/10.52589/AJMSS-F6H03BNE

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 162-171)



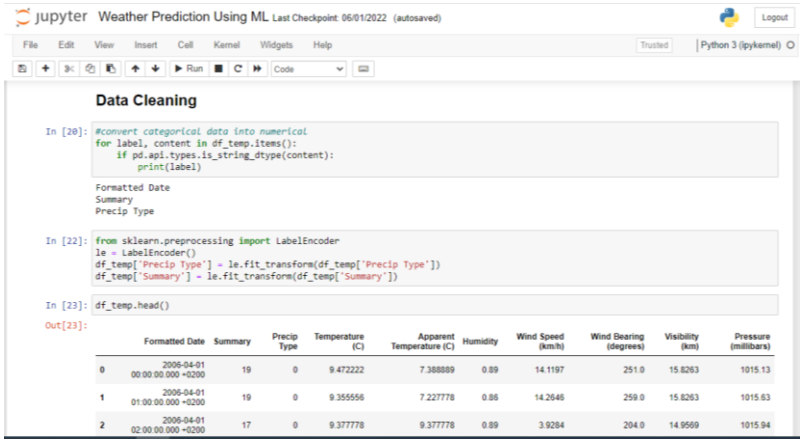
Dataset in Jupyter Notebook

```
#import the data
weather_data = pd.read_csv("dataset/weatherHistory.csv")
#view the data
weather_data.head()
```

Importing dataset

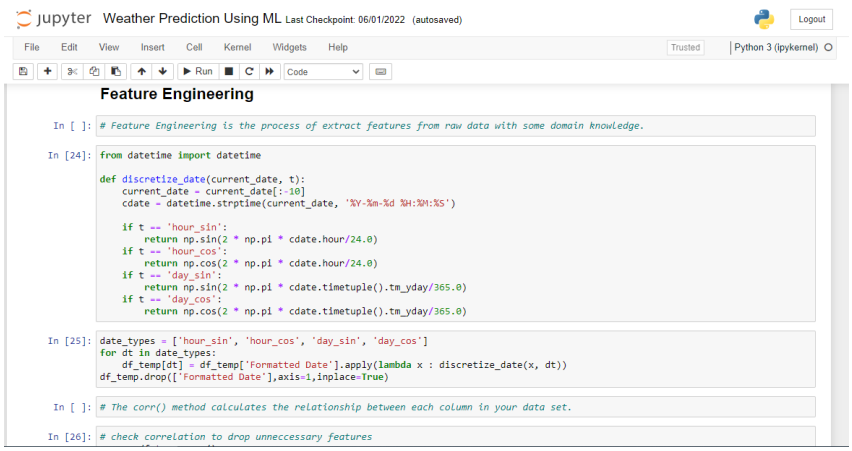
DATA PREPROCESSING

All the white spaces, irrational entries and false data was removed by variable python functions.



Data cleaning

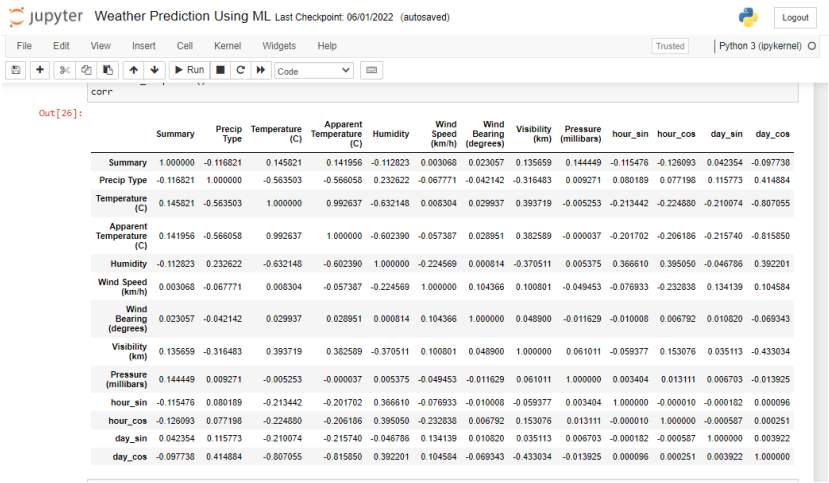
FEATURE ENGINEERING



Feature engineering

Article DOI: 10.52589/AJMSS-F6H03BNE
DOI URL: https://doi.org/10.52589/AJMSS-F6H03BNE

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 162-171)



jupyter Weather Prediction Using ML Last Checkpoint: 06/01/2022 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

```
corr
```

Out[26]:

	Summary	Precip Type	Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Pressure (millibars)	hour_sin	hour_cos	day_sin	day_cos
Summary	1.000000	-0.116821	0.145821	0.141956	-0.112823	0.003068	0.023057	0.135659	0.144449	-0.115476	-0.126093	0.042354	-0.097738
Precip Type	-0.116821	1.000000	-0.563503	-0.566058	0.232622	-0.067771	-0.042142	-0.316483	0.009271	0.080189	0.077198	0.115773	0.414884
Temperature (C)	0.145821	-0.563503	1.000000	0.992637	-0.632148	0.008304	0.029937	0.393719	-0.005253	-0.213442	-0.224880	-0.210074	-0.807055
Apparent Temperature (C)	0.141956	-0.566058	0.992637	1.000000	-0.602390	-0.057387	0.028951	0.382589	-0.000037	-0.201702	-0.206186	-0.215740	-0.815850
Humidity	-0.112823	0.232622	-0.632148	-0.602390	1.000000	-0.224569	0.000814	-0.370511	0.005375	0.366610	0.395050	-0.046786	0.392201
Wind Speed (km/h)	0.003068	-0.067771	0.008304	-0.057387	-0.224569	1.000000	0.104366	0.100801	-0.049453	-0.076933	-0.232838	0.134139	0.104584
Wind Bearing (degrees)	0.023057	-0.042142	0.029937	0.028951	0.000814	0.104366	1.000000	0.048900	-0.011629	-0.010008	0.006792	0.010820	-0.069343
Visibility (km)	0.135659	-0.316483	0.393719	0.382589	-0.370511	0.100801	0.048900	1.000000	0.061011	-0.059377	0.153076	0.035113	-0.433034
Pressure (millibars)	0.144449	0.009271	-0.005253	-0.000037	0.005375	-0.049453	-0.011629	0.061011	1.000000	0.003404	0.013111	0.006703	-0.013925
hour_sin	-0.115476	0.080189	-0.213442	-0.201702	0.366610	-0.076933	-0.010008	-0.059377	0.003404	1.000000	-0.000010	-0.000182	0.000096
hour_cos	-0.126093	0.077198	-0.224880	-0.206186	0.395050	-0.232838	0.006792	0.153076	0.013111	-0.000010	1.000000	-0.000587	0.000251
day_sin	0.042354	0.115773	-0.210074	-0.215740	-0.046786	0.134139	0.010820	0.035113	0.006703	-0.000182	-0.000587	1.000000	0.003922
day_cos	-0.097738	0.414884	-0.807055	-0.815850	0.392201	0.104584	-0.069343	-0.433034	-0.013925	0.000096	0.000251	0.003922	1.000000

Correlation output

jupyter Weather Prediction Using ML Last Checkpoint: 06/01/2022 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

In [27]:
```
#visualize the correlation using seaborn heatmap
plt.figure(figsize=(8,6))
sns.heatmap(corr,
            fmt=".2g",
            cmap="YlGnBu",
            annot=True)
```

Out[27]: <AxesSubplot:>

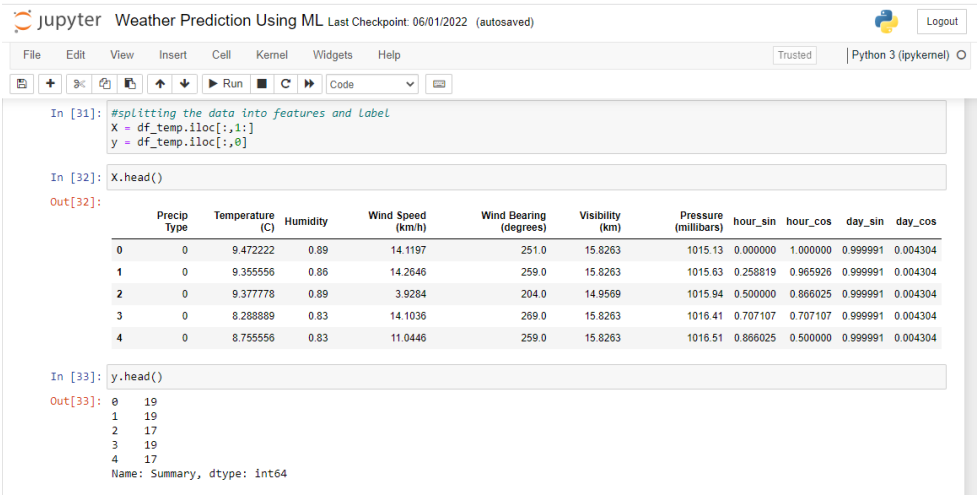
Visualize correlation output with seaborn heatmap

SPLITTING THE DATA INTO FEATURE AND LABEL

1. The data is split into features and label; x: feature, y: label

Article DOI: 10.52589/AJMSS-F6H03BNE
DOI URL: https://doi.org/10.52589/AJMSS-F6H03BNE

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 162-171)



Splitting data into feature and label

DIVIDING THE DATA INTO TRAINING AND TESTING

1. We used Scikit learn's function named train_test_split for this purpose.

2. It divides the dependent variables and the independent variables into 2 parts consisting of 80% training data and 20% testing data out of the total dataset.

Fig 4.2 Dividing the data into training and test data

ASSIGNING PRE-BUILT MODEL BY SCIKIT LEARN

1. Since we intend to only make predictions of the dataset test part, we are making use of classification algorithms named as Random Forest Classifier, Decision Tree Classifier and Extra Tree Classifier.

2. The Scikit learn's library must be imported before initiating the application process. So Scikit learn defines Random Forest Classifier, Decision Tree Classifier and Extra Tree Classifier.

3. The fit_transformfuction is used on the test and train input data to standardize it.

```
In [35]: #standardized the input data
         sc = StandardScaler()
         X_train  = sc.fit_transform(X_train)
         X_test = sc.fit_transform(X_test)
```

Standardize the input data

Article DOI: 10.52589/AJMSS-F6H03BNE
DOI URL: https://doi.org/10.52589/AJMSS-F6H03BNE

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 7, Issue 2, 2024 (pp. 162-171)

The fit_and_train function is use to train and test the models

```
In [36]: #create a dictionary and a function to evaluate the models
models = {"RFC": RandomForestClassifier(),
          "DTC": DecisionTreeClassifier(),
          "ETC": ExtraTreesClassifier()}

def fit_and_train(models, X_train, X_test, y_train, y_test):
    np.random.seed(42)
    model_scores = {}

    for name, model in models.items():
        model.fit(X_train, y_train)
        model_scores[name] = model.score(X_test, y_test)
    return model_scores
```

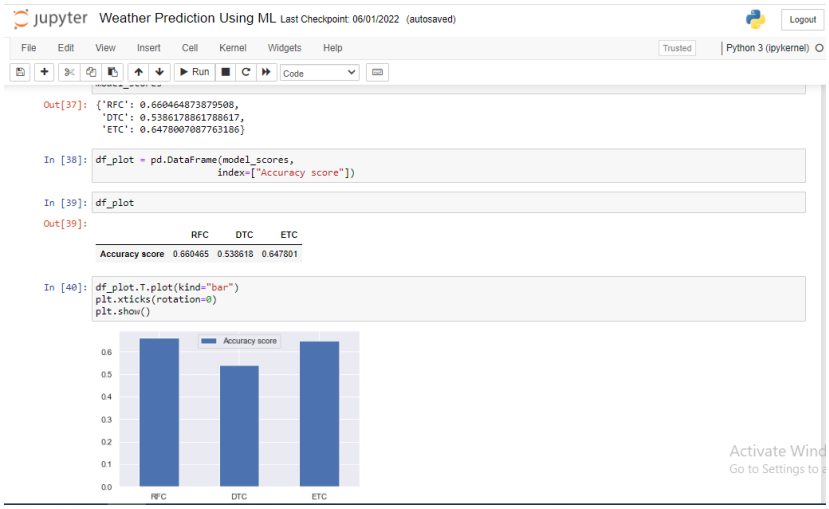
Testing and training the models

EVALUATING THE PERFORMANCE OF THE THREE CLASSIFIERS

After training the data by fitting into Random Forest, Decision Tree and Extra Tree we performed classification metrics from the Scikitlearn module by importing accuracy score, classification report and got the scores below respectively.

PLOTTING OF CHART USING MODELS' ACCURACY SCORE

After training and testing the models, the scores are saved in the variable model_scores, the accuracy score is then used to plot a bar chart showing the score of the three models.



Accuracy score chart

CONCLUSION

Weather prediction is an important research area in the field of machine learning, there are many things which are yet to be discovered and many new algorithms to be developed. In the Random Forest Classifier, we were able to achieve 66% accuracy, but there are some

Article DOI: 10.52589/AJMSS-F6H03BNE
DOI URL: https://doi.org/10.52589/AJMSS-F6H03BNE

limitations which are faced by our research as well. Hence, these issues are to be addressed by future works. We can also improve the accuracy, reliability, efficiency and speed. At the end of this research, we were able to implement Random Forest Classifier, Decision Tree Classifier and Extra Tree Classifier from Scikit-Learn to make weather prediction and using matplotlib to visualize the accuracy score of the implemented models.

REFERENCES

- Abyson Mattew (2022). Weather forecasting using the Random forest Algorithm Analysis Proceedings of the *National Conference on Emerging Computer Application* (NCEACA),*Vol-4, Issue-1, 2022*
- Attar-ur Rahman, Sagheer Abbas(2022). Rainfall Prediction System using Machine Learning Fusion for Small Cities. *Sensors 2022, 22, 3504* *https://doi.org/10.3390*
- Aravind Thilak S Vigneshwarar B, Dr.JB.Jona (2022) Weather Prediction using Random Forest Method. *International Journal of Creative Research Thoughts (IJCRT).Vol-10, Issue-2 february 2022*
- Bodgan Bochenek and Zbigniew Ustrnul (2022) Machine learning in weather prediction and climate analysis application and perspectives. *Atmosphere 2022, 13, 180, https://doi.org/10.3390/atmos13020180*
- Bilal Ahmed (2022).Will It Rain Tomorrow, Department of Computing and Information System.
- Bo xiang, Chunfenzeng Xining Dong and Jiaywe wang School of Geography and Tourism, Chongqing Normal University, Chongqing. The application of Decision Tree and Stochastic forest model in summer precipitation prediction in Chongqing
- E. B. Abrahamsen, O. M. Brastein, B. Lie (2018). Machine Learning in Python for Weather Forecast based on Freely Available Weather Data. *f The 59th Conference on Simulation and Modelling (SIMS) 59, September 2018:*https://doi.org/10.3384/ecp18153169
- Elia Georgiana Petre (2009). Decision Tree For Weather Prediction. *International Journal of Computer Applications (IJCA). Vol. LXI No. 1/2009*
- Garima Jain , Bhawna Mallick (2016). A Review on Weather Forecast Techniques. *International Journal of Advanced Research in Computer and Communication Engineering ISO 3297-2007.Vol-5, Issue-12, December 2016*
- G.Vamsi Krishna (2015). An Integrated Approach for Weather Forecasting based on Data Mining and Forecasting Analysis. *International Journal of Computer Applications (0975 – 8887), Vol.-120, No. 11, June 2015*
- Iseh. A. J., Woma. T. Y. (2013). Weather Forecasting Models, Methods and Applications. *International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 12, December – 2013*
- Ismaila Oshodi (2022). Machine Learning based Algorithm for Weather Forecasting. *doi:10.20944/preprints202206.0428.v1*
- Manepalli Janaki Naga Jyothi (2020). Weather Prediction Through Machine Learning. *Complexity International Journal (CIJ), Vol.-24, Issue-01, March 2020*
- Mohammad Abrar, Alex Tze Hiang Sim, Dilawar Shah, Shah Khusro, Abdusalam (2014). Weather Prediction Using Classification. *Sci.Int.(Lahore),26(5),2217-2223, Nov. – Dec., 2014*

Article DOI: 10.52589/AJMSS-F6H03BNE
DOI URL: https://doi.org/10.52589/AJMSS-F6H03BNE

- Ms. Ashyuni Mandale, Mrs Jadhawar B.A Assistant Professor, Dr. Daulatrao Aher (2015). Weather Forecast Prediction a Data mining Application. *International Journal of Engineering Research and General Science. Volume-3, Issue-2, March-April, 2015*
- Munmun Biswas, Tanni Dhoom, Sayantanu Barua (2018). Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data. *International Journal of Computer Applications (0975 – 8887), Vol.-182, No. 34, December 2018*
- Nalanda B Dudde,Dr. S.S.Apte (2014). Arbitrary Decision Tree For Weather Prediction. *International Journal of Science and Research(IJSR).ISSN(Online): 2319-7064. Index Copernicus Value(2013):6.14|Impact Factor(2014):5.611*
- N.DIYVA PRABHA , P.RADHA (2019). Prediction of Weather and Rainfall Forecasting using Classification Techniques. *International Research Journal of Engineering and Technology(IRJET). Volume-6, Issue-2 feb 2019*
- N.Priyanka (2021). Weather Prediction Using Deep Learning Technique. *Journal of Engineering Sciences(JES). Vol-12, Issue-5 MAY/2021*
- Pratyush Muthukumar, Emmanuel Cocom, Kabir Nagrecha, Dawn Comer, Irene Burga, Jeremy Taub, Chisato Fukuda Calvert, Jeanne Holm, Mohammad Pourhomayoun (2021). Predicting PM2.5 atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data. *Air Quality, Atmosphere and Health, Received-10, August 2021, Accepted-2, November 2021*
- Raitih Prasetya,Anggreani Ridwan (2019) Data Mining Application on Weather Prediction Using Classification Tree, Naives Bayes and K-Nearest Neighbor Algorithm with Model Testing of Supervised Learning Probabilistic Brier Score. *Journal of Applied Communication and Information Technologies(JAICT). Vol-4 No-2, 2019*
- Rajesh Kumar (2013). Decision Tree for Weather Forecasting. *International Journal of Computer Applications(IJCA). Volume-76 No-2, August 2013*
- R.Meenal, Prawin Angel Micheal *(2021*) . Weather Prediction using Random Forest Machine Learning Model . *Indonesian Journal of Electrical Engineering and Computer Science. Vol-22, No-2, May 2021, pp. 1208-1215*
- Rubhi gupta (*2020*). Review on Weather Prediction Using Machine Learning. *International Journal of Engineering Development and Research (IJEDR).Year 2020 Vol-8, Issue 1*
- Siddharth.S, Bhatkande ,Roopa G, Hubballi (*2016*). Weather Prediction Based on Decision Tree Algorithm using Data mining Techniques. *International Journal of Advanced Research in Computer and Communication Engineering(IJARCE). Vol-5, Issue-5 May 2016*
- S.Karthick, D.Malathi, C.Arun (*2018*). Weather Prediction Analysis using Random Forest. *International Journal of Pure and Applied Mathematics(IJPA). Vol-118, No-20, 2018, 255-262*
- Sudhnya Kashikar1, Sumedha Patil, Ameya Vedantwar, Shivani Katpatal, Sofia Pillai (*2019*). Weather Prediction using Scikit-Learn. *International Journal of Computer Sciences and Engineering, Vol.-7, Issue-4, April 2019*
- Tom Hamill (2019). Introduction to Numerical Weather Prediction and Ensemble Weather Forecasting. *NOAA-CIRES Climate Diagnostics Center Boulder, Colorado USA.*
- V.Sasikala (2020). Weather Predictive System using Machine Learning Algorithms. *Journal of Xi'an University of Architecture & Technology. Vol-XII, Issue-VI, 2020*

Article DOI: 10.52589/AJMSS-F6H03BNE
DOI URL: https://doi.org/10.52589/AJMSS-F6H03BNE

- Yousif Elfaith Yousif (2022). Weather Prediction System using KNN Classification Algorithm. *European Journal of Information Technology and Computer Science. http://dx.doi.org/10.24018/ejcompute.2022.2.1.44*

Article DOI: 10.52589/AJMSS-F6H03BNE
DOI URL: https://doi.org/10.52589/AJMSS-F6H03BNE