International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-9, Issue-3, May 2021
https://doi.org/10.21276/ijircst.2021.9.3.14
Article ID IRP1167, Pages 77-81
www.ijircst.org

Weather Prediction Using Machine Learning

Shivam Kumar Tandon[1], Pawan Kumar Singh[2], Abhishek Patel[3] and Dr. Shiv Kumar Verma[4]

[1,2,3,4] School of Computing Science and Engineering, Galgotias University, Greater Noida, Uttar Pardesh, India

Correspondence should be addressed to Shivam Kumar Tandon; shivam_tandon.scsemca@galgotiasuniversity.edu.in

ABSTRACT— Climate conducts a completely critical function in many key production sectors, e.g. farming. Climate change with high charging these days, which is why old weather forecasts are getting closer and less powerful and continue to be annoying. Miles are therefore very important to decorate and modify the weather forecast model. those predictions affect the country's financial system and people's lives. A system of information and statistics analysis algorithms has been used that includes a wooded area used for weather forecasting. The weather is one of the highest natural barriers in all parts of our lives in the world, we need to look at the weather including temperature, rain, humidity and other protection remarkable. The purpose of our artwork is to format effective weather forecasts. Earth's climate will change over a long period of time and also what kind of impact it will have on the lives of future generations. Our predictive nature of end-of-life climates offers an excellent desire to provide information as a way to allow stadium insurers to make an informed wish for the future of the world. Our approach greatly enhances the model in a positive way to govern the state of staff inconsistencies and inequalities and performs its function of accurately predicting the weather.

KEYWORDS— Digital Technology, Machine Learning, Weather, Data Preprocessing, Humidity, Rainfall

I. INTRODUCTION

Weather conditions change quickly and constantly all over the world. In today's world, accurate predictions are critical. We rely heavily on weather forecasts in everything we do, from agriculture to manufacturing, from travel to regular commuting. Since the entire planet is suffering from ongoing climate change and its consequences, it is critical to accurately forecast the weather in order to ensure day-to-day operations that are simple and smooth, as well as fast and seamless agility. Since weather systems can travel a long distance in all directions over time, the weather of one location can have a significant impact on the weather of others. In this paper, we propose a method for predicting weather conditions by combining historical weather data from nearby cities with data from a specific area[1]. We combine these data and use it to train simple machine learning models that can accurately forecast weather for the next few days. These simple models can be run on low-cost, low-resource computing systems while still providing fast and reliable forecasts that can be used in our daily lives.

The major contributions of this paper include:

1. The use of machine learning in the prediction of weather conditions in short periods of time, which can operate on less resource-intensive machines, is one of the paper's major contributions.
2. Automated systems to gather historical data from a dedicated weather service would be implemented.
3. A thorough assessment of the proposed methodology, as well as a comparison of various machine learning models for forecasting future weather conditions.

II. ARCHITECTURE OF MACHINE LEARNING

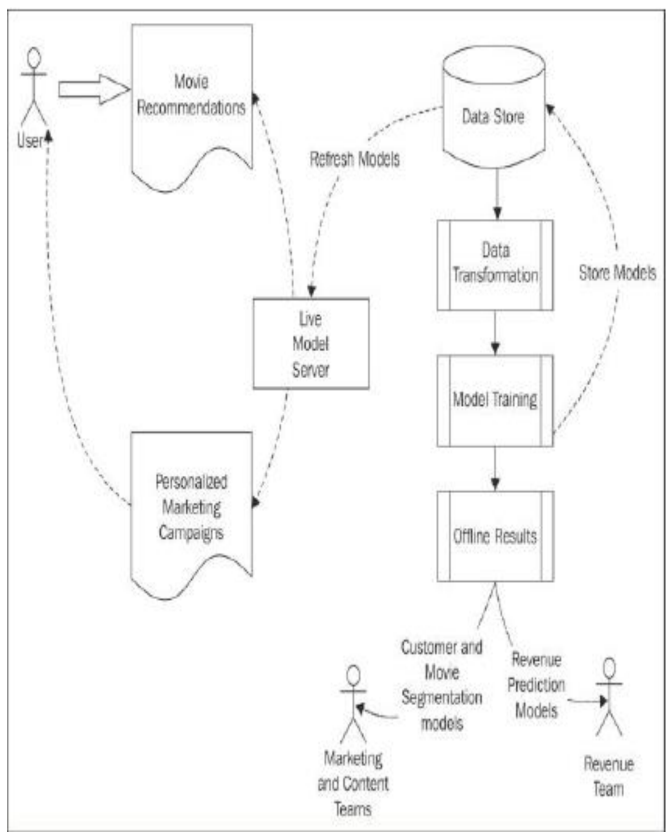

Fig. 1: Architecture of Machine Learning

The architecture of machine learning is elaborated with the aid of Figure 1.

As seen in the diagram,

1. We need to collect data on users, their actions, and content titles.
2. Transform the data into functionality after it has been collected.
3. Next, we'll use training and testing data to train our model. Model selection process includes training as well.
4. Model feeding back is the results of Movie Stream's website via targeting and recommendation pages.
5. Model feeding back is the results of Movie Stream's website via personalized marketing channels.
6. Using offline models to give Movie Stream's various teams tools to better understand user behavior.
7. The qualified model will then be deployed to a live-model serving system.

Machine Learning algorithms are divided into two categories. To construct models from raw data and perform regression and classification, supervised learning is used in this project.[4]

A. Supervised learning

Supervised learning is a method of learning in which the computer receives both the input and the desired output. For classification purposes, input and output data are labelled. A selection of predictors is used to predict the target variable. Random forest, KHH, Decision Tree, Regression, Logistic Regression, Boosting Algorithm, and other supervised learning algorithms include Random Forest, KHH, Decision Tree, Regression, Logistic Regression, and Boosting Algorithm.

Table 1: Supervised Learning Types and Usages

Types	Usage Example in business
Neural Network	Predicting Financial Result & Fraud Detection
Classification & regression	Spam filtering and fraud detection
Decision Tree	Risk Management, Threat management system

B. Unsupervised learning

Unsupervised learning is a form of learning that is used to make inferences from datasets that contain input data but no labelled responses. We don't have a boss to provide us with direction. Unsupervised learning algorithms include the Apriori algorithm, K-means, Adaptive Resonance Theory, and the Self-Organizing Map (SOM Model) [2].

Table 2: UnSupervised Learning Types and Usages

Types	Usage Example in business
Cluster analysis	Financial transactions, streaming analytics in IOT
Pattern recognition	Spam detection, Biometrics, Identify management
Association rule learning	Bioinformatics, manufacturing & assembly

III. LITERATURE SURVEY

Rapid growth in the fields of software, communication technology, and hardware, as we all know. This will make it easier for Internet-connected sensory devices to emerge. This will include physical world data measurements and observations. The total number of internet-connected devices in use by 2020 is expected to be between 25 and 50 billion. In comparison to the past, it is expected that internet-connected devices will rise and technologies will mature. The amount of data that is published will grow. Internet of Things (IoT) technology, which is used in Internet-connected devices, continues to expand the existing Internet by allowing interactions and communication between the physical and cyber worlds [1]. Furthermore, as the internet of things (IOT) becomes more accessible due to cheaper sensors and improved communication, the number of devices and pieces of equipment that can provide useful real-time weather information will likely grow dramatically[3]. In contrast to Weather Forecasting using Machine Learning Algorithms, which is mainly focused on simulation based on Physics and Differential Equations, Artificial Intelligence is often used for predicting weather, and involves models such as Neural Networks, Bayesian Networks, and Vector Machines. Because of its ability to capture non-linear dependencies of past weather patterns and future weather conditions, this algorithm is the only one used. And, in the coming years, more progress will be made in using these technologies to reliably forecast the weather in order to avoid disasters such as hurricanes, tornadoes, and thunderstorms [5].

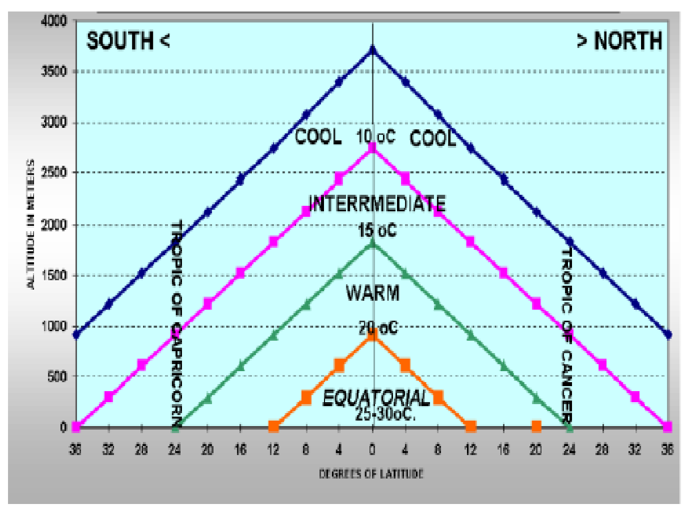

Fig. 2: Temperature Altitude/Latitude Relationship

Figure2 depicts the relationship between temperature and latitude. Low pressure and heavy rain are to blame for these. Kerala received strong monsoon rainfall on the evening of August 8, about 257 percent more than normal, causing dams to overflow; the state received 310 mm (12 in) of rain in the first 24 hours. As the water level has risen close to overflow due to heavy rainfall, flooding local low-lying areas, nearly all dams have been opened. 35 of the state's 54 dams have been opened for the first time in its history.

Table 3: Describes Maximum and Minimum Temperature

S No	Duration	Maximum Temperature	Minimum Temperature
1	June 2020	42.4 C	32 C
2	July 2020	35 C	27 C
3	August 2020	34 C	27 C
4	June 2019	33 C	23 C
5	July 2019	32 C	23 C
6	August 2019	32 C	24 C

IV. PROPOSED MODEL

Our model collects historical weather data that includes a variety of important factors that influence weather change, such as temperature, both maximum and minimum temperatures, atmospheric moisture or humidity, precipitation, the UV Index of the atmosphere, and atmospheric mean pressure.[6]

In our proposed model, the collected dataset is divided into sections that are useful to the machine learning model and parts that aren't. After that, the dataset goes through data preprocessing, which involves passing the data through a process that replaces missing and error values in the dataset with mean values or the most frequently occurring value in that field. Following the data preprocessing, the cleaned dataset is divided into two parts: the training set and the test set.

The training set is used to teach the machine learning model how to compute the results, while the testing set is used to find the results, compare the real and measured values, and use the error value as a benchmark to teach the machine learning model even more.

The training process will also include fold cross validation, in which the dataset is divided into k sets k times and then into test and training sets, with training sets chosen at random in each set and the model trained in each set. The test set for the qualified machine learning model is then the kth set.[8] This technique not only aids in the reduction of underfitting, but also the reduction of overfitting.

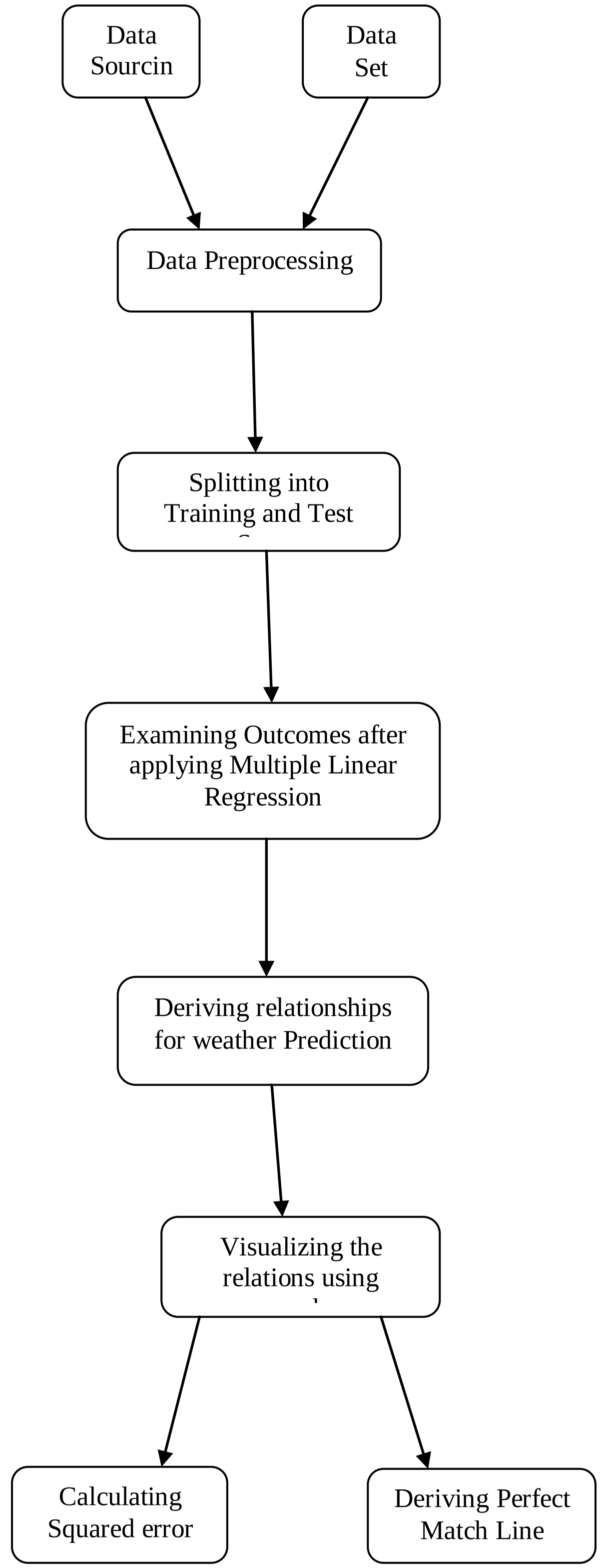

Fig. 3: Flow chart of proposed model

V. IMPLEMENTATION

A. Data Set

1. A website called https://rp5.ru/ obtained the Data Set we're using.
2. Depending on the location of the weather stations, it offers current as well as historical weather data going back many years.
3. The data set given by the source is fairly accurate and includes a variety of features to work with, such as mean temperature, maximum temperature, humidity, and precipitation.
4. The data set we're working with is for a particular area and is organized by date.

B. Tools Used

The University's authorized person who is responsible for the following activities is the Learner:

1. Spyder IDE: For data analysts and engineers, a powerful science environment written in Python.
2. GNU Octave: With built-in plotting and visualisation software, this is a powerful mathematics-oriented syntax.
3. Jupyter Notebook: Build and exchange documents with live code, calculations, visualizations, and narrative text using this open-source web application.

C. Libraries Used

The University's authorised person who is responsible for the following activities is the Examination Cell:

1. Sklearn: A machine learning library that aids in the development of machine learning models.
2. Matplotlib: Used to build intuitive graphs.
3. Pandas: Handle and import datasets.
4. Numpy: It is a library for mathematicians.
5. Future: Acclimating to incompatible updates that will be implemented in future launches.

Our proposed model is useful if we have unanticipated values to consider. We should pay attention to whether the value is very high or very low.

Disadvantage of this model, If we make a single extremely bad prediction, squaring may amplify the error and can skew the metric to overestimate the model's badness. That is an especially problematic behaviour when we have noisy data (data that is not completely accurate for any reason) — even a perfect model may have a high MSE in that case, making it difficult to judge how well the model is doing. On the other hand, if all of the errors are small, or even smaller than one, the opposite effect occurs: we can underestimate the model's poor performance [7].

VI. RESULTS

We discovered that the quadratic hypothesis cannot be used for this reason in the ML model used in this project. This is because a quadratic equation descends, which is something we don't want to happen. The hypothesis function should be cubic or higher degree. Furthermore, feature scaling is critical since the value increases as the degree of the polynomial increases.

VII. FUTURE WORKS

We discovered that feature scaling is an essential aspect of ML models during this project. The basic concept is to make sure that all of the functionality are on the same scale. We're just trying to get things moving faster here; the aim is to get all of the input variables into one of these ranges, give or take a few. In the coming years, we should strive to reduce variance as much as possible, as this will aid in better prediction, resulting in a successful ML model. Since outliers are bad for ML models, they should be avoided or omitted before determining the best match. This not only improves the model's precision, but it also keeps the findings consistent, which could be different if outliers were included. Professional weather forecasters aren't flawless, but they're usually more reliable than this linear regression model’s prediction. Weather is a non-linear system, according to this. Furthermore, unlike most forecasters, I based all of my predictions on data from a single location rather than several locations.

VIII. CONCLUSION

We introduced a technology that uses machine learning techniques to provide weather forecasts in this paper. Intelligent models can be created using machine learning technologies that are much simpler than conventional physical models. They use less resource and can be run on almost any machine, including mobile devices. In the future, we plan to collect weather data from various parts of a city using low-cost Internet of Things (IoT) devices, such as temperature and humidity sensors. The number of local features in the training dataset may be increased by using different sensors. Our prediction models will develop even further as a result of this data, which will be combined with data from weather stations.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1]. Arne Sund. 2015. Using Amazon Machine Learning to Predict the weather.

[2]. Scott C. James Yushan Zhang, Fearghal O’’Donncha, “A machine learning framework to forecast wave conditions”, Coastal Engineering Volume 137, july 2018, Pages 1-10.

[3]. WengianZhang, YomgviChen, “Mathematical Models of Multifactorial Decision and Weather Forecast”, IFAC Proceedings Volume 16, Issue 13, July 1983, Pages 265-269

[4]. Tanzila Saba, Amjad Rehman, Jarallah S. AlGhamdi – Weather forecasting based on hybrid neural model

[5]. Mark Holmstrom, Dylan Liu, Christopher Vo – Machine Learning Applied to Weather Forecasting

[6]. Jitcha Shivang, S.S Sridhar – Weather Prediction For Indian Location Using Machine Learning

[7]. Dixian Zhu, Changjie Cai, - A Machine Learning Approach for Air Quality

[8]. Sanjay D. Sawaitful, Prof. K.P. Wagh, Dr. P.N. Chatur – Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach

ABOUT THE AUTHORS

Shivam Kumar Tandon is a student of Master of Computer Application, School of Computing Science and Engineering Galgotias University Greater Noida , Uttar Pardesh , India

Pawan Kumar Singh, is a student of Master of Computer Application, School of Computing Science and Engineering Galgotias University Greater Noida , Uttar Pardesh , India

Abhishek Patel is a student of Master of Computer Application, School of Computing Science and Engineering Galgotias University Greater Noida , Uttar Pardesh , India

Dr. Shiv Kumar Verma, Program Chair (BCA, IOP, MMA) School of Computing Science and Engineering Galgotias University Greater Noida , Uttar Pardesh , India