# TEMPERATURE PREDICTION

30th Aug 2021-KAUSHAL KUMAR­ <https://www.linkedin.com/in/kaushal-kumar-5502b71a1>

https://github.com/kaush6464

Weather is an important aspect of a person’s life as it can help us to know when it’ll rain and when it’ll be sunny. Weather forecasting is the attempt by meteorologists to predict the weather conditions at some future time and the weather conditions that may be expected. The climatic condition parameters are based on the temperature, pressure, humidity, dewpoint, rainfall, precipitation, wind speed and size of dataset. Here, the parameters temperature, pressure, humidity, dewpoint, precipitation, rainfall is only considered for experimental analysis.

Firstly, the data is trained. For training the data, we will take 15-20% of the data from the data set. For this prediction, we’ll be using Linear regression algorithm and Naïve Bayesian classification algorithm. For the project, we’ll be using python, NumPy, Jupiter Notebook, Spyder, Panda. The project is split into three separate Jupiter Notebooks: one to collect the weather data, inspect it, and clean it; a second to further refine the features and fit the data to a Linear Regression model and Naïve Bayesian model and a third to train and evaluate our output.

**Background**

The application of science and technology that predicts the state of atmosphere at any given particular time period is known as Weather forecasting. There is a many different methods to weather forecast. Weather forecast notices are important because they can be used to prevent destruction of life and environment. The weather forecasting methods used in the ancient time usually implied pattern recognition i.e., they usually rely on observing patterns of events. For example, it is found that the following day has brought fair weather; if the preceding day sunset is particularly red. However, all of the predictions prove not to be reliable.

Weather forecasting is simply the prediction of future weather based on different parameters of the past like temperature, humidity, dew, wind speed and direction, precipitation, Haze and contents of air, Solar and terrestrial radiation etc. Weather forecast is an important factor affecting people’s lives. Once the data is taken, it is trained. The heart of this project is the Linear Regression algorithm which is used to predict the weather using these data. The more parameters considered, the higher the accuracy. This project can help many people finding the weather of tomorrow.

The project simply uses temperature, dew, pressure and humidity for training the data. Here these data are then trained using Linear Regression for the prediction.

**Literature Review**

There are many research papers that have been published related to predicting the weather. A paper was published on ‘The Weather Forecast Using Data Mining Research Based on Cloud Computing’ This paper proposes a modern method to develop a service-oriented architecture for the weather information systems which forecast weather using these data mining techniques. This can be carried out by using Artificial Neural Network and Decision tree Algorithms and meteorological data collected in Specific time. Algorithm has presented the best results to generate classification rules for the mean weather variables. The results showed that these data mining techniques can be enough for weather forecasting. Another paper was published on ‘Analysis on The Weather Forecasting and Techniques’ where they decided that artificial neural network and concept of fuzzy logic provides a best solution and prediction comparatively. They decided to take temperature, humidity, pressure, wind and various other attributes into consideration.

Another research paper titled ‘Issues with weather prediction’ discussed the major problems with weather prediction. Even the simplest weather prediction is not perfect. The one-day forecast typically falls within two degrees of the actual temperature. Although this accuracy isn’t bad, as predictions are made for further in time. For example, in a place like New England where temperatures have a great variance the temperature prediction are more inaccurate than a place like the tropics. Another research paper titled ‘Current weather prediction’ used numerical methods to stimulate what is most likely going to happen based on known state of the atmosphere. For example, if a forecaster is looking at three different numerical models, and two model predict that a storm is going to hit a certain place, the forecaster would most likely predict that the storm is going to hit the area. These numerical models work well and are being tweaked all the time, but they still have errors because some of the equations used by the models aren’t precise.

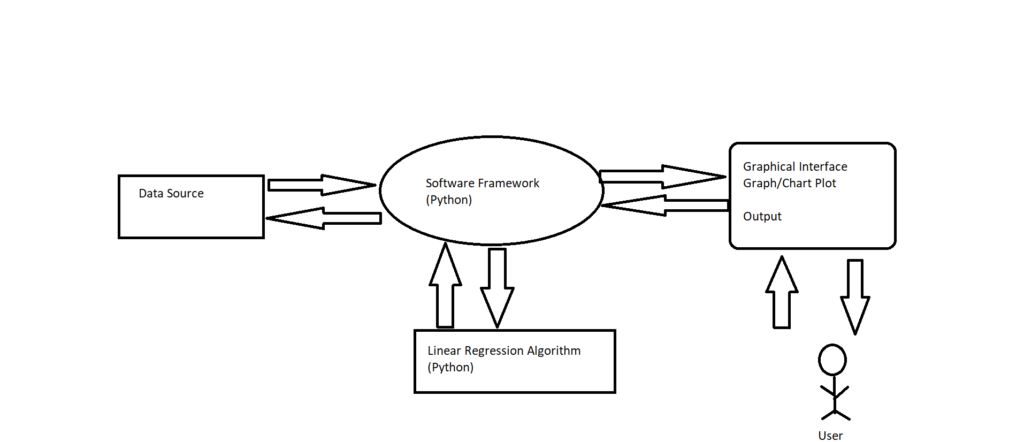
**Software Requirement**

The software used in our projects are:

* Python 3.7: Python is an interpreted, high level, general programming language. Its formatting is visually uncluttered, and it often uses English keywords where other languages use punctuation. It provides a vast library for data mining and predictions.
* Jupiter Notebook/ Spider/ Pycharm: It is an [open source](https://en.wikipedia.org/wiki/Open-source_software) cross-platform [integrated development environment](https://en.wikipedia.org/wiki/Integrated_development_environment) (IDE) for scientific programming in the [Python language](https://en.wikipedia.org/wiki/Python_(programming_language)). Spyder integrates with a number of prominent packages as well as another open source software.
* Numpy: Numpy was used for building the front-end part of the system.
* Pandas: Pandas was used for the data preprocessing and statistical analysis of data.
* Matplotlib: Matplotlib was used for the graphical representation of our prediction.

**Functional Requirements**

* The system must provide the predicted weather.
* The system must have an easy to use interface for using the system for all the users.
* The Admin must be able to update/modify the Dataset.
* The Dataset of the weather must be available for the system.

Block Diagram

**DATA COLLECTION**

The data of weather forecast was obtained from Kaggle. We took about 4000 trained data and 800 test data. Parameters are :-

* Temperature
* Pressure
* Humidity
* Dewpoint
* Rainfall
* Precipitation

**The steps involved in preprocessing are: –**

* **Features selection**

The data we have collected has many unwanted attributes which will not be needed in our project. Hence, we use the attributes which we need only.

* **Normalization**

The data we collected from internet should be first normalized. Normalization refers to rescaling real valued numeric attributes into the rage or 0 and 1. After the data are filtered it is then normalized.

* **Machine Learning**

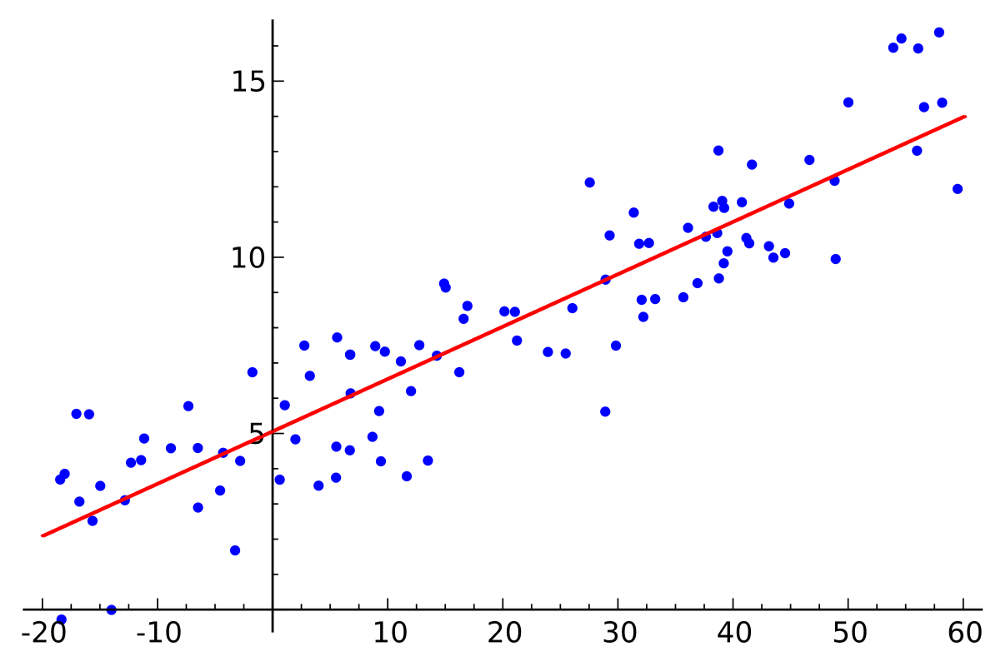
Training a model is the process of iteratively improving your prediction equation by looping through the dataset multiple times, each time updating the weight and bias values in the direction indicated by the slope of the cost function (gradient). Training is complete when we reach an acceptable error threshold, or when subsequent training iterations fail to reduce our cost.

**Implementation of Algorithms**

The algorithm used in our project are Linear Regression and Naïve Bayes Algorithm.

**1. Linear Regression**

Regression is a method of modelling a target value based on independent predictors. This method is mostly used for forecasting and finding out cause and effect relationship between variables. Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables.



**2. Naïve Bayes Algorithm**

Naïve Bayes Algorithm is a probabilistic machine learning algorithm which can be widely used in various classification tasks which is based on Bayes Theorem. The term naïve is given because it assumes the data that is given to the model are independent of each other, that is they have independent distribution. So, if we change the value of one feature than it doesn’t affect the value of other features used in the algorithm.

There are many applications of Naïve Bayes Algorithm like real time prediction, multi class prediction, text classification, spam filtering, recommendation system etc.

However, the algorithm is getting its popularity because of its robustness ability to noise and outliers as well as to irrelevant attributes. The missing values are easily handled. The predictions are made real-quick because of which, it is easily scalable

**Result**

The prediction system works properly. The values of all the attributes were properly preprocessed. After all the preprocessing was completed, model was implemented and it was trained using train data. The GUI for the system was made with tkinter. The coding was done in Pycharm/Spyder. After completing all the process, we connected the front-end with the back-end.

Our accuracy was found to be around 82%.

**Conclusion**

The weather prediction done using linear regression algorithm and Naïve Bayes algorithm are very essential for improving the future performance for the people. For predicting the weather, the linear regression algorithm and Naïve Bayes algorithm was applied to the datasets of the weather. We made a model to predict the weather using some selected input variables collected from Kaggle. The problem with current weather scenario is that we are not able to prepare our self and not able to do some important works. So, for knowing the weather scenario at high accuracy considering every factor that affects in the weather scenario, this model is created.