Weather Forecasting using Machine Learning Techniques

Siddharth Singh(sid912005@gmail.com)[*], Mayank Kaushik(mayank7533@outlook.com)[*], Ambuj Gupta(ambujgupta007@gmail.com)[*], Anil Kumar Malviya(anilkmalviya@yahoo.com)[**]
[*]B. Tech Students, CSED, KNIT, Sultanpur
[**]Professor and Head, CSED, KNIT, Sultanpur

Abstract
This paper explores three machine learning models for weather prediction namely Support Vector Machine(SVM), Artificial Neural Network(ANN) and a Time Series based Recurrent Neural Network(RNN). It also discussed the steps followed to achieve results. RNN using time series along with a linear SVC and a five-layered neural network is used to predict the weather. The results of these models are analyzed and compared on the basis of Root Mean Squared Error between the predicted and actual values. For weather Forecasting, this paper uses Pandas, NumPy, Keras, Git, Matplotlib, TensorFlow, Anaconda and Google Cloud Services[4]. It is found that Time Series based RNN does the best job of predicting the weather.

Keywords
Machine learning, weather forecasting, artificial neural network, time series analysis using recurrent neural network, support vector machine. Time series analysis.

1. Introduction
Weather forecasting is the application of scientific techniques and technology to predict the conditions of the atmosphere at a certain location and time. Weather Forecasting in old time is carried out by hand, using changes in barometric pressure, current weather conditions, and sky condition or cloud cover, weather forecasting now relies on computer-based models that take many atmospheric factors into accounting now relies on computer-based models that take many atmospheric factors into account.

For a long time, the researcher had attempted to establish a linear relationship between the input weather data attributes and the corresponding target attribute. But the discovery of nonlinearity within different attributes of weather data, the focus has shifted towards the nonlinear prediction of the weather. Weather forecasts are made by collecting quantitative data about the current state and previous trend of the atmosphere and using scientific understanding of atmospheric processes to predict how the atmosphere will evolve. The weather warning is important for the protection of life and property. Rain predictions can be used by farmers.

In order to analyze how the different machine learning techniques will perform in the forecasting of weather. We have trained different types of machine learning models on data collected from the airport weather station of several cities.

2. Previous Work
Weather prediction has seen a variety of approaches in recent years based on Genetic Algorithms and Neural networks but these fail to capture the complex relationships between various factors which affect weather. In the neural network approach, Kaur[2] and Maqsood[1] describes a model that predicts the hourly temperature, wind speed and relative humidity 24 hours ahead. The authors have made a comparison of Multilayer Perceptron Networks (MLP), Radial Basis Function Network (RBFN).

All these models do a good job in identifying the seasonal variations but fail in trend and random variations. A variety of time series models have also been developed[6] like Autoregression, Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA) etc. but combining time series with the neural network[7] is mostly unexplored. We found that RNN with time series performs better than the classical time series models.

3. Methodology

The following steps are used to achieve the objectives of this paper-

1. Setup
2. Data collection
3. Data preprocessing
4. Training models

Figure 1 is a flowchart of the steps followed during the research.

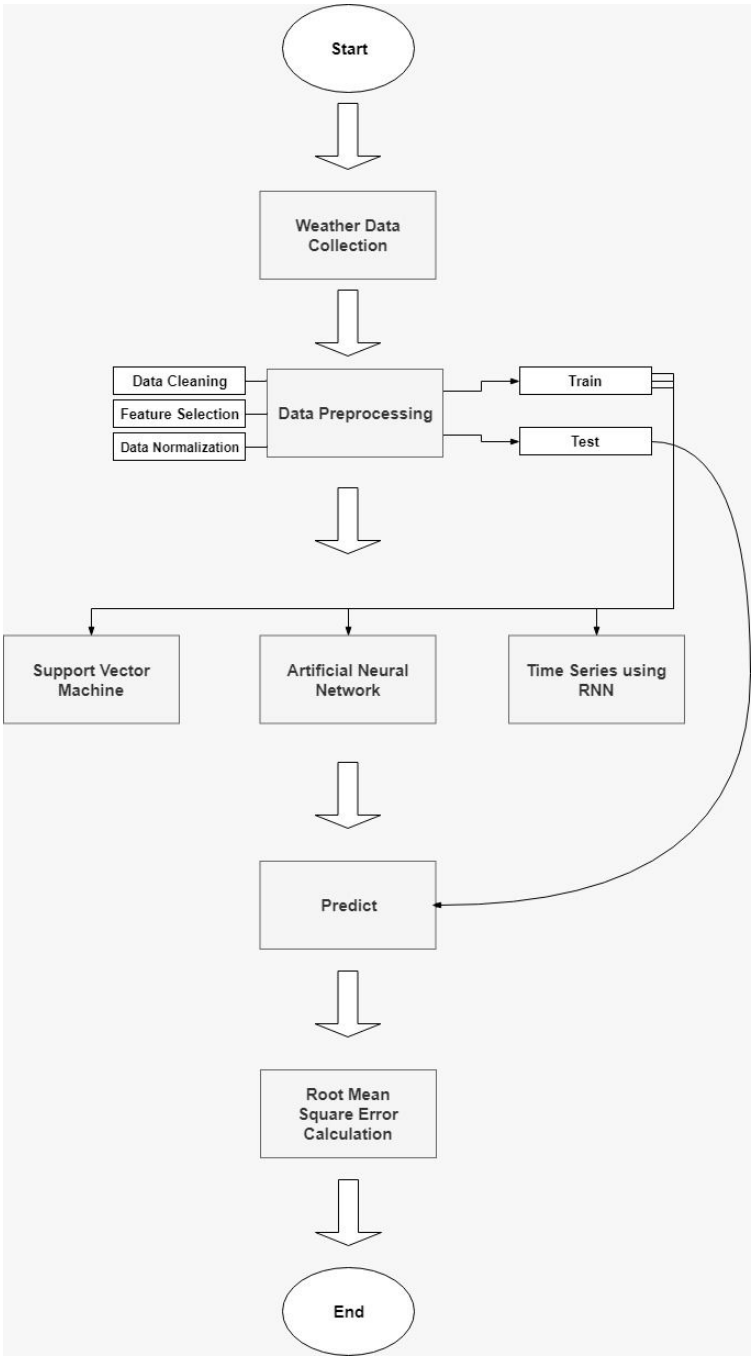

Figure 1: Flow chart of research methodology

3.1 Setup

Complete data analysis and preprocessing in Anaconda's Python 3.6 environment. Using libraries like Pandas, Sklearn, NumPy, Matplotlib.For training purposes, we used Python 3.6 environment provided by Google colaboratory[4]. We used libraries like TensorFlow, Keras.

3.2 Data collection

In order to analyze different machine learning techniques. We have collected data[3] from different airport weather station of India. This datasets include several attributes Air temperature at 2 metre height above the earth's surface, Atmospheric pressure at weather station level, Atmospheric pressure reduced to mean sea level, Relative humidity (%) at a height of 2 metres above the earth's surface, Mean wind direction at a height of 10-12 metres above the earth's surface, Total cloud cover, Horizontal visibility, Dew point temperature at a height of 2 metres above the earth's surface. This dataset contains data of previous 12 years 2006-2018. Data for some days are missing in this dataset but this can be compensated by the large size of the dataset

3.3 Data preprocessing

Data preprocessing is a data mining technique that involves transforming raw data into a usable format. The dataset contains some missing values. In order to train the model on this dataset, we used the following preprocessing techniques.

3.3.1 Data Cleaning:

No real-world dataset is complete so we have to fill empty values and columns, texts need to mapped with numbers before feeding dataset to different models. First, we assign different numbers to different wind direction and map it to wind directions columns. All columns have numerical values

For columns dew point and rain we fill not available rows with 0 representing no dew and rain respectively. We then used linear interpolation to fill other columns

3.3.2 Feature Selection:

The relationship between various columns of the weather data is considered and their covariance matrix is generated showing linear dependence between different columns. Columns with high linear dependence do not provide important information in prediction thus removed from the dataset

3.3.3 Data Normalization

The goal of normalization is to change the values of numeric columns in the dataset to a common scale, without affecting differences in the ranges of values or losing any information. Normalization help in faster training of models. We normalized the dataset except for the temperature in 0-1 range.

3.4 Training models

We used different machine learning techniques for weather forecasting. For analysis purpose we only predicted temperature but the same algorithms can be used for predicting other attributes of weather like precipitation, pressure, snowfall etc

We used the following methods for prediction:-

3.4.1 Support Vector Machine

For this method, we only used time, rain, snow and humidity as input features and temperature as output feature.

We used a linear SVC to predict the temperature. Although SVC is suitable for classification it can be used for regression.

3.4.2 Artificial Neural Network

A typical feedforward with back propagation network should have at least three layers- an input layer, a hidden layer, and an output layer. We have selected the appropriate number of hidden layers and the number of neurons in each of them as shown in table 1.

Layer	No.of Neuron	Activation
1	16	relu
2	32	relu
3	16	relu
4	5	relu
5	1	linear

Table 1: Details of each layer in ANN

We used Adam optimizer for faster training and trained for 100 iterations using batch size 32, It took us near 300s for all iterations. For regression, the loss used is generally root mean square error so we used the same

Figure 2 depicts loss with each iteration.

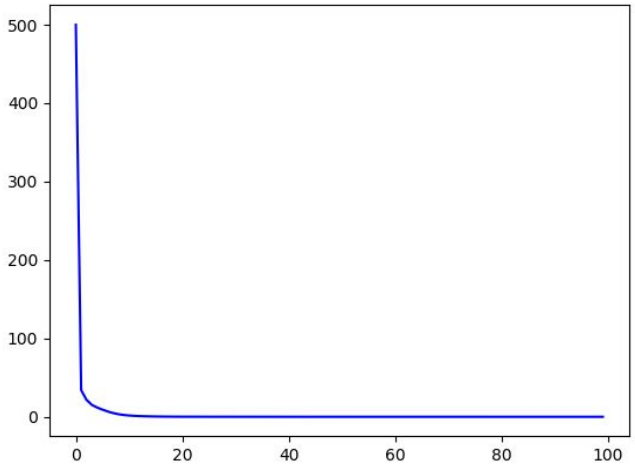

Figure 2: Decrease in the loss with iteration

So we can see around iteration 15, 18 so we do not need that much iteration it can be reduced to 20.

3.4.3 Time series using RNN:

We first formatted the data to make it a time series, which involved indexing the data based on timestamp. Then we normalized the data so that we can use the sigmoid activation function. We used a batch generator that gave the model input with batch size 256 for training the model.

Then we used the Keras Sequential model with Sigmoid activation function to predict the temperature, pressure, and humidity. We ran this model for 20 epochs with 100 steps in each epoch.

We then used the model to predict the temperature, pressure, and humidity for the period of eight weeks.

4 Tools Used

- Pandas - Open source data structuring and analysis tool
- Keras - An open source neural network library in Python.
- NumPy - Used for scientific computing in Python.
- Git - Version control system
- Matplotlib - Python library for visualization of data
- TensorFlow - Python library used for dataflow programming.
- Anaconda - Free and open source distribution of Python for scientific computing.

- Google Cloud Services - Provides free GPU accelerated platform for machine learning

5 Results

After analysis of all the models it is observed that time series RNN is better than SVM and ANN for this problem, Also the prediction window is very crucial as large prediction windows result in a high error. Table 2 depicts is the comparison between the three models.

Model	Prediction Window	RMS error
SVM	8 weeks	6.67
ANN	8 Weeks	3.1
Time Series RNN	8 Weeks	1.41

Table 2: comparison of model

Figure 3 shows temperature on the y-axis and sequence no of test data on the x-axis, the temperature predicted by SVM is represented by blue color and the actual temperature is represented by orange

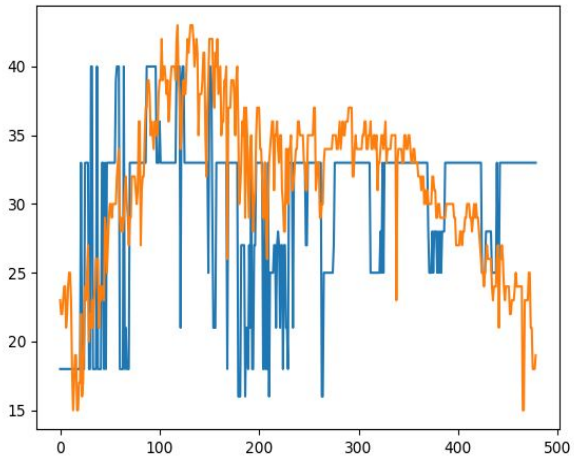

Fig 3: actual vs predicted temperature by SVM

Figure 4 shows temperature on y-axis and sequence no of test data on the x-axis, the temperature predicted by ANN is represented by red color and the actual temperature is represented by blue color

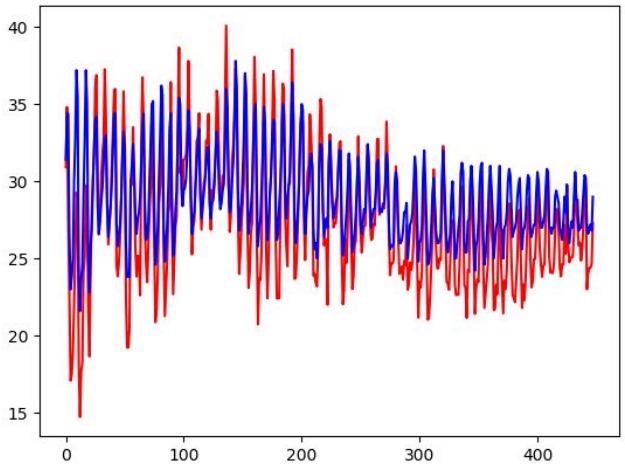

Fig 4: actual vs predicted temperature by ANN

Figure 5 shows temperature on y-axis and sequence no of test data the temperature predicted by RNN is represented by green color and the actual temperature is represented by blue.

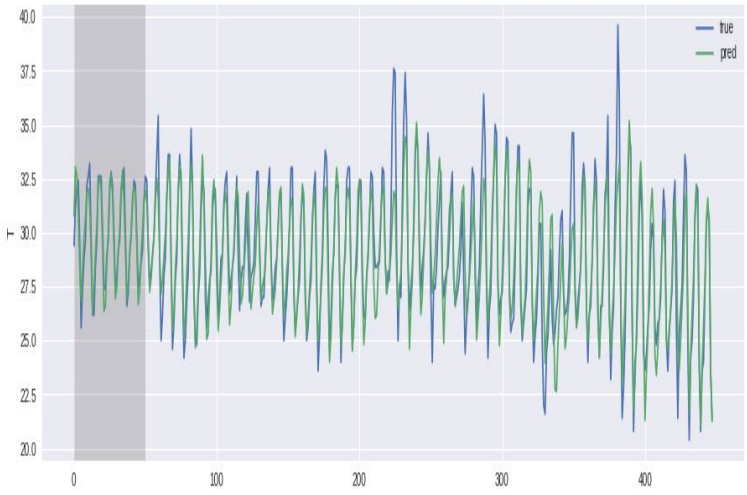

Fig 5: actual vs predicted temperature by RNN

6 Conclusion

In this paper, weather data is considered with different attributes for weather forecasting. The weather forecasting experiment was carried out to analyze the

performance of different machine learning techniques. We trained three different models on this data SVM, ANN
and time series RNN. We then used these models to predict weather and calculated root mean square error from the actual temperature. From observation of this project, we
found out that time series using RNN is a better method for weather forecasting.

7 Acknowledgment

We acknowledge with gratitude to our HOD Dr. Anil Kumar Malviya, for his constant encouragement, patience and valuable supervision at every stage of our work. This work would not have been materialized at the present from without his incisive observation and intellectual direction in the course of completion

We also acknowledge the different books, different websites and a research paper which help us to get acquainted with new topic and prerequisites for writing this paper.

At last, we would like to thank google colab for providing the hardware which made this research possible

8 References

[1] Imran Maqsood, Muhammad Riaz Khan, and Ajith Abraham, "An ensemble of neural network for weather forecasting", Neural Comput & Applic (2004) 13: 112–122

[2] Amanpreet Kaur, J K 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/ a Russian website with weather data 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] G.Vamsi Krishna, "An integrated approach for weather forecasting based on data mining and forecasting analysis".International Journal of Computer Applications (0975 – 8887) Volume 120 – No.11, June 2015

[6] G.Vamsi Krishna, "An integrated approach for weather forecasting based on data mining and forecasting analysis". International Journal of Computer Applications (0975 – 8887) Volume 120 – No.11, June 2015

[7] Neeraj Kumar, Govind Kumar Jha, "A time series ANN approach for weather forecasting". International Journal of Control Theory and Computer Modeling (IJCTCM) Vol.3, No.1, January 2013