Proceedings of IOE Graduate Conference, 2019-Winter
Peer Reviewed
Year: 2019 Month: December Volume: 7
ISSN: 2350-8914 (Online), 2350-8906 (Print)

Weather Forecasting Application using Linear and Logistic Regression Algorithm

Astha Sharma

Himalaya College of Engineering, Tribhuvan University, Institute of Engineering, Chyasal, Lalitpur, Nepal
Corresponding Email: asthasharmabhattarai@gmail.com

Abstract
Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location. This system will predict weather based on parameter such as temperature, humidity, wind speed, wind direction, pressure, precipitation and the probability of rainfall. This system is a web application with effective graphical user interface. The algorithms that have been used are multi-linear regression and logistic regression for the forecast. The final system forecasts average minimum and maximum temperature, windspeed and probability of rainfall in Kathmandu city for 5 upcoming days. This system can be used in sectors like air traffic, marine, agriculture, forestry, military, navy, individual, etc.

1. Introduction

Weather forecasting is the application of current technology and science to predict the state of the atmosphere for a future time and a given location. Different algorithms like Neural Network, Sliding Window, Bayesian, Decision Tree, etc. have been used to develop such forecasting applications in today's era. Most of the existing applications use data from the satellites, sensors and weather stations made for capturing weather data. Then, they analyze and predict weather with the help of computer models.

But in this project looking at the most affordable means of data collection and prediction, past 23 years daily weather data have been collected from Department of Meteorology and Hydrology, Naxal. Linear and Logistic regression algorithms have been used to predict the weather on the basis of minimum and maximum temperatures, wind speed, humidity, precipitation and chances of rainfall.

2. Objective

The general objective of our project is to develop a web based weather forecasting application that forecasts 5 days weather for Kathmandu city in Python programming language.

3. Methodology

3.1 System architecture

As the project encompasses different technical fields and requires specific handling for each subsection, the project is divided into six different sections.

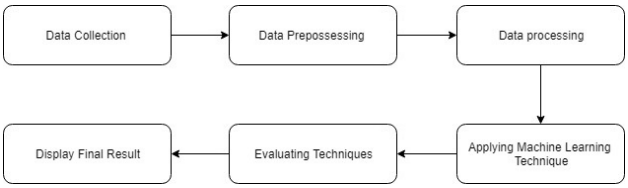

Figure 1: System Architecture

3.2 System Algorithm

3.2.1 Linear Regression

Linear Regression is a machine learning algorithm used for the prediction of parameter which is in continuous nature. In this project, linear regression has been used for forecasting the minimum and maximum temperature and wind speed.

Major objectives of Linear Regression Function:

Linear regression has been used for the following two objectives:

- In order to find the relationship among variables (here maximum temperature rainfall and minimum temperature, etc.).

- In order to forecast or estimate the values of some variables so that new observations are entertained.

3.2.2 Logistic Regression

Logistic regression is used when the output are in categorical form. In this project, logistic regression has been used for forecasting the probability of rainfall which in turn decides whether it will rain or not.

Major objective of Logistic Function:

For this project, logistic regression is being for the categorizing the data and to predict the probability of rainfall.

Major formulae used in the project:

The fundamental equation of generalized linear model is:

$$g(E(y)) = \theta_0 + x_1\theta_1 + x_2\theta_2 + x_3\theta_3 + x_4\theta_4 + x_5\theta_5 + \cdots \quad (1)$$

Where, $g()$ is the link function,

$E(y)$ is the expectation of target variable and is the linear predictor i.e. minimum and maximum temperature, humidity, precipitation and wind speed.

The role of link function is to link the expectation of y to linear predictor.

The cost function is use to predict the optimum value of $\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5$.

For those values cost function has minimum value and the predicted line is best fit.

Logistic regression is only concerned about the probability of outcome dependent variable (success or failure). As described above, $g()$ is the link function. This function is established using two things: Probability of Success (p) and Probability of Failure ($1-p$).

To make the probability less than 1, we must divide $g(E(y))$ by a number greater than $g(E(y))$ This can simply be done by:

$$L(x) = \frac{\exp(g(E(y)))}{1+\exp(g(E(y)))} \quad (2)$$

Using the above formula L(x) can be written as:

$$L(x) = \frac{e^y}{e^1 + e^y} \quad (3)$$

where,

$g(E(y)) = \theta_0 + x_1\theta_1 + x_2\theta_2 + x_3\theta_3 + x_4\theta_4 + x_5\theta_5$

$L(x)$ is the probability of success of rainfall.

This (3) is the Logit Function

If $L(x)$is the probability of success, $1-L(x)$ will be the probability of failure which can be written as:

$$q = 1 - L(x) = 1 - \frac{e^y}{e^1 + e^y} \quad (4)$$

Where q is the probability of failure

Let $L(x) = p$

On dividing, Equation 3 by 4, we get:

$$\frac{p}{1-p} = e^y \quad (5)$$

After taking log on both side, we get

$$\log\left(\frac{p}{1-p}\right) = y \quad (6)$$

$\log\left(\frac{p}{1-p}\right)$ is the link function. Logarithmic transformation on the outcome variable allows us to model a non-linear association in a linear way.

After substituting value of y, we'll get:

$$\log\left(\frac{p}{1-p}\right) = \theta_0 + x_1\theta_1 + x_2\theta_2 + x_3\theta_3 + x_4\theta_4 + x_5\theta_5 \quad (7)$$

where $x_1, x_2, x_3, x_4,$ and x_5 are different variables.

4. Results and Discussion

The web based weather forecasting application forecasts the average maximum and minimum temperature, wind speed and probability of rainfall in Kathmandu city for the upcoming 5 days. The expected min-max temperature must be within range of the -3°C and 40°C. The outputs are obtained in the expected range and are 94% accurate when compared with the forecast by Google. Similarly, wind speed is forecasted in the range of 1 km/hr to 10 km/hr. The probability of rainfall is obtained on the range of 0 to 100% which is resulted in 90% accurate forecasts.

One of the reasons why this system could not predict 100% accurately is because of the effect of global warming. The weather seems to fluctuate and not follow the exact trend of previous years. Furthermore, the data collected needed thorough preprocessing which also may be another reason for deducting the accuracy. At last, it is a fact that in machine learning, the more and better is the data, the perfect is the result. Since the total data collected were from only last 23 years, the algorithm could not forecast the weather with 100% accuracy.

4.1 Result Analysis using Chart

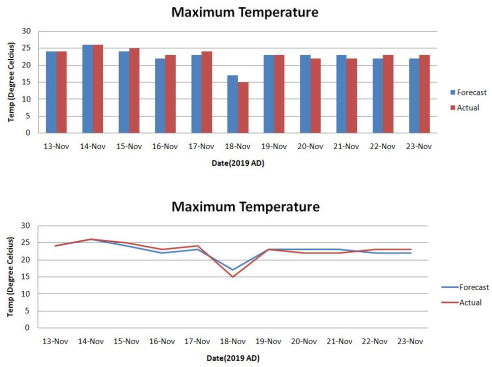

Figure 2: Bar diagram and line chart of forecast and actual maximum temperatures in Kathmandu city

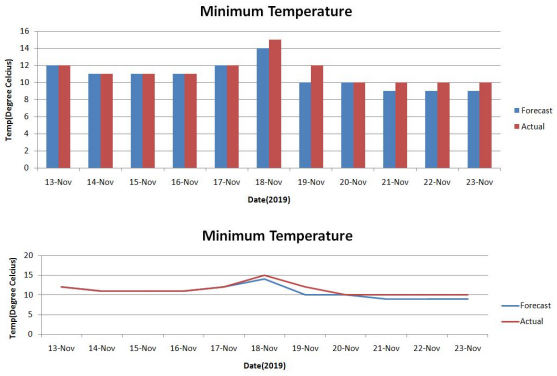

Figure 3: Bar diagram and line chart of forecast and actual minimum temperature in Kathmandu city

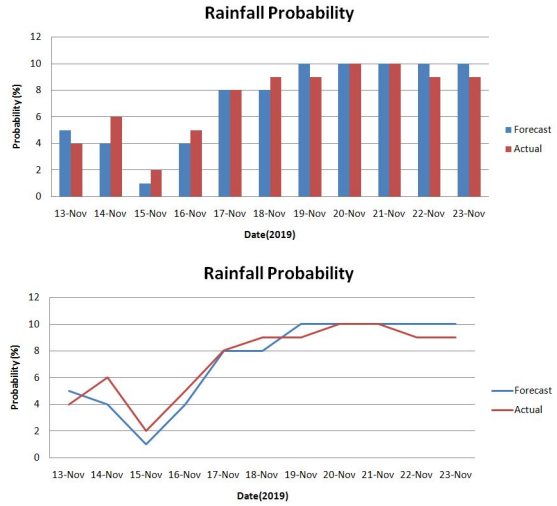

Figure 4: bar diagram and line chart of forecast and actual rainfall probability in Kathmandu city

Figure 5: Bar diagram and line chart of forecast and actual wind speed in Kathmandu city

4.2 Synapsis of Results

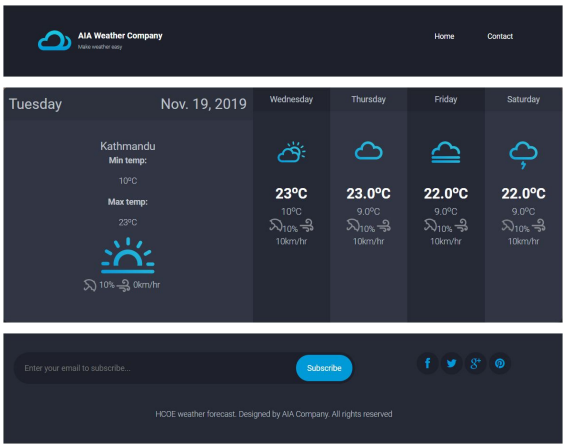

Figure 6: Pictures of the Home page with header, forecast and footer

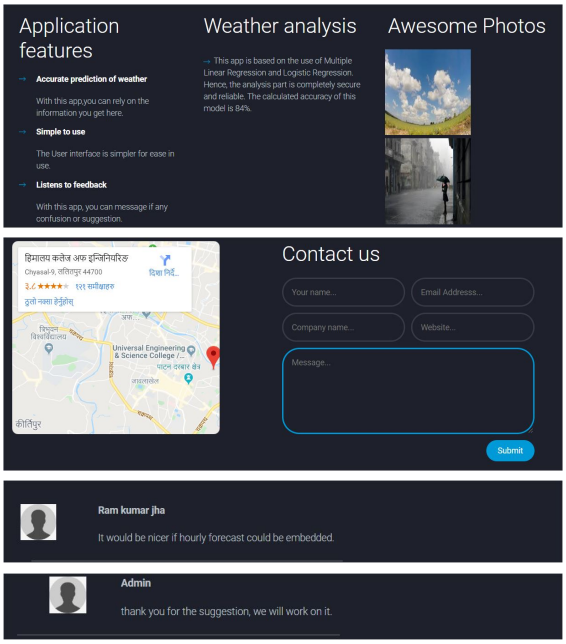

Figure 7: Pictures of the features including feedback system, contact page and weather related photos

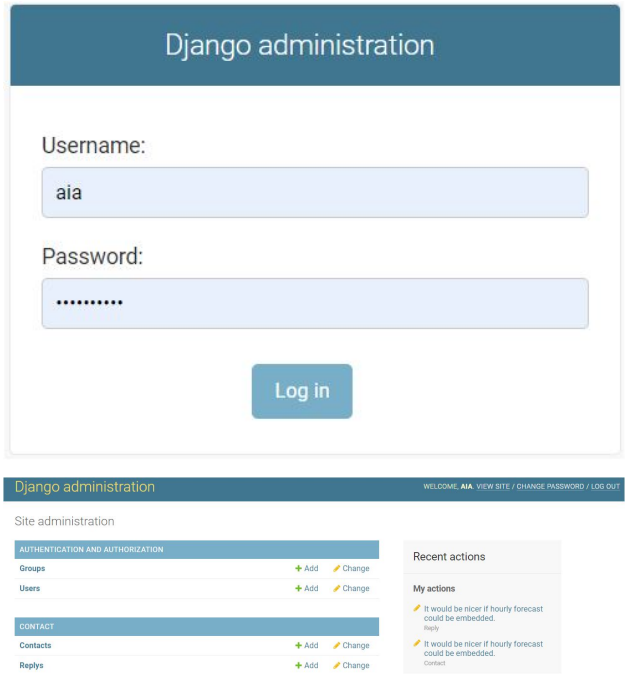

Figure 8: Screenshots of Django administration including login page (with username and password) and contacts and replies management

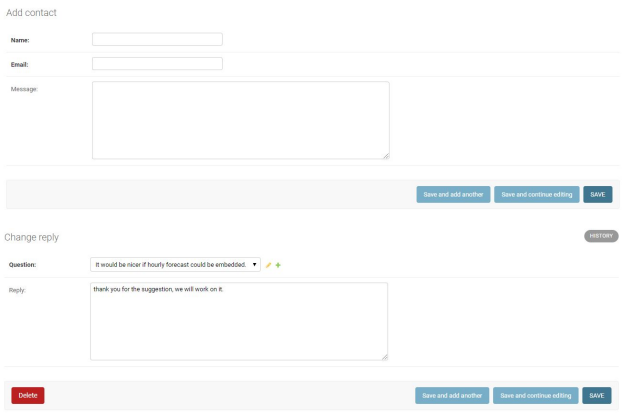

Figure 9: Screenshots of the site administration including contact addition and reply changing procedures and templates

5. Conclusion

The final web based weather forecasting application is now ready and can forecast the weather condition in context of Kathmandu city. The application can predict maximum temperature, minimum temperature, wind speed, and the probability of rainfall in Kathmandu valley for coming 5 days. The application has successfully embedded the provision of receiving feedbacks from the users and analyzes them for the improvement and growth of the system.

Acknowledgments

The author is grateful to the Department of Electronics and Computer Engineering of Himalaya College of Engineering for this opportunity. Without the support of the supervisor, Er. Ashok G.M. and the coordinator, Er. Narayan A.C., this would not have been possible.

References

[1] Randy Olson. Accuracy of three major weather forecasting services. [Online] http://www.randalolson.com/2014/06/21/accuracy-of-three-major-weather-forecasting-services/, June 2014. [Accessed: 03/06/2019].

[2] Sam Savage. Firm expands ways to get weather. [Online] https://www.redorbit.com/news/technology/752863/firm_expands_ways_to_get_weather/index.html/, December 2006. [Accessed: 03/06/2019].

[3] Jason Samenow. Accuweather extends its controversial, 45-day weather forecasts to 90 days. [Online] https://www.washingtonpost.com/news/capital-weather-gang/wp/2016/04/11/accuweather-extends-its-controversial-45-day-weather-forecasts-to-90-days/, April 2016. [Accessed: 03/06/2019].

[4] John Nese. Students put accuweather long-range forecasts to the test. [Online] https://www.washingtonpost.com/news/capital-weather-gang/wp/2013/12/26/students-put-accuweather-long-range-forecasts-to-the-test/, December 2013. [Accessed: 03/06/2019].

[5] Jason Samenow. Accuweather: You cannot be serious(new 45-day forecasts). [Online] https://www.washingtonpost.com/news/capital-weather-gang/wp/2013/08/06/accuweather-you-cannot-be-serious-new-45-day-forecasts/, August 2013. [Accessed: 03/06/2019].

[6] Nellie Andreeva and Mike Fleming Jr. Byron allen's entertainment studios acquires the weather channel tv network for $300 million. [Online] https://deadline.com/2018/03/byron-allen-entertainment-studios-acquisition-the-weather-channel-tv-network-1202351426/, March 2018. [Accessed: 03/06/2019].