# Temperature Prediction for Ahmedabad using Machine Learning

Umang Kamdar
*AU1841069*
Ahmedabad University
umang.k@ahduni.edu.in

Jainesh Patel
*AU1841101*
Ahmedabad University
jainesh.p@ahduni.edu.in

Vatsal Patel
*AU1841103*
Ahmedabad University
vatsal.p1@ahduni.edu.in

Shubh Shah
*AU1841122*
Ahmedabad University
shubh.s1@ahduni.edu.in

## I. Abstract

Temperature prediction and forecasting are becoming important day by day, as it's applications makes human life more comfortable in dealing with uncertain events like: Heatwaves, Droughts, Blizzards, Hurricanes, etc. In this paper, we have predicted temperature by implementing different varieties of Machine Learning algorithms like: Multiple Linear Regression, Polynomial Regression, Ridge Regression, Lasso Regression. Furthermore, the data-set features involved for predicting temperature tended to change as per the location, consequently we had to restrict our model to data-set of a particular region i.e, Ahmedabad City. We further restricted the use of certain features for forecasting the temperature, some of them were immaterial, according to our brainstorming sessions and some of them were found unassociated, according to mathematical analysis. After a series of actions, we were able to successfully execute various ML algorithms on the data-set and generate a good accuracy.

## II. Key Words

Machine Learning, Correlation, Regression, Temperature Prediction.

## III. Introduction

In real life we need to deal with many uncertain weather problems. We use the temperature prediction in Agriculture sector, Airport environments and some other important events. Thus, in order to forecast the chaotic nature of atmosphere we need to use some statistical approach. Only mathematical approach will not give us the insights of the temperature prediction. We need past data of the temperature and other features like Humidity, Wind index and others. We need to use algorithms which will include the past data and statistical calculations for predicting the temperature.

Our project is to predict the weather specifically for Ahmedabad city. We took a data-set of Ahmedabad city from [1] using API key. We collected the historical weather data of each day from 2015 to 2020. The data-set included various features related to weather which were: humidity, totalSnow, sunHour, precipMM, uvIndex, moon illumination, moonrise, moonset, sunrise, sunset, DewPointC, visibility, HeatIndexC, WindChillC, WindGustKmph, cloudcover, pressure, tempC, winddirDegree, windspeedKmph.

For weather prediction we chose Regression as the machine learning algorithm, as our weather data has real values, so the best algorithm for predicting temperature was Regression. There are many regression techniques and we tried some of them for testing and compare the accuracy of the prediction. First of all we tried Multiple Linear regression and in which we got good accuracy and after that we tried Polynomial, Lasso and Ridge regression. We got very interesting results in all of them.

## IV. Literature Survey

A simple weather prediction was performed using Multiple Linear Regression. In that model, a time-series data of weather was first collected. Temperature and relative humidity were predicted after doing feature selection based on correlation results [2]. Pearson correlation is used to find linear relationship between continuous variables [3]. Most of the time the data is noisy and it has null values and missing values so it needs to clean the data first [4]. To generalize the model well, Regularization is used to reduce the overfitting problem. [5]. Three Machine Learning models - Multiple Linear Regression (MLR), Support Vector Machine (SVM) and Artificial Neural Network (ANN) were compared for finding better Weather Prediction model. It was concluded that MLR was better for temperature prediction than others [6].

## V. Implementation

- Data prepossessing: First of all, data need to be processed before applying to the model to train the model well. Data may have null values or missing values which impact the performance of the model. So Initially we find out the features which have zero values and also find out the features which are redundant to predict the temperature. To make the feature independent of each other we performed correlation between the features and set the threshold to 0.9 to obtain the features which are dependant on some other features of the dataset. We found that the some of the columns have correlation value greater then 0.9. We removed those features as feature selection process and reduce the features from our dataset.
- Multiple linear Regression: Initially, we used inbuilt function of sklearn library to predict the temperature.

For that, we import all required libraries and split the data set into training and testing data set. Then by using inbuilt function of linear regression, we trained the model. Based on predicted parameters, we test the model on testing dataset and find the accuracy. After understanding of mathematics behind the multiple linear regression, we build custom function for regression. We used cost function under gradient descent method to find the good parameters.

$$cost = \frac{1}{2n} * \sum_{i=1}^{n}(f(x(i)) - y(i))^2$$

Here $f(x)$ is hypothesis function which is $\theta^T X$. Here actual value is subtracted from the hypothesis function value and its value is squared. n is the total number of training data.

To get good parameters we should minimise the cost function. For that gradient descent method is used.

$$\theta_{new}(i) = \theta(i) - \alpha * \frac{d}{d\theta}(cost)$$

$$\theta_{new}(i) = \theta(i) - \alpha * \frac{1}{n}(f(x) - y)X^T$$

Here $\alpha$ is learning rate and we iterate over multiple times such that we get at the situation where $\theta_{new} = \theta$ for gradient descent algorithm. By using this method we can find best parameters to predict the temperature accurately.

- We also tried polynomial regression to predict the temperature. We used inbuilt functions and the libraries. We choose the degree 2 and degree 3 for this model.
- We also tried to introduce regularization by performing regression using Ridge regression and the Lasso regression. The cost function for the Ridge regression is described as

$$cost = \frac{1}{2n} * \sum_{i=1}^{n}(f(x(i)) - y(i))^2 + \sum_{j=1}^{m}\frac{\lambda}{2n} * (\theta(j)^2)$$

here $\lambda$ is regularisation parameter. By minimising the cost function using gradient descent algorithm, we find the parameters which fit the model well and get good accuracy.

Cost function for Lasso regression is described as

$$cost = \frac{1}{2n} * \sum_{i=1}^{n}(f(x(i)) - y(i))^2 + \sum_{j=1}^{m}\frac{\lambda}{2n} * (\theta(j))$$

We still left to implement Lasso regression because it does not have closed form solution so we cannot apply gradient descent algorithm directly.

## VI. RESULTS

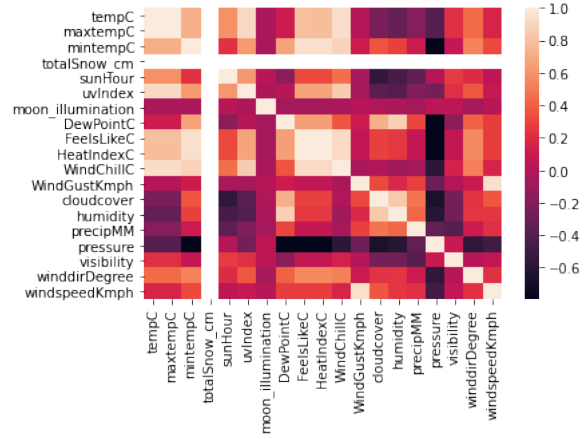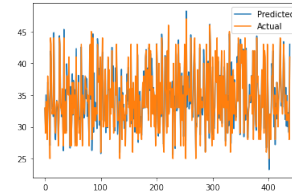| Machine Learning Algorithm | Accuracy |
|---|---|
| Multiple Linear Regression(Inbuilt) | 0.94187 |
| Multiple Linear Regression(Custom) | 0.9238 |
| Polynomial Regression(Degree=2) | 0.958642 |
| Polynomial Regression(Degree=3) | 0.956624 |
| Ridge Regression(Inbuilt) | 0.941834 |
| Ridge Regression(Custom) | 0.92578 |
| Lasso Regression | 0.94168 |



Fig. 1. Correlation of the features



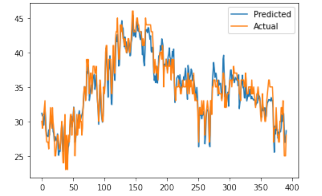Fig. 2. Ridge Regression (Inbuilt)



Fig. 3. Ridge Regression (Custom)

## VII. CONCLUSIONS

We performed four regression techniques for temperature prediction. We found out that out of all regression techniques, Polynomial Regression gives the best accuracy for our data. We found that degree 2 and degree 3 doesn't affect the accuracy much, but degree 2 is best. Next, we found that the accuracy of Multiple Linear Regression, Lasso Regression and Ridge Regression were almost similar. The accuracy result of custom made regression is also somewhat similar to inbuilt library functions.

## REFERENCES

[1] Ahmedabad Historical Weather. Worldweatheronline.Com, 2021, https://www.worldweatheronline.com/ahmedabad-weather-history/gujarat/in.aspx. https://dzone.com/articles/hashmap-
[2] Paras, Sanjay Mathur. "A simple weather forecasting model using mathematical regression." Indian Research Journal of Extension Education 12, no. 2 (2016): 161-168.
[3] "What Is The Pearson Coefficient?". Investopedia, 2021, https://www.investopedia.com/terms/p/pearsoncoefficient.asp.
[4] Folorunsho, Olaiya Adeyemo, Adesesan. (2012). Application of Data Mining Techniques in Weather Prediction and Climate Change Studies. International Journal of Information Engineering and Electronic Business. 4. 10.5815/ijieeb.2012.01.07.
[5] "Mathematics for Machine Learning". Copyright 2020 by Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. Published by Cambridge University Press.
[6] T. Anjali, K. Chandini, K. Anoop and V. L. Lajish, "Temperature Prediction using Machine Learning Approaches," 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 2019, pp. 1264-1268, doi: 10.1109/ICICICT46008.2019.8993316.