

# Temperature Analysis in India Using Machine Learning

G. Deepeka<sup>1</sup>, N. Ashulin Nivisha<sup>2</sup>

<sup>1,2</sup>B.E. Student, Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Erode, India

**Abstract:** Temperature analysis in India using machine learning is to analyse the temperature from the past to the present. In earlier days the temperature was moderate and there was no global Warming but Now-a-days lots of pollution has been caused due to industries, deforestation, burning of fossil fuels etc. Because of such calamities global warming has been gradually increased. Due to this problem it's necessary to know the land average temperature. It explains the prospects of analyzing climate change recorded trends in a biotic factors such as land surface temperature using data mining. The approach used here is graphing and clustering techniques. India's financial is the customarily farming in nature and extraordinary inconsistencies of the atmosphere dramatically affect the economy and in addition on the living states of the occupant's confirmations of anticipated atmosphere by various atmosphere models shows that the worldwide atmosphere is changing in an exceptional way because of quick increment in barometrical convergence of ozone harming substances.

**Keywords:** temperature analysis, machine learning

## 1. Introduction

Early data was collected by technicians using mercury thermometers, where any variation in the visit time impacted measurements. In the 1940s, the construction of airports caused many weather stations to be moved. In the 1980s, there was a move to electronic thermometers that are said to have a cooling bias. Environmental change is a standout amongst the most vital overall issues among the researchers and specialists. Since, the precipitation is altogether influenced by temperature changes and such changes are considered a vital for water assets arranging, agribusiness generation and hydrological cycles. Indian is particularly defenseless against the effects of environmental change essentially because of fast industrialization, populace weights and high rate of financial advancement in the nation. India's financial is the customarily farming in nature and extraordinary inconsistencies of the atmosphere dramatically affect the economy and in addition on the living states of the occupant's confirmations of anticipated atmosphere by various atmosphere models shows that the worldwide atmosphere is changing in an exceptional way because of quick increment in barometrical convergence of ozone harming substances.

Numerous specialists have been examined on temperature

pattern to evaluating the atmosphere under changing situations and revealed warming in the South Asian area including India.

A huge warming pattern was recognized in the normal most extreme, normal least and normal mean temperatures of the bowl. It detailed that adjustments in yearly temperatures might be ascribed mostly to the ascent in the base temperature caused by escalated urbanization.

## 2. Literature survey

1. *Armstrong in (1989) ascertains that the forecast by combining gives constant and sufficient increase in the accuracy:* In this paper the researcher also reveals that the Combining forecasts by Clemen, (1989) is a landmark on the area of forecasting by combining and it is also an eye opener for all the researchers in the area of forecasting. The researcher on the wide angle concentrated on three types of forecasting analysis such as meta analysis, realistic simulations and rule based forecasting.
2. *Assaad et al in (2008) have stated the importance of ensemble methods for solving classification and regression problems:* In this paper the researchers reveal that the ensemble model performs better for both theoretical and empirical than single models. The authors have developed an ensemble model by using the base learner as Recurrent Neural Networks (RNNs). During the learning process, the boosting algorithm is used for difficult points of the time series data set and a new parameter is used for tuning the boosting influence on the data sets.
3. *Kurniady and Kosala in (2011) have built a model of financial forecasting for time series:* It is a derived time series data and the news acquired from the internet which uses multilayer neural network taught with genetic algorithm. For the purpose of classification of the news, incrementally trained Naive Bayesian classifier is used. The outcome of the investigational result shows that the addition of news does increase the accuracy of prediction for certain stocks and the index.

## 3. Problem definition

This project explains the prospects of analyzing climate change recorded trends in a biotic factors such as land surface

temperature using data mining. The approach used here is graphing and clustering techniques. The temperature data came from 395 met stations across the country. Maximum and minimum and daily mean temperature data was analyzed for summer, monsoon and winter periods. The maximum and minimum temperature data of 37 year has been analyzed by using Mann-kendall statistical test to assess the trends under changing climate. The season were classified as per Indian meteorological department criteria. These gases have increased by anthropogenic activities such as burning fossil fuels, farming activities and land use changes. The carbon-dioxide has been increased about 40% since pre-industrial time of this, 30% was observed in the ocean. It results in ocean acidification and rapid rise in surface temperature in the past 70 years.

#### 4. Machine learning and its algorithms

##### A. Predictive analysis

The predictive analysis system architecture comprises various phases like data collection, missing values imputation, pattern discovery, pattern matching and result analysis. Fig. 1, describes overall architecture of proposed work.

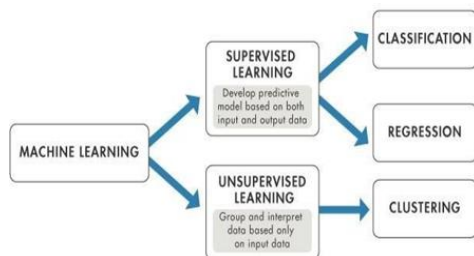


Fig. 1. Predictive analysis system architecture

Use K-means algorithm with different values of seed. we are using unsupervised learning. Logistic Regression Algorithm used for supervised classification.

##### B. Machine learning

Machine learning is a collection of methods that can automatically identify patterns in data, and then use those patterns to predict future outcomes, or to perform other types of decision making below certain conditions. Machine learning introduces various algorithms, those enable machines to understand the current situations and on the basis of that machines can take appropriate decisions. Machine learning works independently and takes decision at its own.

The main two types of machine learning are, supervised learning and unsupervised learning.

##### 1) Supervised Learning

In supervised learning, the input and its corresponding output is already known. This is called supervised learning because it learns from training data set and creates model from it and when this model applies on new data set it gives predicted results. Decision Tree, naive Bayes etc. are the examples of supervised learning.

##### 2) Unsupervised Learning

Unsupervised learning is where we have only input data and no corresponding output variable. The main job of unsupervised learning is to build up class labels automatically. The relationship between the data can be found using unsupervised learning algorithms to discover whether the data can characterize to form a group. This group is known as clusters. Unsupervised learning can be also described as cluster analyses”. K Means Clustering, KNN etc are the examples of unsupervised learning.

##### C. Selected online dataset

India’s land temperature is used for the experimentation. The data set are available at UCI Machine Learning Repository. The data set contains attributes and one class variable. Early data was collected by technicians using mercury thermometers, where any variation in the visit time impacted measurements. In the 1940s, the construction of airports caused many weather stations to be moved. In the 1980s, there was a move to electronic thermometers that are said to have a cooling bias. Given this complexity, there are a range of organizations that collate climate trends data. The three most cited land and ocean temperature data sets are NOAA NASA GISTEMP and the UK’s HadCrut. The Berkeley Earth Surface Temperature Study combines 1.6 billion temperature reports from 16 pre-exist it.

Date starts in 1750 for average land temperature and 1850 for max and min land temperatures and global ocean and land temperatures Land Average Temperature-global average land temperature in Celsius Land Average Temperature Uncertainty-the 95% confidence interval around the average Land Max Temperature-global average maximum land temperature in Celsius. Land and Ocean Average Temperature Uncertainty-the 95% confidence interval around the global average land and ocean temperature.

##### D. Selected algorithm for implementing

##### 1) K-means algorithm

The k-means clustering algorithm and machine learning tool used to cluster observations into groups of related observations without any prior knowledge of those relationships. By sampling, the algorithm attempts to show in which category, or cluster, the data belong to, with the number of clusters being defined by the value k.

The k-means algorithm is one of the simplest clustering techniques and it is commonly used in medical imaging, biometrics, and related fields. The advantage of k-means clustering is that it tells about your data (using its unsupervised form) rather than you having to instruct the algorithm about the data at the start (using the supervised form of the algorithm).

K-means clustering is a data mining/machine learning algorithm used to cluster observations into groups of related observations without any prior knowledge of those relationships. The k-means algorithm is one of the simplest clustering techniques and it is commonly used in medical imaging, biometrics and related fields.

2) Working of k-means algorithm

The k-means calculation is a developmental calculation that picks up its name from its strategy for task. The calculation bunches perceptions into k gatherings, where k is given as an info parameter. It at that point allots every perception to groups dependent on the perception's vicinity to the mean of the bunch. The bunch's mean is then recomputed and the procedure starts once more. Here's the means by which the calculation works.

Each point in the dataset is allocated to the shut group, in view of the Euclidean separation between each point and each bunch center. Each group focus is recomputed as the normal of the focuses in that cluster. Steps 2 and 3 rehash until the bunches converge. convergence might be characterized contrastingly relying on the usage, yet it ordinarily implies that either no perceptions change groups when stages 2 and 3 are rehashed, or that the progressions don't have a material effect in the meaning of the groups.

3) Decision tree

Choice tree learning utilizes a choice tree as a prescient model which maps perceptions around a thing to decisions about the thing's objective esteem. More distinct names for such tree models are order trees or relapse trees. In these tree structures, leaves speak to class marks and branches speak to conjunctions of highlights that prompt those class names.

In information mining, a choice tree depicts information yet not choices; rather the subsequent arrangement tree can be a contribution for basic leadership. J48 are the enhanced adaptations of C4.5 calculations or can be called as upgraded execution of the C4.5. The yield of J48 is the Decision tree. A

Decision tree is like the tree structure having root hub, moderate hubs and leaf hub. Every hub in the tree comprise a choice and that choice prompts our outcome. Choice tree isolate the info space of an informational collection into fundamentally unrelated regions, every zone having a mark, an esteem or an activity to depict its information focuses. Part measure is utilized to figure which credit is the best to part that bit tree of the preparation information that achieves a specific hub.

5. Output graph

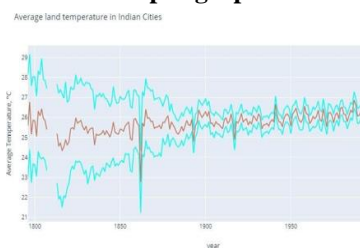


Fig. 2. Average land temperature in Indian cities



Fig. 3. Temperature shows upward trend in India

6. Results of implemented algorithms in python

The statistically significant of the trend was evaluated on the basis of Z value. A positive value of Z indicates an upward trend in the data with time series. The results revealed that, the annual mean maximum and minimum temperature was found 33.8, 19.2 0C respectively. Annual mean maximum and minimum temperatures showed increasing trend with almost same rate (0.0020C/year). But statistically non-significant. Khavse It also observed increasing trend of mean maximum temperature in Raipur. The increasing trend in maximum and minimum temperature has been also reported by Solomon in Ethiopia.

A comparable pattern in most extreme temperature was seen by Sharad contemplated on fluctuation and time arrangement slant examination of precipitation and temperature in north focal Ethiopia and found critical expanding pattern for least normal temperature while the pattern for greatest temperature displayed non- noteworthy expanding. Least temperatures indicated expanding pattern over the north Indian states in all seasons while an expanding pattern over the west shore of India in winter and southwest storm seasons. It might be accounted for from the previous outcomes that both greatest and least temperature observed to be expanded in every one of the seasons might be because of the reality of environmental change. The expanding pattern in most extreme and least temperature has been likewise announced by Solomon in Ethiopia. It considered on changeability and time arrangement slant investigation of precipitation and temperature in north focal Ethiopia and found critical expanding pattern for least normal temperature while the pattern for most extreme temperature showed non-huge expanding.

7. Conclusion

After implementing algorithm, From the long term analysis of temperature at India, it may be concluded that the annual mean maximum and minimum temperatures exhibited increasing trend. In all the seasons except monsoon, maximum temperature showed increasing trend. Minimum temperature exhibited increasing trend in all the seasons. Since, significant positive trend showed in minimum temperature during monsoon season.

References

- [1] Amogne Asfaw a, Belay Simane, AliHassen, Amare Bantider (2018). Variability and time series trend analysis of rainfall and temperature in North Central Ethiopia: A case study in Woleka sub- basin. Weather and Climate Extremes 19, 29–41.
- [2] Arora M., Goel N.K., Singh R. (2005). Evaluation of temperature trends over India. Hydrol Sci J 50: 81-93.
- [3] Bhutiyani M. R., Kale V. S. and Pawar N. J. (2007). Long-term trends in maximum, minimum and mean annual air temperatures across the North western Himalaya during the twentieth century. Climatic Change, 85,159–177.
- [4] Ceppi P., Scherrer S.C., Fischer A.M. and Appenzeller C. (2008). Revisiting Swiss temperature trends 1959–2008.

- [5] Cheema M.A., Farooq M., Ahmad R. and Munir H. (2006). Climatic trends in Faisalabad (Pakistan) over the last 60 years (1945- 2004). *J Agri Soc Sci* 2: 42-45.
- [6] Hamid A.T., Sharif M. and Archer D. (2014). Analysis of Temperature Trends in Satluj River Basin, India. *Earth Sci Clim Change*, Vol. 5, Issue 8, 1000222.
- [7] Hingane L. S., Rup Kumar K. and Ramanamurthy B. V. (1985). Long term needs of surface air temperature in India. *Int J Climatol*.