IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

WEATHER CLUSTERING USING K-MEANS

M.Yashaswini, G. Sowmya, P. Jai deep
Department of Computer Science and Engineering
St. Peter's Engineering College
Hyderabad, India

Shaik Mohammed Ilias (Assistant Professor)
Department of Computer Science and Engineering
St. Peter's Engineering College
Hyderabad, India

Abstract – Weather Prediction is for the most part worried about the forecast of climate condition in the given future time. Weather prediction give basic data about future climate. Predicting the weather is fundamental to help planning generally advantageous and the most exceedingly terrible atmosphere. Here K-Means Clustering algorithms is used to identify the variation in the weather condition in terms of Temperature, Humidity, Wind Speed, Pressure and Weather. Weather predictions are significant for arranging our everyday exercises. Farmers need data to enablethem to anticipate the planting and gathering of their yields. Prediction of weather condition is essential for variousapplicationslike Agricultural, Industry, Air Traffic, Marine, Forestry, Army and Navy etc.

Keywords – Clustering, Forecasting, Incremental, K-means.

I.INTRODUCTION

Many weather predictions like rainfall prediction, thunderstorm prediction, predicting cloud conditions are major challenges for atmospheric research. In Here K-Means Clustering and Naive Bayes Algorithms is used to identify the variation in the weather condition in terms of Temperature, Humidity, Wind Speed, Pressure and Weather.

Predicting the weather is fundamental to help planning generally advantageous and the most exceedingly terrible atmosphere. We should be on caution to the unfriendly climate conditions by adjusting a few safeguards and utilizing expectation instruments for early cautioning of dangerous weather wonders.

Weather Prediction is for the most part worried about the forecast of climate condition in the given future time.
Weather prediction give basic data about future climate. Predicting the weather is fundamental to help planning generally advantageous and the most exceedingly terrible atmosphere.

EXISITNG SYSTEM:

Weather Prediction is for the most part worried about the forecast of climate condition in the given future time. Weather prediction give basic data about future climate. Predicting the weather is fundamental to help planning generally advantageous and the most exceedingly terrible atmosphere.

Weather Prediction an is done manually these tasks are time consuming and not reliable It may vary actual situation and it is not convenient because in the manual system there are lot of difficulties in conducting.

PROPOSED SYSTEM:

Weather Prediction an is done manually these tasks are time consuming and not reliable It may vary actual situation and it is not convenient because in the manual system there are lot of difficulties in conducting.

i. Enables organizations and individuals to get ready for power generation and how much capacity to utilize.

ii. Enables individuals to plan in the event that they have to take additional rigging to get ready for the climate.

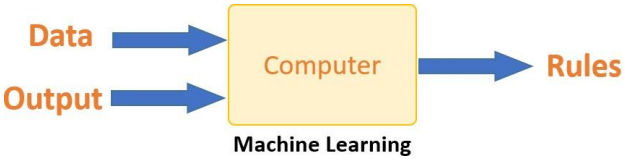
iii. Enables organizations to get ready for transportation risks that can result from the climate.

II.TECHNOLOGY

What is Machine Learning?

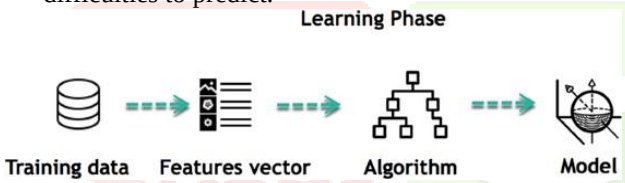
Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed

The primary aim of ML is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.



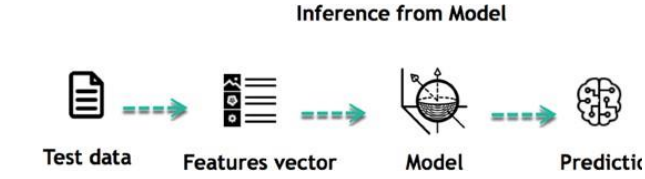
How does Machine learning work?

Machine learning is the brain where all the learning takes place. The way the machine learns is similar to the human being. Humans learn from experience. The more we know, the more easily we can predict. By analogy, when we face an unknown situation, the likelihood of success is lower than the known situation. Machines are trained the same. To make an accurate prediction, the machine sees an example. When we give the machine a similar example, it can figure out the outcome. However, like a human, if its feed a previously unseen example, the machine has difficulties to predict.



Inferring

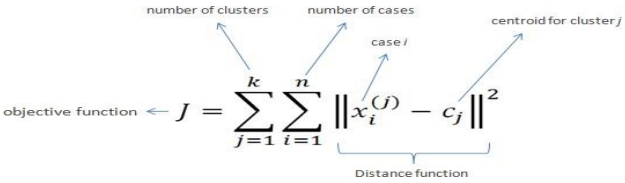
When the model is built, it is possible to test how powerful it is on never-seen-before data. The new data are transformed into a features vector, go through the model and give a prediction. This is all the beautiful part of machine learning. There is no need to update the rules or train again the model. You can use the model previously trained to make inference on new data.



III. METHODOLOGY

k-means algorithm

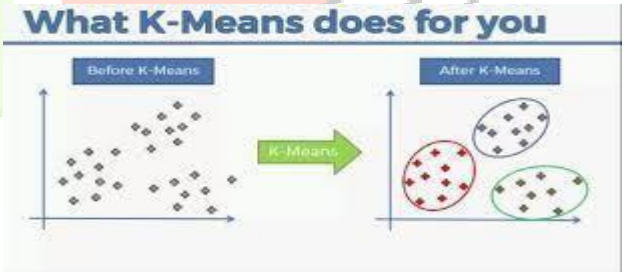
k-means is a widely used clustering algorithm. It creates 'k' similar clusters of data points. Data instances that fall outside of these groups could potentially be marked as anomalies. Before we start kmeans clustering, we use elbow method to determine the optimal number of clusters.



How the K-means algorithm works

To process the learning data, the K-means algorithm in data mining starts with a first group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative (repetitive) calculations to optimize the positions of the centroids It halts creating and optimizing clusters when either:

- The centroids have stabilized — there is no change in their values because the clustering has been successful.
- The defined number of iterations has been achieved.



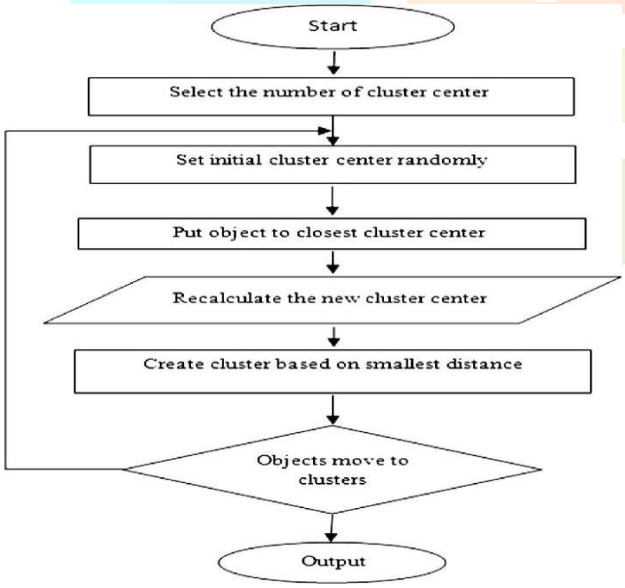
IV. IMPLIMENTATION

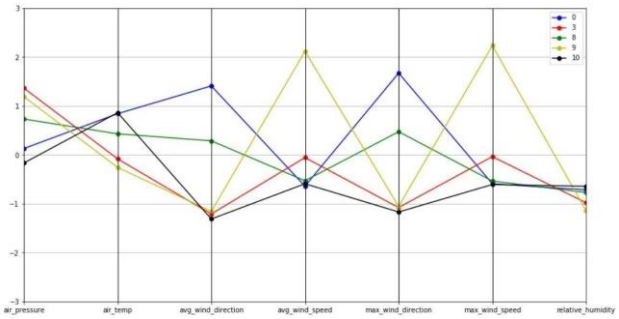
Clustering with scikit-learn Importing the Necessary Libraries from sklearn. preprocessing import StandardScaler from sklearn.cluster

Each row in **minute_weather.csv** contains weather data captured for a one-minute interval. Each row, or sample, consists of the following variables:

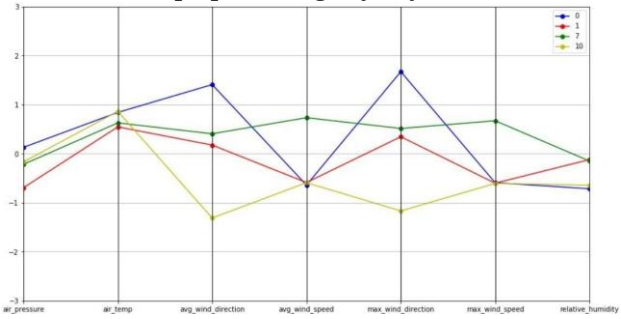
- **rowID:** unique number for each row (Unit: NA)
- **hpwren_timestamp:** timestamp of measure (Unit: year-month-day hour:minute:second)
- **air_pressure:** air pressure measured at the timestamp (Unit: hectopascals)
- **air_temp:** air temperature measure at the timestamp (Unit: degrees Fahrenheit)

- **avg_wind_direction:** wind direction averaged over the minute before the timestamp (Unit: degrees, with 0 means coming from the North, and increasing clockwise)
- **avg_wind_speed:** wind speed averaged over the minute before the timestamp (Unit: meters per second)
- **max_wind_direction:** highest wind direction in the minute before the timestamp (Unit: degrees, with 0 being North and increasing clockwise)
- **max_wind_speed:** highest wind speed in the minute before the timestamp (Unit: meters per second)
- **min_wind_direction:** smallest wind direction in the minute before the timestamp (Unit: degrees, with 0 being North and inceasing clockwise)
- **min_wind_speed:** smallest wind speed in the minute before the timestamp (Unit: meters per second)
- **rain_accumulation:** amount of accumulated rain measured at the timestamp (Unit: millimeters)
- **rain_duration:** length of time rain has fallen as measured at the timestamp (Unit: seconds)

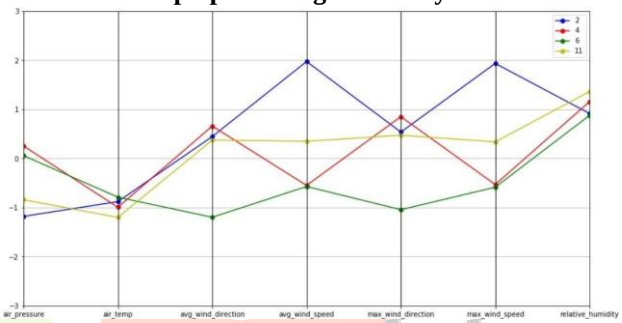




Graph predicting dry days



Graph predicting warm days



Graph predicting cool days

V. CONCLUSION AND FUTURE SCOPE

The project has been successfully completed. The goal of the system is achieved and the problems are solved. This project is developed in this manner that is user friendly and required help is provided at different levels.

In wake of the new and related trends, it is imperative frequent upgrades to a new models and algorithms to make it easier according to address new business needs.

REFERENCES

[1] Chakraborty, S. and Nagwani, N.K. , " Analysis and study of Incremental K-Means clustering algorithm",Communication in Computer and Information Science. 1, *International conference in High Performance Architecture and Grid Computing (Springer Germany), Vol. 169*, part 2, 2011, pp.338-341.

[2] Mumtaz, K. and Duraiswamy K. , " A Novel Density based improved k-means Clustering Algorithm – Dbkmeans ", *IJCSE*, Vol. 2, No. 02, 2010, pp.213-218,.

[3] Kanungo, T. and Mount D.M. : " An Efficient k-Means Clustering Algorithm : Analysis and implentation" , *IEEE Transaction*, Vol. 24, No. 7, 2002.

[4] Ordonez, C. and Omiecinski E.: "An Efficient Disk-Based K-Means Clustering for Relational Databases, *IEEE transaction on knowledge and Data Engineering,*Vol.16,2004.

[5] Vipul K., Vamsidhar T., Kamalakar K., "Time Series Forecasting through Clustering - A Case Study", In Proceedings of COMAD'2005, pp.183~191.

[6] Xiang L., Beth P., Nithya V., Rahul R., Sara G. and Helen C., "Real-Time Storm Detection and Weather Forecast Activation through Data Mining and Events Processing",Vol.1, No.2, May 2008, pp. 49-57.

[7] Dr. Santhosh B. S. and Kadar Shereef I., "An Efficient Weather Forecasting System using Artificial Neural Network", International Journal of Environmental Science and Development, Vol. 1, No. 4, October 2010.

[8] Godfrey C. O., Peter B. , Sitaram G., Visagaperuman R., Viti B. A. A., " Self-organizing Data Mining for Weather Forecasting", IADIS European Conference Data Mining, ISBN: 978-972-8924-40-9, 2007.

[9] "Air Pollution" , Wikipedia free encyclopaedia http://en.wikipedia.org/wiki/Air_pollution#mw-head.

[10] "Air pollution and Climate change", published by Science for Environment Policy, European Commission, issue 24, November-2010.

[11] Hans E., Markus A., Jelle G. van M., " The ETC-ACC framework for base-line scenario development in the context of integrated assessment for air pollution and climate change" , European Topic Centre on Air and Climate Change(ETC/ACC), 2002.