



## SYSTEMATIC MAPPING STUDY ON DATA MINING PROCESS TO WEATHER DATA ANALYSIS

<sup>1</sup>R. Jayakumar, <sup>2</sup>Dr. R. Annamalai Saravanan,  
<sup>1</sup>Research Scholar, <sup>2</sup>Associate Professor and Head,  
<sup>1</sup>Department of Computer Science, <sup>2</sup>Department of Information Technology,  
<sup>1,2</sup>Sankara College of Science and Commerce,  
<sup>1,2</sup>Coimbatore, Tamilnadu, India.

**Abstract** - Data mining gives the methodology and technology to change huge measure of data into helpful information for decision making. High throughput weather data, which is obtained by remote detecting innovation, gathered by neighborhood weather stations, or assembled via autonomous sensors, is the establishment for modern weather forecast and environmental change prediction. Data mining advances, which is the computer helped process that uncovers valuable examples from underneath large-scale data sets, is broadly recognized as an extremely encouraging bearing in weather data analysis. Approaches, for example, Active Methodologies are known as having the capacity to include and propel understudies. This survey presents a systematic planning expecting to identify current weather Data Mining and Learning Analytics methods.

**Keywords:** [Systematic mapping, Data mining, Weather data analysis.]

### 1. INTRODUCTION

The effective analysis of weather data is grabbing increasing eye among scientists and investigates over the new years. The weather limits recorded over the new many

years has experienced a quick increase in both recurrence and extent, prompting tremendous financial misfortune as well as undermining individuals' life. Accordingly, it is essential for the public authority, research establishments, as well as disaster the board associations to comprehend the regular peculiarities and predicting the weather limits accurately. Weather data analysis, which removes the valuable information behind the noticed data, establishes the framework of the prediction and forecasting of weather limits. Weather data analysis has been generally contemplated since years and years prior. A systematic planning study (SMS) is an interaction for the investigation of the circumstance of a wide exploration region with a high level of granularity, permitting us to recognize areas in the space where it could be fascinating to investigate in more detail. In any case, ongoing examination uncovers that the yield environment models neglect to catch numerous drivers and experience the ill effects of a misstatement of likely unfriendly impacts of weather limits. Factors, for example, the weather limits like dry seasons and floods, the surrounding temperatures, and the dirt dampness, are not completely thought to be in these models.

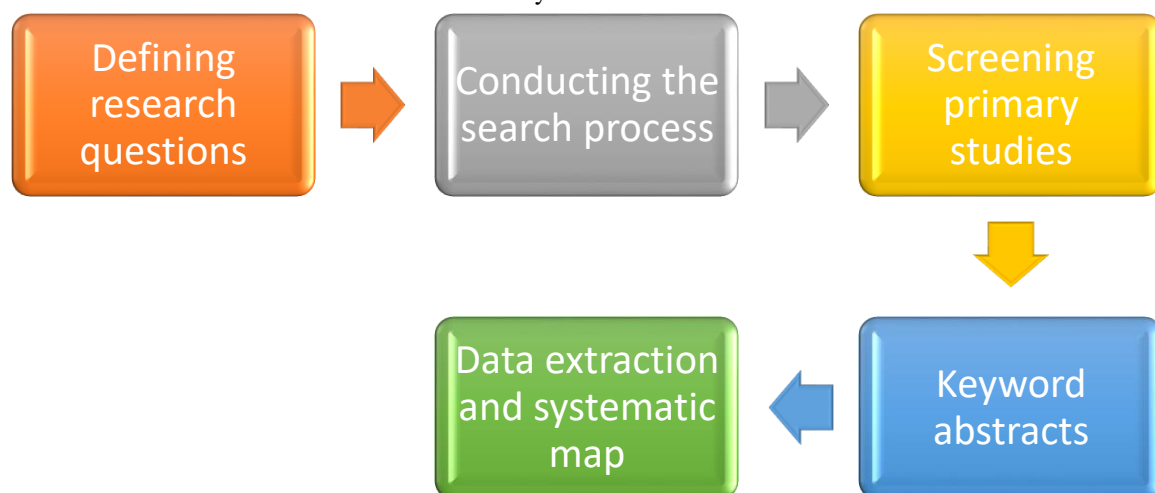


Figure 1. Systematic mapping process

Data mining techniques, with the guide of the computers to burrow valuable information/designs out of large-scale data sets, is generally utilized in many fields and is perceived as an extremely encouraging bearing for weather data analysis. For example, artificial neural network and decision tree algorithms are among the famous models

to predict weather and environment changes. In this study portray a distance-metric-learning based data digging procedure for weather data analysis. Traditionally, data is physically gathered which is both time and cost consuming, and accordingly it is the data assortment stage that is generally considered as the most significant and rate-

restricting advance of data analysis. Be that as it may, with the new improvement of computer science and gadgets innovation, for example, remote detecting innovation and remote sensors, the data assortment has become moderately easy, and the accessible data set is extending at unprecedented rates. Presently, the focal point of data analysis has moved to uncovering the valuable information behind the immense data sets and understanding them in an effective and unbiased manner. Data digging techniques take into account breaking down tremendous dataset and removing valuable information.

## 2. LITERATURE SURVEY

**1. Chammas M, Makhoul A, Demerjian J.** (2019) et.al proposed an efficient data model for energy prediction using wireless sensors. One of the key factors for coordinating systems and making homes and urban communities wiser is the energy reserve funds that can be accomplished. In this work, they propose to demonstrate an energy prediction system in view of a Multilayer Perceptron neural network (MLP). A few data classifications have been utilized, i.e.; light energy, temperature, humidity, day of the week, and so on, gathered from a Wireless Sensor Network introduced in a two story building. They looked at the presentation of the system against four order algorithms, specifically: Linear Regression, Support Vector Machine, Gradient Boosting Machine and Random Forest. They evaluated the significance of transient information and light energy as extra highlights for the prediction model. The transient highlights slightly worked on the exhibition for all systems, while the lights energy diminished the presentation. For the strategy, they note a diminishing by 10% in relative execution while the thing that matters was unimportant for different classifiers (under 1% for Gradient Boosting Machine). They system outflanked the other four classifiers in all situations. In light of the led experiments, they infer that climate information would be to the point of anticipating energy consumption. This empowers minimal expense answers for energy predictions. Further experiments ought to be performed on various datasets to approve they results.

### Merits

The temporal features slightly improved on the performance for all systems.

### Demerits

Improve energy and delay

**2. Janković R, Čosović M, Amelio A.** (2019) et.al proposed Time series prediction of air pollutants: a case study for Serbia, Bosnia and Herzegovina and Italy. This paper presented a NARX neural network model to foresee the values of NO<sub>2</sub> and CO, as far as meteorological parameters. About Serbia, a more exact prediction is acquired for NO<sub>2</sub>, while as far as Bosnia and Herzegovina and Italy; more precise predictions are gotten for CO. A correlation of the NARX model with SVM regression and regression trees demonstrated that the proposed model can defeat other notable regression strategies for time series prediction of CO outflow in all nations, and time series prediction of NO<sub>2</sub> in Bosnia and Herzegovina. In show disdain toward the promising outcomes got by the model, researching about the impact of interaction of among the

utilized meteorological variables (pressure, relative humidity, temperature, wind speed) on anticipating the air contaminations (NO<sub>2</sub> and CO) is past the extent of this work. Preliminary tests demonstrated that extra exertion in the NARX model is expected to acquire cutthroat outcomes while thinking about such interaction, because of production of instable arrangements when the meteorological variables are utilized related. In like manner, an expansion of the proposed model will be given as a future work bearing for exploring and assessing such interaction in the prediction process.

### Merits

The NARX model with SVM regression and regression trees demonstrated that the proposed model can beat other notable regression strategies.

### Demerits

A direction will investigate and assessing such interaction in the prediction process.

**3. Finamore A, Calderaro V, Galdi V, Piccolo A, Conio G** (2019) et.al proposed A Day-ahead Wind Speed Prediction based on Meteorological Data and the Seasonality of Weather Fronts. A reliable and accurate forecasting model is quite possibly the best answer for manage the issue of renewable energy sources incorporation. In this paper, a model for the medium-long haul wind speed prediction, in view of spatiotemporal advancement of weather fronts and Multi-Layer Perceptron Neural Network (MLP NN) data mining model, is created. In this paper, ANN MLP based way to deal with anticipates one day wind speed is showed. The key thought of the proposed predictor depends on the building of a dynamic training data set that incorporates the physical peculiarities related with the meteorological fronts and their development, in view of a seasonal clustering of data. To acquire a single out, the neural predictor with the best presentation, in term of MAPE, ANN MLP is prepared with three distinct data sets: one for the midyear, one for the winter and the final remaining one for the exchange seasons (spring and autumn). The MLP, prepared with the best data set, gauges wind speed in the most basic months and where weather fluctuations are ordinary of the seasonal changes, further developing significantly the past results acquired utilizing the musicale approach, with a prediction blunder under 15%.

### Merits

The MLP, prepared with the best data set, figures wind speed in the most basic months and where weather fluctuations.

### Demerits

The decrease of the greenhouse gas discharges combined with the energy consumption.

**4. Edona Doko, Lejla Abazi Bexheti** (2018) et.al proposed A Systematic Mapping Study of Educational Technologies based on Educational Data Mining and Learning Analytics. The traditional flipped classroom is characterized in light of the video directions of flipped class at home. What's more, have opportunity and willpower in class to work with their instructor on key learning exercises. With presentation of

Flipped Classroom model and quick improvement of data sciences for taking the advantages of the learners' data to enhance learning process there has been an expanding thought of taking advantage of Educational Data Mining and Learning Analysis. Huge measure of data made from performed moves of understudies and instructors in FC has made data-driven learning and self-guided learning conceivable to perform. The objective in this paper is to devise a systematic planning study, to investigate and break down existing explores about executions of learning procedures and learning algorithms. They chose 122 papers and grouped by field of interest, video field of interests and technology utilized in FC.

#### Merits

A systematic to further developed learning interactivity by means of videos in educational technologies.

#### Demerits

Teacher not knows precisely what portions of the substance in every theme are generally hard for learners.

**5. A. Alvarez-Ayllon, M. Palomo-Duarte and J. -M. Dodero** (2019) et.al proposed Interactive Data Exploration of Distributed Raw Files: A Systematic Mapping Study. While exploring enormous measures of data without a reasonable objective, giving an interactive encounter turns out to be truly troublesome, since this conditional examination ordinarily overcomes any early choice on data structures or ordering procedures. This is likewise evident in the physics area, explicitly in high-energy physics, where the enormous volume of data produced by the finders are typically investigated through C++ code utilizing cluster processing, which presents an impressive inertness. An interactive instrument, when incorporated into the current data the board systems, can increase the value of the ease of use of these stages. Here, they mean the present status of-the-art of interactive data investigation, targeting fulfilling three necessities: admittance to crude data records, put away in a disseminated climate, and with a sensibly low inactivity. This paper observes the rules for systematic mapping studies, which is appropriate for social occasion and ordering accessible studies. They sum up the results subsequent to grouping the 242 papers that passed them consideration measures. While there are many proposed solutions that tackle the issue in various habits, there is little proof accessible about their execution practically speaking. Practically every one of the solutions found by this paper cover a subset of the necessities, with just one partially fulfilling the three. The solutions for data investigation flourish. It is an active research region and, thinking about the nonstop development of data volume and assortment, is just to become more earnestly. There is a specialty for research on an answer that covers them prerequisites, and the necessary building blocks are there.

#### Merits

There distributed on their down to earth execution confirm a successful introduction to clients.

#### Demerits

Not equipped for adapting themselves to changes on the workflow.

**6. L. Dong, B. Liu, Z. Li, O. Wu, M. A. Babar and B. Xue** (2017) et.al proposed A Mapping Study on Mining Software Process (MSP). Mining Software Process distills significant information about software process enactment from software data archives. An expanding measure of research effort is being committed to MSP. These studies contrast in different viewpoints (e.g., topics, data, and techniques) of MSP. Objective: They expect to study the state of the art on MSP from following viewpoints, i.e., research topics, data sources, data types, mining techniques, and mining apparatuses. Strategy: They led a systematic mapping study on the research applicable to MSP at both microprocess and macroprocess levels. The recognized mining techniques have been planned onto the related mining devices that fall into four sorts. Driven by the three research questions which addressed in a meta-model, the discoveries uncovered the relationships among the research topics, data sources, data types, mining techniques, and mining devices. End: It is seen that to find the software process model or guide, the primary data source is from modern task. Momentum mining techniques for microprocess research are for the most part business process mining or sequence mining techniques used to recuperate elucidating software process. Likewise, different machine learning algorithms and novel proposed techniques are utilized to work on the accuracy of macroprocess level factors (e.g., software effort estimation).

#### Merits

Prefer to utilize sequence mining and some high level process disclosure techniques that can be supported by the full grown mining tools for business process like ProM and Disco.

#### Demerits

Evaluate the process mining not track down proper techniques for mining business projects.

**7. Wang H, Gu C, Zhang X, Li F, Gu L.** (2018) et.al proposed Identifying the correlation between ambient temperature and gas consumption in a local energy system. In this paper, an original strategy is proposed to analysis the relationship between's gas consumption and temperature. Mahalanobis distance is a promising mean for outlier recognition and EMD considerably further develops the relationship level between gas load and temperature. By observing relationship coefficient between continuous gas consumption and temperature, it is seen that gas load is contrarily connected with temperature. What's more, such relationship changes with the general temperature level. In winter, the gas and temperature are in most noteworthy connection and this relationship drops when the temperature is higher in spring and autumn. By isolating the data into weekday/end of the week and day/night, it is figured out that during various opportunity periods, the connection is more grounded during weekday and night than that during end of the week and day time, individually. In the wake of applying the EMD procedure, relationship levels are significantly expanded in all of the different contextual analyses. To sum up, the original strategy proposed in this paper distinguishes the temperature sensitive part of the gas load and measures the connection between's the gas load and temperature as the vast majority of the other research centers just around the connection between electricity load

and temperature and not very many spotlight on the gas load and temperature. This gives a reference to weighting temperature as a variable in gas load forecasting.

#### Merits

Identifies the temperature sensitive part of the gas load and measures the connection between's the gas load and temperature.

This gives a reference to weighting temperature as a variable in gas load forecasting.

#### Demerits

Polynomial regression techniques will be applied to break down weather factors like humidity and wind.

**8. Thi Thi Shein, Sutheera Puntheeranurak** (2018) et.al proposed Incremental Clustering Approach for Evolving Trajectory Data Stream. Trajectory data stream contain an enormous measure of data about spatial and temporal information of moving items. Clustering the trajectory might carry advantages to a few applications, for example, traffic monitoring system and conduct analysis of animal movement design, weather forecasting. In numerous genuine applications, trajectory data continue to come into the database or server for guaranteed analysis. Most existing methodologies investigate the entire article trajectory from the static database rather than the current movement dynamic data. These methods can't get a brief outcome on the grounds that the items generally move then their position has changed over the long haul. In this paper, they address the issue of monitoring the development of moving articles over the long haul and propose incremental Sub-Trajectory Clustering in view of Micro-group (iSTCM) system to lessen computational time complexity. As a test, the presentation of they proposed algorithm will direct on genuine taxicab datasets and look at the productivity and bunch quality as viability with one more state of the art methods.

#### Merits

By utilizing micro-group, they can be refreshed rapidly with any new information.

This makes it more appropriate for genuine applications.

#### Demerits

The issue of efficiently keeping a clustering of a dynamic arrangement of data focuses that move constantly with temporal consideration.

**9. Daraio E, Corso ED, Cerquitelli T, Chiusano S.** (2018) et.al proposed Characterizing Air-Quality Data through Unsupervised Analytics Methods. A few cities have based on-the-ground air quality monitoring stations to quantify every day concentration of air pollutants, as PM10 and NO2. The recognizable proof of the causalities for air pollution will help governments' direction on mitigating air pollution and on prioritizing suggestions. This paper presents a two-level methodology in light of unaided investigation methods, named PANDA, to find interesting bits of knowledge from air quality-related data. In the first place, PANDA finds groups of pollutants that have happened with comparative concentrations. Then, at that point, each group is privately portrayed through three types of comprehensible information to give interesting

correlations between air pollution and meteorological conditions at various abstraction levels. As a contextual analysis, PANDA has been approved on genuine contamination estimations gathered in a significant Italian city. Preliminary experimental results show that PANDA is compelling in finding strong and very much isolated groups of comparative concentrations of pollutants.

#### Merits

PANDA a two-level methodology ready to extricate groups of pollutants.

Predictable for correlations between poison concentrations and meteorological features.

#### Demerits

The PANDA motors with prediction capacities not conjecture toxin concentrations.

**10. Niu Z, Zhu Y, Jiang L.** (2018) et.al proposed Distance Metric Learning Approach for Weather Data Mining. High throughput weather data, which is obtained by remote sensing technology, gathered by local weather stations, or accumulated via independent sensors, is the establishment for present day weather conjecture and climate change prediction. Weather data analysis is a basic advance in weather estimate and climate change prediction. With the quickly growing data sets on account of remote sensing and wireless sensor technologies, data mining techniques are profoundly compelling and efficient in weather data analysis. Such data set regularly contains multidimensional information on viewpoints, for example, temperature, humidity, wind speed/bearing, environmental pressure, and so forth, which can be incredibly enormous scope and tangled. In this way, powerful and efficient methods for weather data analysis are significant and critically required. Data mining technologies, which are the PC, helped process that uncovers valuable examples from underneath enormous scope data sets, is generally recognized as an extremely promising course in weather data analysis. In this paper, an original methodology is proposed approach for weather data mining. Such a technique is applied to weather data set gathered at JFK, MCO and SFO air terminal in 2016, and shows an extremely promising benefit in classification accuracy contrasted and other conventional methods.

#### Merits

It is convincingly shown that distance metric learning technique reduces the misclassification rate and the most elevated accuracy is accomplished.

#### Demerits

These viewpoints to the model won't without a doubt help further genuine application.

**11. Prashanthi B, Meganathan S, Krishnan RB, Varahasamy R, Swaminathan S.** (2016) et.al proposed Data Mining as a Tool for Hot Day Prediction during Summer Monsoon. Human health, energy, and solace are impacted more by climate than by some other component of the physical environment. Physiological elements of the human body react to changes in the weather with climate and the seasons. The emphasis of present work is to give the hot day temperature designs during April and May months

and concentrate its social perspectives utilizing association rule mining of data mining strategy. The data mining methodology in view of association rule mining for extricating connections among climate parameters over Cuddalore station was applied to separate the serious summer day (hot day) designs during late spring months. The proposed data mining methodology is more helpful to apply with threshold values. As confirmed in the results, the methodology is reasonable for monitoring and foreseeing the temperature days 48 hours to come. This technique vows to be a valuable one for tropical coastal stations. By expecting the outrageous summer temperature, the day to day practice will likewise be arranged ahead of time founded on human.

#### Merits

This method promises to be a useful one for tropical coastal stations.

#### Demerits

The extreme summer temperature not worked the day to day practice.

**12. M. N. Wibisono and A. S. Ahmad (2017)** et.al proposed Weather Forecasting Using Knowledge Growing System (KGS). The presence of valuable weather forecast information can help different parties in doing their exercises. In this study, weather forecasts were made utilizing another Artificial Intelligence technique called Knowledge Growing System. The Knowledge Growing System uses weather sign models as the premise of weather forecasts. This paper conveyed the consequence of Knowledge Growing System executed on a weather forecast. This study shows that Knowledge Growing System is fit for forecasting weather utilizing synoptic weather data. As the after-effect of this study, The A3S algorithm on the KGS can identify occasions in light of the hypothesis set. A3S Algorithm had the option to anticipate the event as indicated by the standards data set at the level of Accuracy 70.68%. The best situation of the study had the option to anticipate with an accuracy of 79%. This demonstrates that dynamic should be finished considering the expected OMA3S results and readings from A3S for ideal prediction results.

#### Merits

Knowledge Growing System is fit for forecasting weather utilizing synoptic weather data.

#### Demerits

Weather measures dynamically founded on data not obtained progressively.

**13. Wang Z, Mujib AM (2017)** et.al proposed The Weather Forecast Using Data Mining Research Based on Cloud Computing. Weather forecasting has been a significant application in meteorology and one of the most logically and innovatively testing issue all over the planet. This paper proposes a cutting edge strategy to foster a service arranged architecture for the weather information systems which forecast weather utilizing these data mining techniques. In this study they likewise adjusted the technique for Artificial Neural Networks; it can identify the relationships between the information variables and create outputs in light of the

noticed examples inborn in the data without any requirement for programming or creating complex conditions to model these relationships. Weather conditions are essential to climatic change studies on the grounds that the variety in weather conditions in term of temperature and wind speed can be concentrated on utilizing these data mining techniques. ANNs are carried out, to look at their effectiveness in changing the network topology and the training mode. The results acquired from genuine data depend on time series of meteorological data given by the Dalian Meteorological Bureau. The experiments brought up that the proposed approach gives an exceptionally interesting presentation of the executed network and shows great execution in term of MSE. For future point of view, there is as yet critical potential for development in weather forecasting by utilizing ANN model, through presenting climate change and global warming variables, to forecast more realistic weather parameters.

#### Merits

The experiments called attention to that the proposed approach gives an extremely interesting exhibition of the carried out network and shows great execution in term of MSE.

#### Demerits

There is as yet critical potential for development in weather forecasting by utilizing ANN model.

**14. Krishna Kumar Sharma, Ayan Seal (2019)** et.al proposed modeling uncertain data using Monte Carlo integration method for clustering. Nowadays, data clustering is a significant errand to the mining research local area since the accessibility of uncertain data is expanding quickly in numerous applications. In this work, proposed Monte Carlo integration based uncertain objects modeling technique. This work centers on clustering uncertain article in view of their conveyances utilizing two similitude gauges in particular, KL-uniqueness and J-dissimilarity. These two measures alongside the item rule and the larger part casting ballot techniques are converged with DBSCAN and k-means clustering algorithms to show the effectiveness of Jdivergence over KL-uniqueness. A heuristic algorithm is proposed to observe ideal span, which is one of the essential parameters of DBSCAN. Monte Carlo integration put together technique depends with respect to the estimate of b-spline function and the exhibition of the proposed strategy relies upon the legitimate selection of request of b-spline function. Presently, value 4 is considered for request of the b-spline function. The proposed strategy is contrasted and three state-of-the-art methods genuinely on three genuine and one engineered datasets. They might want to concentration to foster an algorithm for discovering the ideal request of the b-spline function theoretically. This work could be broadened further for uncertain data streams.

#### Merits

That a large portions of the cases the proposed strategy outperforms state-of-the-art methods genuinely.

#### Demerits

Not discovering the ideal request of the b-spline function theoretically.

**15. Torabi M, Hashemi S, Saybani MR, Shamshirband S, Mosavi A.** (2019) et.al proposed A Hybrid clustering and classification technique for forecasting short-term energy consumption. This paper presents a hybrid way to deal with anticipate the electric energy use of weather-sensitive loads. The presented technique utilizes the clustering perspective close by ANN and SVM approaches for accurate short-term prediction of electric energy use, utilizing weather data. Since the methodology being summoned in this research depends on CRISP data mining, data planning has gotten a lot of consideration in this research. The proposed approach (CBA-ANN-SVM) was applied to genuine load data and coming about higher accuracy contrasting with the current models. In this research they attempted to perceive every single efficient boundary and with the utilization of CBA-ANN-SVM model, the pace of error has been minimized. The last model has the advantages from the two models and the advantages of clustering. Clustering algorithm with finding data structure, isolates data into a few clusters in view of likenesses and contrasts between them. Since data inside each cluster are more comparable than whole data, modeling in each cluster will pre-sent better results.

### Merits

The proposed approach (CBA-ANN-SVM) was applied to genuine load data and coming about higher accuracy contrasting with the current models.

### Demerits

Forecasting models are and Challenges local they are explicitly not intended for a particular region and can't be applied globally.

### CONCLUSIONS

This paper presented a Systematic Mapping of survey about the world situation of research on the utilization of the Active Methodology to moderate dropout and inspires the perpetual quality Weather. Weather data analysis is a basic advance in weather forecast and environmental change prediction. The admittance to Systematic Mapping data sources is frequently for analyzing and mining of these information are more important to comprehend and further develop the Weather data analysis. With the rapidly growing data sets because of remote detecting and remote sensor advancements, data mining techniques are highly effective and proficient in weather data analysis.

### REFERENCES

[1]. Chammas M, Makhoul A, Demerjian J. An efficient data model for energy prediction using wireless sensors. *Computers & electrical engineering*. 2019 Jun 1;76:249-57.  
 [2]. Janković R, Čosović M, Amelio A. Time series prediction of air pollutants: a case study for Serbia, Bosnia and Herzegovina and Italy. In2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH) 2019 Mar 20 (pp. 1-6). IEEE.  
 [3]. Finamore A, Calderaro V, Galdi V, Piccolo A, Conio G. A day-ahead wind speed prediction based on meteorological

data and the seasonality of weather fronts. In2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia) 2019 Mar 19 (pp. 915-920). IEEE.  
 [4]. E. Doko and L. A. Bexheti, "A systematic mapping study of educational technologies based on educational data mining and learning analytics," 2018 7th Mediterranean Conference on Embedded Computing (MECO), 2018, pp. 1-4, doi: 10.1109/MECO.2018.8406052.  
 [5]. A. Alvarez-Ayllon, M. Palomo-Duarte and J. -M. Doderó, "Interactive Data Exploration of Distributed Raw Files: A Systematic Mapping Study," in IEEE Access, vol. 7, pp. 10691-10717, 2019, doi: 10.1109/ACCESS.2018.2882244.  
 [6]. L. Dong, B. Liu, Z. Li, O. Wu, M. A. Babar and B. Xue, "A Mapping Study on Mining Software Process," 2017 24th Asia-Pacific Software Engineering Conference (APSEC), 2017, pp. 51-60, doi: 10.1109/APSEC.2017.11.  
 [7]. Wang H, Gu C, Zhang X, Li F, Gu L. Identifying the correlation between ambient temperature and gas consumption in a local energy system. *CSEE Journal of power and energy systems*. 2018 Dec 31;4(4):479-86.  
 [8]. T. T. Shein and S. Puntheeranurak, "Incremental Clustering Approach for Evolving Trajectory Data Stream," 2018 International Electrical Engineering Congress (iEECON), 2018, pp. 1-4, doi: 10.1109/IEECON.2018.8712334.  
 [9]. Daraio E, Corso ED, Cerquitelli T, Chiusano S. Characterizing air-quality data through unsupervised analytics methods. InEuropean Conference on Advances in Databases and Information Systems 2018 Sep 2 (pp. 205-217). Springer, Cham.  
 [10]. Niu Z, Zhu Y, Jiang L. Distance metric learning approach for weather data mining. InProceedings of the 2018 2nd International Conference on Big Data and Internet of Things 2018 Oct 24 (pp. 33-36).  
 [11]. Prashanthi B, Meganathan S, Krishnan RB, Varahasamy R, Swaminathan S. Data mining as a tool for hot day prediction during summer monsoon. In2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT) 2016 Mar 18 (pp. 1-3). IEEE.  
 [12]. M. N. Wibisono and A. S. Ahmad, "Weather forecasting using Knowledge Growing System (KGS)," 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017, pp. 35-38, doi: 10.1109/ICITISEE.2017.8285526.  
 [13]. Wang Z, Mujib AM. The Weather Forecast Using Data Mining Research Based on Cloud Computing. InJournal of Physics: Conference Series 2017 Oct 1 (Vol. 910, No. 1, p. 012020). IOP Publishing.  
 [14]. Sharma KK, Seal A. Modeling uncertain data using Monte Carlo integration method for clustering. *Expert systems with applications*. 2019 Dec 15;137:100-16.  
 [15]. Torabi M, Hashemi S, Saybani MR, Shamshirband S, Mosavi A. A Hybrid clustering and classification technique for forecasting short-term energy consumption. *Environmental progress & sustainable energy*. 2019 Jan; 38 (1):66-76.