Literature Review on Weather prediction using Data Mining

Course: Data Mining

Professor: Dr. Tjortjis

Students: Efthymios Koutsimanis, Christos Galanis

Abstract

It is undeniable that extreme weather events can impact the environmental and social system. Data mining and Machine learning methods can mitigate the risks of extreme weather. That is why, in the context of Data mining course we conducted a literature review regarding weather prediction using data mining. In this paper, it is highlighted the importance of data mining and machine learning algorithms since they can accurately predict weather parameters more efficiently compare to other traditional weather prediction models.

Table of Contents

List of Figures

1. Extreme weather events

An increased number of extreme weather events have been observed around the globe in the last decade (Cho et al., (2018) referred by Saura, et al., (2023)). Natural and human factors change the earth's surface cover or land cover during time (Hassan, et al., (2016) highlighted by Kaya, et al., (2023)). Natural phenomena such as fires, pollution, tsunamis, earthquakes, floods and more are results from the continental movements (Breuste, et al., (2021) as pointed out by Kaya, et al., (2023) and high temperatures (Keles¸ et al., 2018; Shafi and Mohammadi, 2020; UNDRR, (2020) referred by Saura, et al., (2023)). Urbanization constitutes a major driver behind human land usage according to the IPCC session (Kaya, et al., 2023). This has effects on the climate. There is an increasing worrying about climate change globally (Nyasulu, et al., 2022). According to Stott (2016) and Chik and Xue (2021) extreme weather and global warming impact both the natural environment and social system (Saura, et al., 2023). In the meantime, Saura, et al., (2023) noted from the papers of Lee et al., (2014) and Huang et al., (2018) that despite the hot temperatures recorded on our planet, economic output and profitability are at high levels too.

Increasing natural disasters such as droughts, floods and more have negative consequences on flora and fauna leading to the extinction of species. Furthermore, economic production changes due to weather events (Choi et al., (2012) as described by Saura, et al., (2023)). In addition, they highlighted from the paper of Jian-Bin et al. (2012) that greenhouse gases emit in the air and on the earth's surface resulting in high temperatures. These extreme phenomena affect millions of people regarding their food and water insecurity (Schenker, M. (2000) pointed by Kaya, et al., (2023)). Simultaneously, the population of the planet rises over time meaning the natural resources are decreasing (Lam et al., (2016) as cited by Saura, et al., (2023)). Therefore, global warming that affects both animal and human beings need to be addressed with regards to rational management of natural resources and decrease of pollution and gases (Lam et al., (2016) as described by Saura, et al., (2023)). Thus, in order to take effective measures and minimize the risks of extreme natural phenomena, the prediction of weather events is deemed necessary (Saura, et al., 2023).

2. Weather forecasting

"Weather prediction is the scientific prediction of the state of atmospheric conditions such as temperature, humidity, dew point, rainfall and wind speed based on reliable data" (Jaseena & C. Kovoor , 2020, p. 3394). Another definition is the combination of science and technology in order to forecast atmospheric conditions for a specific location at a specific time (AbdulRaheem, et al., 2022). Since 19th century people were trying to predict the weather formally (C. A. Anton et al., (2019) and B. Das et al., (2018) reported by AbdulRaheem, et al., (2022)). Weather data collection can be achieved by radars, barometers and thermometers. Information regarding previous weather patterns, current weather conditions, air pressure and motion of air and clouds in the sky is essential for weather prediction (Jaseena & C. Kovoor , 2020). According to Purwandari, et al., (2020) nowadays people can be informed of weather conditions about four-five times a day by exploiting devices such as mobile phones, newspapers, TV and more. Furthermore, they pointed out in their paper that having accurate weather information affects human activities and this results in economic and social development. Weather knowledge can on the one hand keep people safe from extreme weather events such as floods, tornadoes, storms and more. On the other hand, it has a high impact in a variety of sectors like Sports, Tourism, Agriculture, Business, Food

industry and more. Having prior knowledge of weather, for instance wind conditions can help farmers take the right decisions for their crop fields. Therefore, it is vital to have accurate predictions on a hourly, daily, weekly, monthly or yearly basis since this can provide environmental information either for the climate change or for business purposes (Jaseena & C. Kovoor , 2020).

3. Increasing in technologies and Big Data

In some cases, weather forecasts are not reliable since the climate models do not have the ability to replicate sub-daily precipitation and there are not sufficient predictions daily. This results in setting precipitation patterns daily according to Kendon, et.all, (2014) as cited by Purwandari, et al., (2020). Jaseena & C. Kovoor, (2020) referred that due to the fast evolution of technologies such as Wireless Sensor Network, Internet of Things and Cloud Computing weather information has entered the era of Big Data. Enormous amount of weather data is generated now through the fast increase in radars, satellites, weather balloons and more (Saha Roy, 2020). In this context, people use weather information models precisely and in real time (Purwandari, et al., 2020). This helps to predict future climate patterns more accurately.

Being in the 21st century and with the evolution of Big Data there is a raise of scientific interest in new methods alongside with supercomputers having Graphics Processing Units (GPU) (Bochenek & Ustrnul, 2022). Due to the development of climate observing systems such as satellites meteorological systems the weather data have been exploded in terms of volume and variety. Thus, in order to make accurate predictions one can use deep learning techniques since the traditional computational weather-based systems are not adequate to handle these kind of Big Data according to Jaseena & C. Kovoor , (2020). New technologies like deep learning and data visualization methods can result in more effective and precise weather and climate predictions. In addition, many sectors are using deep learning and the motivation of using this kind of technique to predict the weather has started as mentioned by Jaseena & C. Kovoor , (2020).

However, there are some challenges that comes along with the wide availability of weather data specifically in the prediction context such as identifying the relationship between weather features and creating a powerful forecasting model to utilize the underlying structure hidden in the datasets according to Saha Roy, (2020). Also, in his paper referred to the importance of using enormous datasets involving a variety of weather features such as humidity, temperature, air pressure and more since they can support accurate predictions. This results in learning how the temperature changes over time.

4. Emerging new methods

This was achieved conventionally by the numerical weather prediction (NWP) models which are computer models using physical equations (Marchuk (2012) reproduced by Saha Roy, (2020)). In contrast, Saha Roy, (2020) noted from Pathak et al., (2018) and Reichstein et al., (2019) that these physics-based systems need structured prior information and high computational capacity. A reference needs to be made here with regards to prediction technology which currently uses simulations focusing on physical methods of the atmosphere (Krasnopolsky et al., (2006), Marchuk (2012) and Richardson (2007) mentioned by Saha Roy, (2020)).

The work of Pauzi, et al., (2020) as mentioned by Kaya, et al., (2023) highlighted the importance of machine learning algorithms which are better than the traditional statistical methods for predicting the weather. These models, especially deep learning as recognized by Saha Roy, (2020) do not need prior information in order to identify the underlying structure and patterns in the dataset. That is, they are suitable for weather forecasting instead of using physical models. Furthermore, Jaseena & C. Kovoor, (2020) stated that deep learning can support weather forecast by identifying and finding useful patterns from weather data. Especially, neural networks with deep architectures can find useful features of Big Data efficiently (Gheisari et al., (2017) as referred by Jaseena & C. Kovoor , (2020)).

5. Importance of Data Mining

A notion here to be made clear is the importance of data mining in weather prediction and climate change. It is used for research in order to extract helpful information from huge datasets (H. A. Issad et al., (2019) and K. Soomro et al., (2019) as reproduced by AbdulRaheem, et al., (2022)). Studies in weather prediction and climate change with the support of data mining are important since people can access weather understanding and usability and therefore plan weather-related activities. In addition, both Z. Ge et al., (2017) and A. M. Abdu et al., (2020) as cited by AbdulRaheem, et al., (2022) referred to the fact that data mining is also called as knowledge discovery since it analyzes datasets from a variety of perspectives and then makes a synopsis of the important information. Data mining has a variety of applications that range from blocking spam emails to predicting if tomorrow will rain. That is, it detects the repetitive patterns and tries to make forecasts for the future (AbdulRaheem, et al., 2022). Nevertheless, as it was stated previously, weather predictions are in some cases inaccurate, not reliable and complex. This is due to the methods that are used. In order this obstacle to be addressed, an accurate prediction needs to have the right weather variables, a suitable data mining model and the appropriate equipment to run the process (AbdulRaheem, et al., 2022). As A. Joshi et al., (2015) have stated, throughout the world, one of the most difficult scientific and technological issues of the previous century is weather prediction (AbdulRaheem, et al., 2022).

6. Literature review process

With that in mind, in the context of data mining course we have been assigned to do a literature review regarding weather prediction using data mining. Specifically, we have searched in 3 online databases used for academic research. That is, ProQuest, ScienceDirect and Scopus. In order to find what we were looking for; we used the following search strings:

(weather prediction or weather forecasting) and (data mining)

The above search strings did not provide us with a suitable number of appropriate articles regarding data mining applications in weather forecast. Therefore, we altered the search strings and we included one more term with the hope to result in more appropriate articles given as below:

(weather prediction or weather forecasting) and (data mining or machine learning)

The term “machine learning” which is more or less the same as data mining term helped us to find more articles for our goal. The results were useful because more articles included in their title or abstract the

term “machine learning” than the phrase “data mining”. We believe this was done since the first term is more widespread nowadays and it is utilized by many people.

Throughout the process of filtering, we used the following criteria:

- Full text / Open access
- Peer Reviewed
- Last 5 years
- Review, Scientific and Journal articles

Subsequently, after screening the most relevant remaining articles by their title we briefly evaluated them based on their abstract. Finally, the most suitable articles for our research objective were kept. That is to write a literature review on weather prediction using data mining.

7. Extreme weather repercussions on global economy

Before we proceed, to how data mining and machine learning can help people understand and predict the weather we need once again to comprehend that weather forecasting is vital for human beings. Saura, et al., (2023) presented in their paper an established data driven model that recognizes the consequences of extreme weather events on economic products. This was applied using User-Generated Content (UGC) obtained from Twitter source. They identified 7 topics associated with 3 sentiments as shown below:

- Positive (Sustainable energies, Green Entrepreneurs)
- Neutral (Climate economy, Producer’s productivity, Stock market)
- Negative (Economy and policy, Climate emergence)

First of all, in the recent years the combination of extreme weather events and meteorological uncertainty has led to negative output for economies of the world and the business sector. They found that there is a new fast growing climate economy that is associated with economic and political decision driven by public institutions. Its goal is to prevent climate change, global warming and the detriment left by extreme weather events. In this stage, they pointed out the negative opinion that the public has with regards to policies implemented so far aiming to counteract the meteorological issues. Furthermore, extreme weather events with global warming influence economic production meaning the growth in supply chains, logistics and / or production. As a result, the stock market is in an uncertainty mode leading to unstable economy. Lastly, taking this into consideration, sustainability and renewable strategies are the main factors to address extreme weather events. In addition, sustainable actions taken by green entrepreneurs is perceived as a leading choice for the evolvement of economy and industry. This results in more flexible management when dealing with weather emergencies around the planet (Saura, et al., 2023).

8. Proposed data preprocessing steps

The main objective of the weather prediction models is the forecast of temperature, dew point, humidity, rainfall and wind speed as referred by Jaseena & C. Kovoor , (2020). Furthermore, they have reviewed and classified weather forecasting models from a variety of papers. Also, in their research is presented the weather prediction framework. This is a variety of steps including data collection, data preprocessing, model selection and training, model evaluation and visualization of the results. These steps are done since

the data are obtained from a variety of sources and they have different formats such as satellites, internet of thing, Wireless Sensor networks and cloud computing. Thus, the data contained are both useful and not. Jaseena & C. Kovoor , (2020) presented the steps as follows:

After the data acquisition from a various resources a preprocess step is applied. That is, data cleaning, data integration, data reduction and data transformation. Regarding data cleaning techniques if there is noise is removed, missing values or inconsistencies. Most of the real world data come with missing values. Therefore, there are techniques to deal with these problems such as imputing the mean of the features with missing values. Data cleaning methods are important techniques since they determine the quality of the data and then these cleaned data are passed as input for the next steps.

When data come from different sources we need to merge all the data together. In order to reduce the size of the data there are methods for reduction of the features such as clustering and more. The next step is to transform the data. That is, to normalize or standardize the data. With these tricks not only the accuracy of the model increases but the training time of the model decreases. After that, the data is ready to be trained and tested by an appropriate model.

Model selection and training is a vital part in any prediction system. There are many parameters where they determine which model to choose from the variety of the existing ones. The user needs to have knowledge of the models regarding their advantages and shortcomings. Furthermore, domain knowledge is important for selecting the suitable algorithm. Subsequently, after the model training the next stage involves the model performance evaluation. This is done using statistical quantitative error indicators for example MAE, RMSE, MAPE and R^2. The results are visualized using appropriate plots such as scatter plots, line plots and semilog plots. All plots depict the difference between the actual and predicted value. However, as Jaseena & C. Kovoor , (2020) highlighted that the last kind of plot referred previously is more suitable for observing the variation of the actual and predicted values. Below is illustrated the blueprint of the weather forecasting algorithm according to Jaseena & C. Kovoor , (2020):

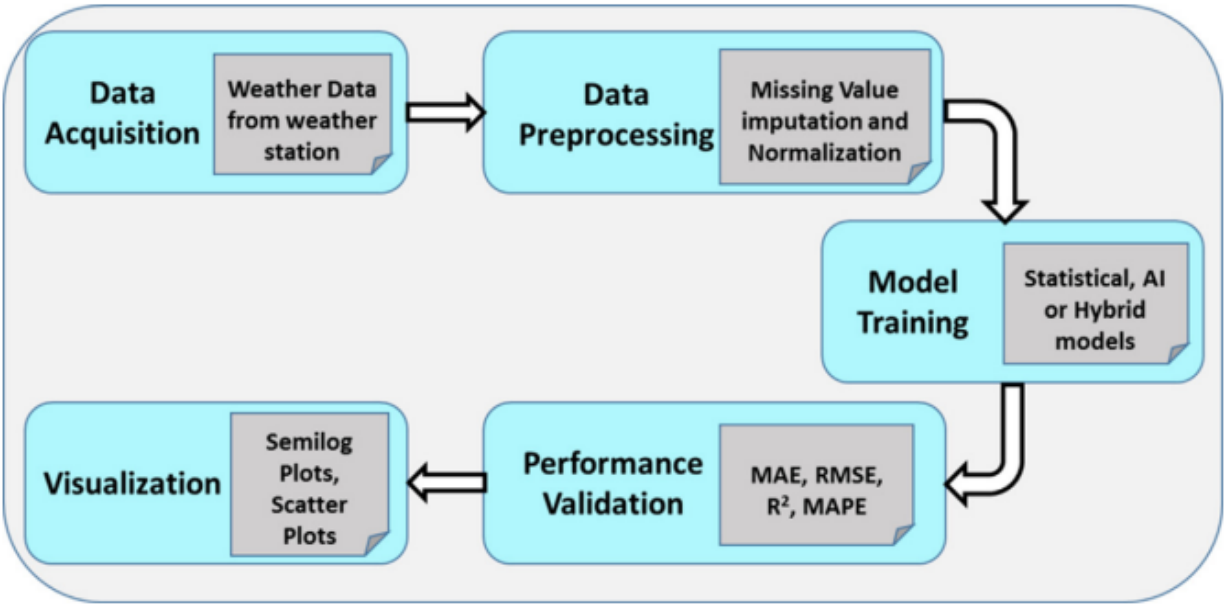

Figure 1: Blueprint of weather prediction algorithm (Jaseena & C. Kovoor , 2020)

9. Categorization of weather prediction models

The weather forecasting models are classified based on multiple criteria as mentioned by Jaseena & C. Kovoor , (2020):

Number of variables: prediction models can be classified into univariate or multivariate. The first term implies that the model predicts the output based on a single variable whereas the second term means that the model considers a variety of features

Number of time steps forecasted: the forecasting models can be single step or multi step where as the name implies the single-step predicts a single future value and the multi-step forecasts a sequence of future values

Methodology used: the models can be two types. The one is deterministic where it predicts values for a specific location while the other is probabilistic where it provides probabilities of weather events. The deterministic methods are classified as statistical models, Artificial intelligence models and hybrid models

Prediction horizon: forecasting models can be categorized into 4 scales meaning very short term, short term, medium term and long term. The very short term and short term are more accurate than the two last range scales

Predicted parameter: the models are categorized as temperature, wind speed, dew point, rainfall prediction models and more

Below is illustrated a diagram of the classification of the forecasting algorithms:

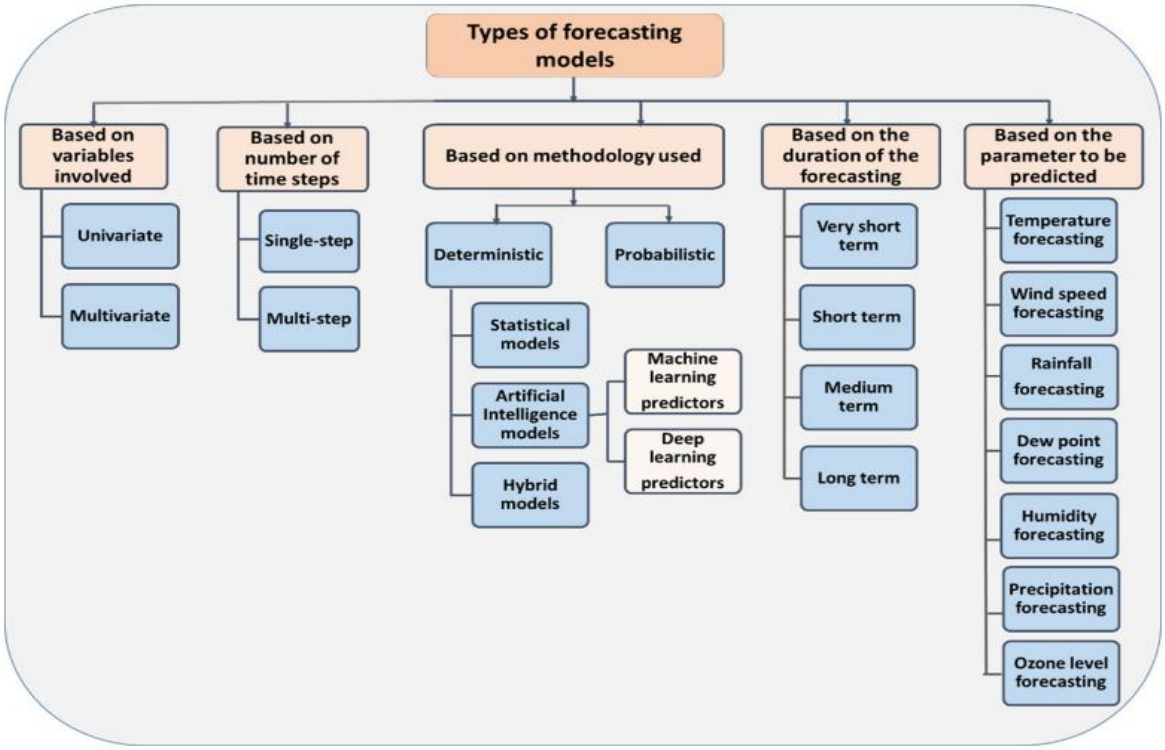

Figure 2: Classification of prediction algorithms (Jaseena & C. Kovoor , 2020)

10. Synopsis of weather prediction articles

Results from the research of Jaseena & C. Kovoor , (2020)

After reviewing all the relevant articles gathered in their research, Jaseena & C. Kovoor , (2020) made some useful conclusions. First of all, they highlighted the importance of hybrid models as more reliable than other types for weather prediction. In addition, a high proportion of wind speed prediction models are univariate since they forecast the output using only one variable. The following figures are borrowed from the paper of Jaseena & C. Kovoor , (2020). It can be observed that most of the forecasting models found in their literature are predicting wind speed. The second parameter that these models predict is the temperature. Furthermore, the most used quantitative metric is RMSE with a proportion of 30%. Jaseena & C. Kovoor , (2020) noted that RMSE do not take the absolute values and therefore gradient and distance can be calculated effectively.

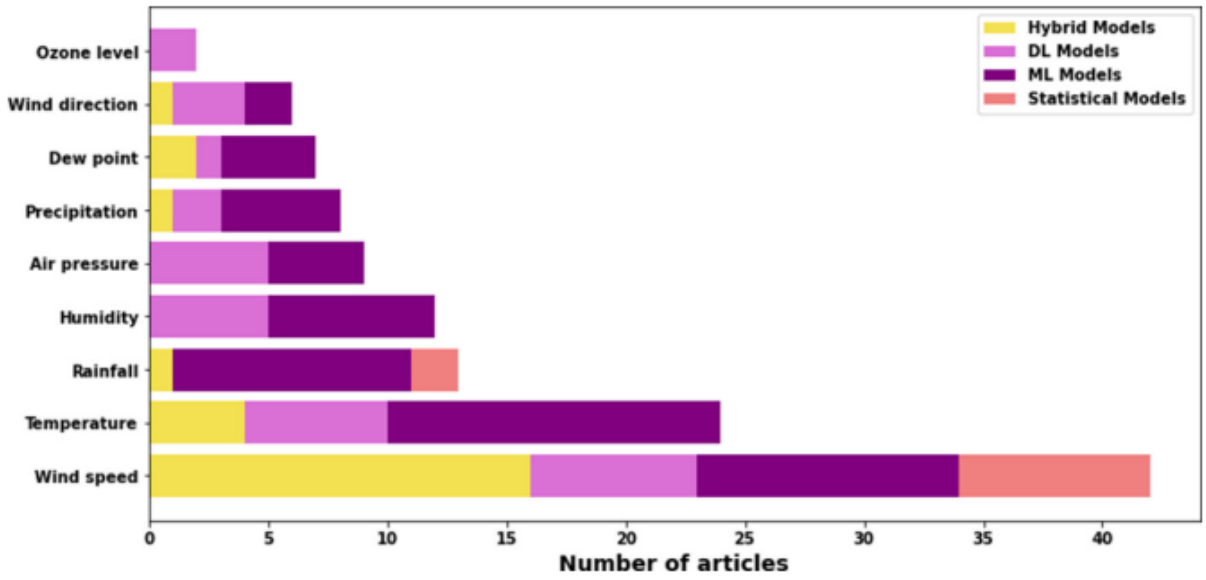

Figure 3: Distribution of prediction models based on parameters (Jaseena & C. Kovoor , 2020)

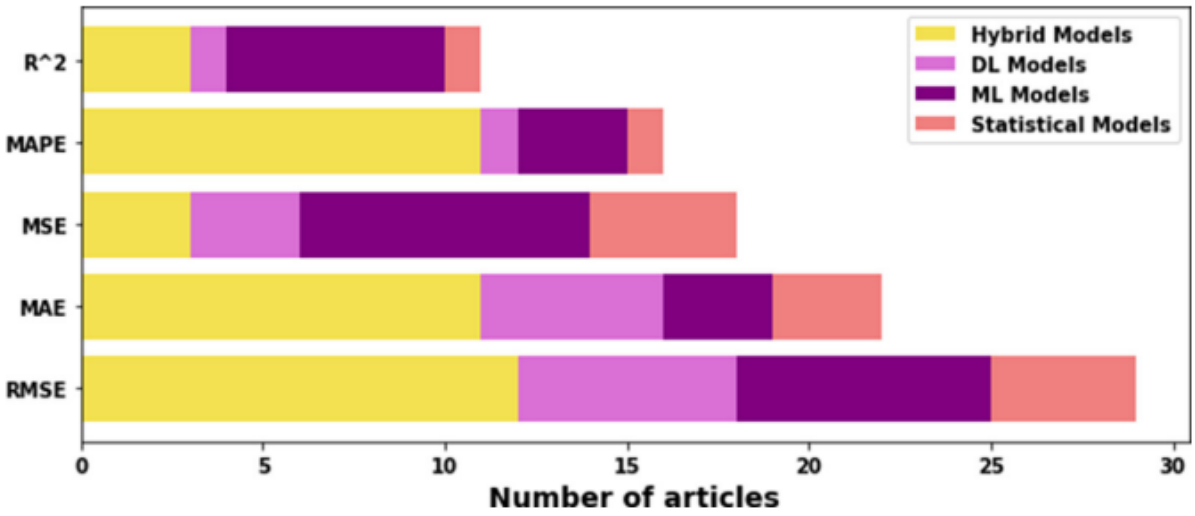

Figure 4: Distribution of prediction models based on the quantitative indicator used (Jaseena & C. Kovoor , 2020)

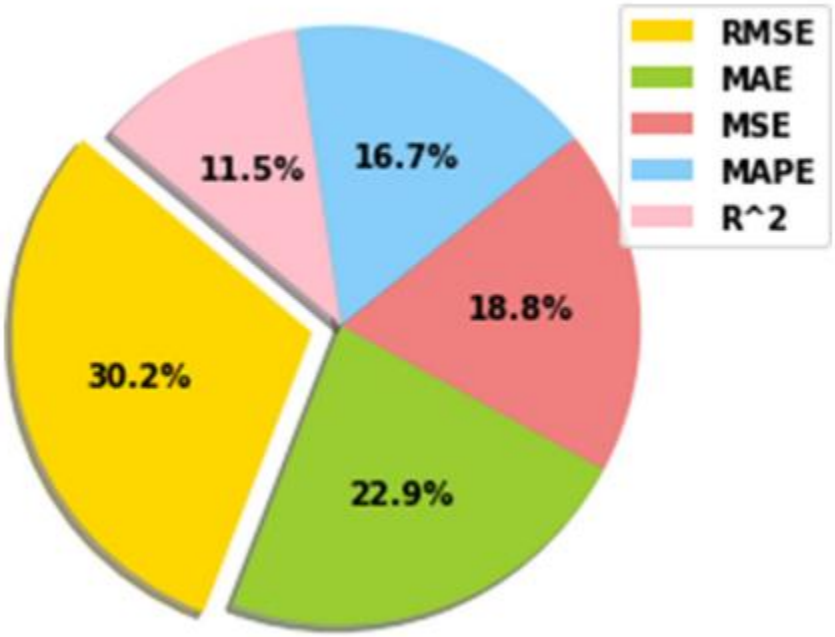

Figure 5: Proportion of the quantitative indicators used (Jaseena & C. Kovoor , 2020)

Results from the research of Bochenek & Ustrnul, (2022)

The paper of Bochenek & Ustrnul, (2022) presented a review of the article that used machine leaning and artificial intelligence approaches in meteorology and climatology. Specifically, they searched in Google Scholar and found 500 articles published since 2018. The first 250 papers included the combination of phrases “numerical weather prediction” and “machine learning” and the second 250 articles had the phrases “climate” and “machine learning”.

The figure bellow depicts the most common phrases found in the articles associated to numerical weather prediction and machine learning. The phrase that was mostly used by the scientists is “Wind Forecasting”.

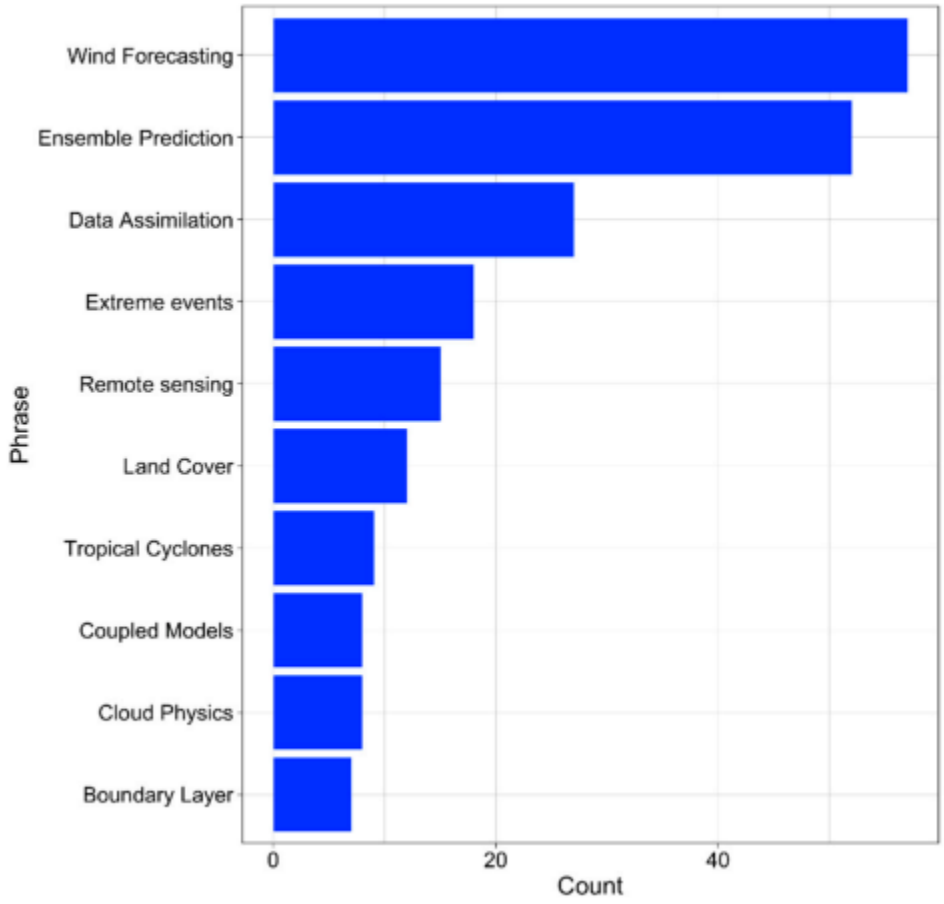

Figure 6: Count of common phrases from papers in relation to the numerical weather prediction and machine learning (Bochenek & Ustrnul, 2022)

The second group of acquired articles related to “climate change” and “machine learning” showed that the most common phrase was unsurprisingly “Climate Change”. These results can be found below:

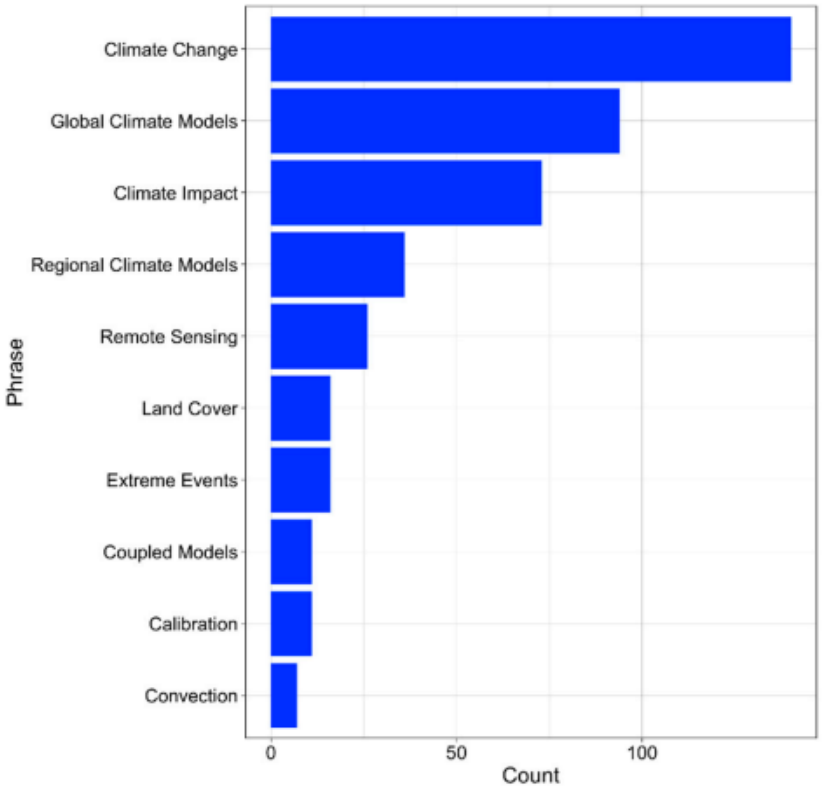

Figure 7: Count of common phrases from papers in relation to the climate change and machine learning (Bochenek & Ustrnul, 2022)

Some other important results found by the Bochenek & Ustrnul, (2022) was the most common meteorological parameter that scientists are focusing. This parameter is the wind which was mentioned more than 200 times in the papers. In addition, they included the most common methods that the scientists are using when applying machine learning in NWP researchers. That is, deep learning is the dominant method. Lastly, it is important to learn what are the most common countries that appeared in papers regarding climate studies and machine learning. Bochenek & Ustrnul, (2022) highlighted that only 25% of the articles referred a specific geographical region in their abstracts. That is, 62 articles from the 250 obtained with relation to the second group. It is illustrated that climate studies with machine learning applications are mostly focused on the country of China. All the above results, can be seen in the figures below:

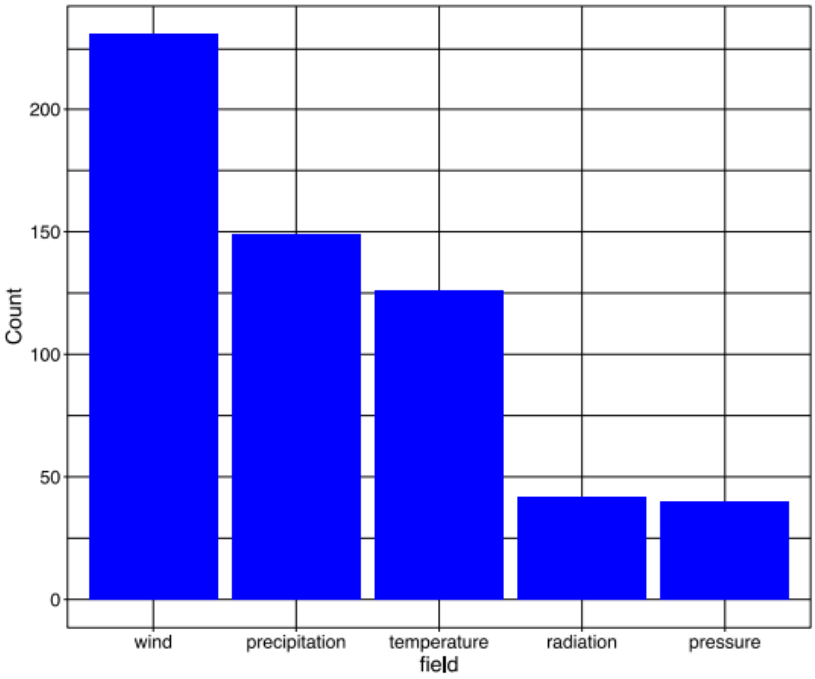

Figure 8: Most common prediction parameter in NWP articles (Bochenek & Ustrnul, 2022)

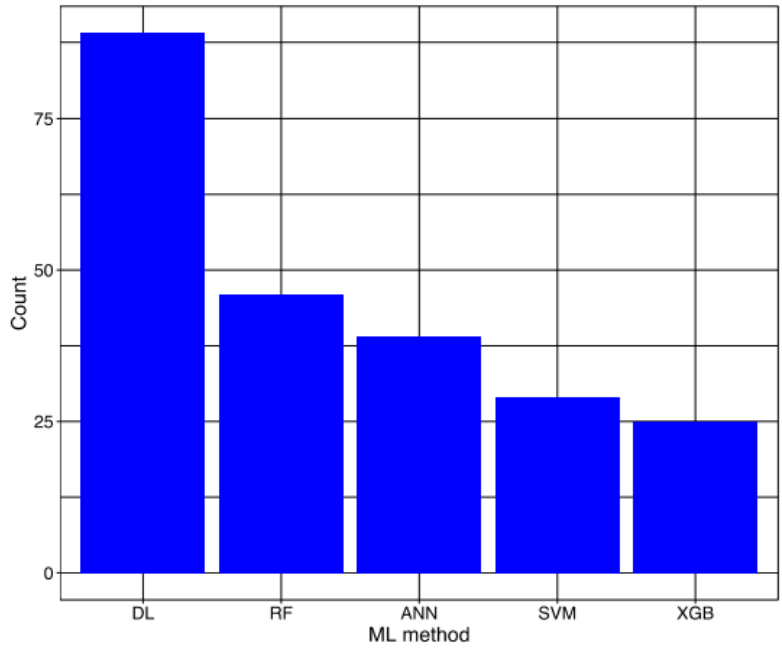

Figure 9: Most common algorithms used in NWP articles (DL – Deep Learning, RF – Random Forest, ANN – Artificial Neural Networks, SVM – Support Vector Machines, XGB – XGBoost) (Bochenek & Ustrnul, 2022)

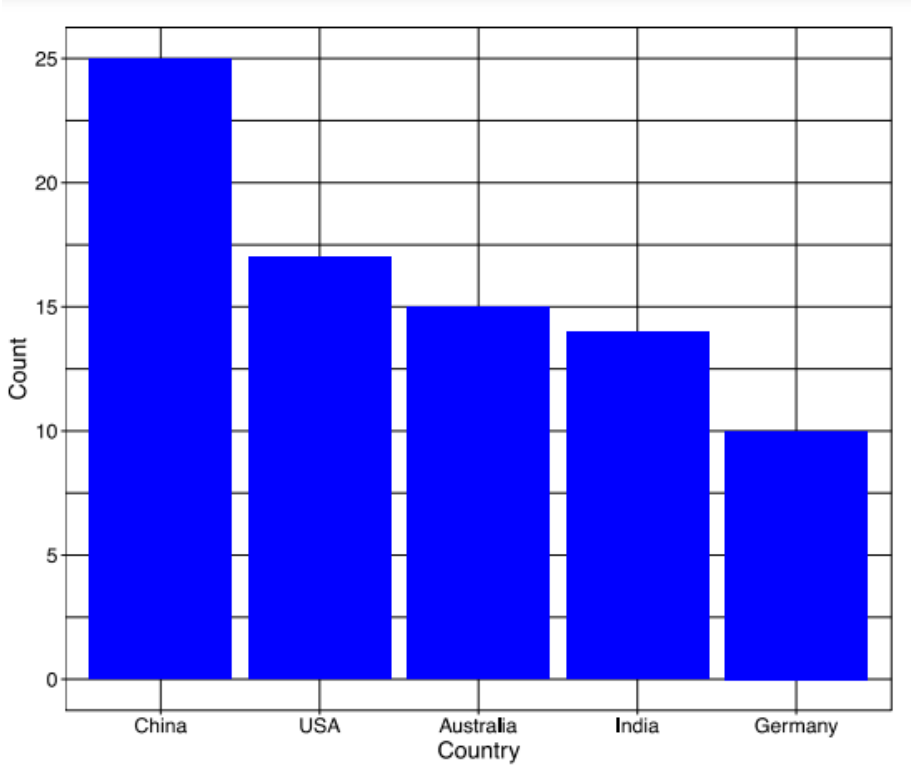

Figure 10: Most common countries found in the articles' abstract in climate papers (Bochenek & Ustrnul, 2022)

11. Internet of Things

Nowadays, with the rapid increase of technology the number of Internet of Things (IoT) devices is rising too. Data collected from small devices operating as weather stations on the rooftops of homes, mobile phones that periodically transmit environmental data, larger stations and government sources collecting large amount of weather data are some examples of obtaining data from multiple sources. These sensors capture a variety of data such as temperature, wind speed, wind-direction, air-pressure, humidity and more. However, according to Montori et al., (2017) as mentioned by Behanan, et al., (2021) there is a phenomenon called "IoT islands" where there are multiple IoT devices. These devices are built to operate for specific networks. However, when it comes to operating on different networks some problems appear. This results in collecting redundant and/ or duplicate data from the sensors from multiple networks (Behanan, et al., 2021). The objective is to specify the nature of the data and pass them for extra processing. Usually, the data obtained do not follow a uniform system of labeling, therefore it is not easy to process them. To overcome this problem, Behanan, et al., (2021) proposed the concept of Mobile Crowd Sensing (MCS) which stores the data coming from the sensors to a cloud-based server periodically. This is vital for the accuracy of ML models since they require large datasets.

In this context, S Patil & Vidyavathi, (2022) analyzed a dataset with 2.3 million records collected from 54 Wireless Sensor Networks from the Intel Berkeley Research lab. The sensors recorded time-stamped information. Furthermore, the dataset had the following format: date, time, epoch, mote ID, temperature,

humidity, light and voltage. After the application of some preprocessing techniques by the researchers, they used multiple linear regression (MLR) algorithm with the support of some key features such as temperature, humidity, light and voltage. This method is a statistical approach which considers multiple independent features with the aim to predict the dependent one which is the output. The model was evaluated by a variety of indicators such as R^2, MAE and more. The resulted model as it is stated it has high accuracy and it is noted that it is statistically more significant than other existing approaches (S Patil & Vidyavathi, 2022).

12. Data mining and Machine learning applications on weather prediction

The paper of Saha Roy, (2020) utilized deep learning techniques in order to predict the average air temperature using data for a duration of 10 years from John F. Kennedy International Airport, NY. As the researcher highlighted this airport was selected since it is the busiest international air passenger gateway in North America. Therefore, making accurate predictions can support a smooth air traffic. The features of the dataset included average wind speed, snowfall, snow depth, precipitation, average temperature, maximum temperature and minimum temperature. The three deep learning models used were Multi-Layer Perceptron (MLP), Long Short Term Memory Network (LSTM) and a combination of Convolutional Neural Network (CNN) and LSTM. Furthermore, the evaluation metrics applied were Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The results showed that the performance combination of the CNN and LSTM model was better than the other two models (Saha Roy, 2020).

Another recent paper regarding weather prediction was made by AbdulRaheem, et al., (2022). These researchers obtained data from Kaggle regarding the year 1978 until 2017 from Seattle-Tacoma International Airport. The dataset consisted of approximately 25.500 records and 5 features namely day, month, prcp (precipitation for that day), tmax (max temperature for that day), tmin (min temperature for that day) and the target variable which was rain (true or false). Three models were utilized: a decision tree (DT), k-nearest neighbor (k-NN) and logistic regression (LR). The decision tree model outperforms the other two classifiers since it performed with accuracy, precision, recall and F1-Score a solid 100%. After that, comes the LR algorithm and then k-NN method.

Sethupathi.M, et al., (2021) pointed out the importance of machine learning techniques when it comes to analyzing hidden patterns from large datasets. The conventional models fail to operate when they are fed with massive data. That is why these researchers applied the random forest and logistic regression algorithms for a classification task. The dataset used included information from many regions within India from 2015 to 2018. The features included were data, location, min temperature, max temperature and more. The records were 966. The models were evaluated on accuracy meaning if it will rain tomorrow or not. Specifically, the algorithm predicted better the no-downpour class and no so well the downpour class. As they highlighted in their paper, this was probably due to some missed qualities, lower precipitation in the areas or limitation of climatic features in the dataset. Overall, both algorithms were very efficient (Sethupathi.M, et al., 2021).

Very interesting research was done for the Ethiopian Dodota Woreda region from Shekana, et al., (2020). This region was selected since there is high weather variability in this area and this results in frequent occurrence of repeated droughts. The dataset was obtained from the national meteorology agency of the

Dodota Woreda area and AwashMelkasa station. There were 7 features and 5282 records. The instances were recorded for a duration of 10 years from 2006 to 2016. The researchers applied data mining tools such as Multi-Layer perceptron (MLP), Naïve Bayes and multinomial logistic regression models. The results were evaluated based on precision and recall and using roc graphs. In the end, they found that MLP model was the most appropriate algorithm for classification and weather variability prediction between the selected ones (Shekana, et al., 2020).

A different paper was introduced by Purwandari, et al., (2020) where they use text mining by employing Support Vector Machine (SVM), MultinomialNaive Bayes (MNB) and Logistic Regression (LR) models. With the support of Twitter, they extracted tweets containing keywords such as sunny, cloudy, rainy, heavy rain, and thunderstorms. The data was collected from 1 January to 31 January 2020 focusing on the Indonesian region. The instances obtained were a total of 412 and were classified manually by a BMKG expert judgment. Using the evaluation metrics accuracy, precision, recall and f-measure they concluded that the SVM model outperformed the other methods. They highlighted the importance of SVM as it is very effective in classifying text. In addition, this algorithm provides speed even with high dimensional data and it is a great choice for the information retrieval problems (Purwandari, et al., 2020).

Agricultural activities are impacted by rainfall quantities and distribution. Therefore, climate change affects global agricultural productivity and food security especially in Africa which is less resilient to climate change. In order to tackle this problem and bring food security and economic progress in the region of Africa, Nyasulu, et al., (2022) employed ten ML Regressors and their Ensemble Model for daily rainfall, relative humidity, minimum and maximum temperature prediction in Senegal. They obtained a dataset from the NASA Langley Research Center's POWER Project regarding ten selected stations from South, Center and North of Senegal with a duration from 1982 to 2020. After the application of the models, they concluded that the Ensemble Model outperformed the Regressors. As quantitative indicators they used MAE, MSE, RMSE and R^2. They highlighted the importance of machine learning models in weather forecasting with regards to spatial distribution and annual cycle.

13. Necessity of forecasting Tropical Cyclones

Tropical Cyclones (TC) have always been in the interest of researchers and there has been many studies regarding forecasting, dynamic techniques and more the last 100 years (Chen, et al., 2020). These are extreme weather events that can cause catastrophes in coastal regions around the world. Nevertheless, despite the research made so far, there are problems with the predictive skills in TC genesis, intensity and risk forecasts. According to Ma, L.-M. (2014) as mentioned by Chen, et al., (2020) the most well-known tropical cyclone dynamical prediction models has resulted in low accuracy. As they highlighted, this is due to incomplete representation of complex procedures, inaccurate vortex initialization of TCs and more. To overcome this issue and improve the accuracy of the Tropical Cyclone models, researchers started using Machine Leaning methods to investigate satellite, radar etc. data. The paper of Chen, et al., (2020) explored both the challenges of tropical cyclones predictions and successful situations of machine learning regarding this topic. Furthermore, they described the progress done so far on machine learning in genesis, track, intensity and extreme weather predictions and storm surge forecasts and how to improve the performance of numerical forecast methods. Finally, they pointed out that there are many opportunities to investigate when it comes to Tropical Cyclones research such as analyzing and predicting techniques based on multi-source data particularly real-time data from satellites and more. Also, there are challenges

since tropical cyclones have complex dynamic mechanisms and are affected often by many factors. That makes them different from other conventional weather phenomena and oceanographic procedures. Therefore, the predictable period and stability of tropical cyclone forecast is difficult.

14. Conclusion

In summary, it is undeniable that nowadays extreme weather events have increased. As we have seen earlier, these events have a direct influence in a variety of sectors such as tourism, agricultural, naval and more. Predicting accurately the weather can result in increasing the economy production and the prosperity of human beings. In this literature review, we identified the importance of using data mining and machine learning tools in order to forecast weather parameters. Furthermore, we specified how IoT and new technologies can enhance the accuracy of the predicting models. In addition, relevant works regarding weather prediction using machine learning tools were referred. Finally, a reference made to the prediction of Tropical Cyclones using machine learning techniques and subsequently noted the difference between these phenomena and the conventional weather events.

All in all, from the literature obtained for the purpose of this course we can summarize that there is a smooth transition from conventional weather prediction models to new algorithms such as machine learning and data mining. These new approaches support the processing of enormous, generated by IoT technologies, datasets nowadays.

Bibliography

AbdulRaheem, M. et al., 2022. Weather prediction performance evaluation on selected machine learning algorithms.

Behanan, N., Ramesh, R. & A, V., 2021. Internet of Things and Machine Learning for Weather Prediction.

Bochenek, B. & Ustrnul, Z., 2022. Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives.

Chen, R., Zhang, W. & Wang, X., 2020. Machine Learning in Tropical Cyclone Forecast Modeling: A Review.

Jaseena, K. & C. Kovoor , B., 2020. Deterministic weather forecasting models based on intelligent predictors: A survey.

Kaya, S. M. et al., 2023. An Intelligent Anomaly Detection Approach for Accurate and Reliable Weather Forecasting at IoT Edges: A Case Study.

Nyasulu, C. et al., 2022. Towards Resilient Agriculture to Hostile Climate Change in the Sahel Region: A Case Study of Machine Learning-Based Weather Prediction in Senegal.

Purwandari, K., Sigalingging, J. W. C., Cenggoro, T. W. & Pardamean, B., 2020. Multi-class Weather Forecasting from Twitter Using Machine Learning Aprroaches.

S Patil, S. & Vidyavathi, B., 2022. A Machine Learning Approach to Weather Prediction in Wireless Sensor Networks.

Saha Roy, D., 2020. Forecasting The Air Temperature at a Weather Station Using Deep Neural Networks.

Saura, J. R., Ribeiro-Navarrete, S., Palacios-Marques, D. & Mardani, A., 2023. Impact of extreme weather in production economics: Extracting evidence from user-generated content.

Sethupathi.M, G., Ganesh , Y. S. & Ali, M. M., 2021. Efficient Rainfall Prediction and Analysis using Machine Learning Techniques.

Shekana, S., Mulugeta, A. & Sharma, D. P., 2020. Weather Variability Forecasting Model through Data Mining Techniques.