



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 6, Issue 5, May 2023



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.54



6381 907 438



6381 907 438



ijmrset@gmail.com



www.ijmrset.com



A Review on Seattle Weather Prediction for Climate Change

Prof.Pranali Vhora¹, Prof. Veena Badgujar²

Assistant Professor, Shah and Anchor Kutchhi Engineering College, Mumbai, India¹

Assistant Professor, Shah and Anchor Kutchhi Engineering College, Mumbai, India²

ABSTRACT: Weather prediction systems are an essential tool for many industries, including agriculture, aviation, transportation, and emergency management. For farmers the system can assist in optimizing crop management and irrigation schedules based on predicted rainfall patterns and temperature fluctuations. This information enables them to plan planting, harvesting, and other agricultural activities more effectively, ultimately increasing crop yields and reducing losses. The aviation industry heavily relies on weather prediction systems to ensure safe and efficient flights. Pilots can access real-time weather updates, including wind patterns, precipitation, and visibility, which help them plan flight routes, estimate fuel consumption, and avoid turbulent areas. Airports can also utilize this information to schedule operations, such as runway maintenance and deicing activities, to minimize disruptions caused by adverse weather conditions. Transportation companies, including trucking and logistics firms, can benefit from weather prediction systems to optimize routes, manage delivery schedules, and mitigate risks associated with severe weather events. By anticipating weather conditions along their routes, they can make adjustments in real-time, ensuring the timely and safe transportation of goods. Emergency management agencies heavily rely on accurate weather forecasts to prepare for and respond to natural disasters. Whether it's a hurricane, severe thunderstorm, or heavy snowfall, having precise information about the expected weather conditions allows these agencies to issue timely warnings, plan evacuations, and allocate resources effectively. This proactive approach can save lives and minimize damage caused by extreme weather events.

KEYWORDS: Datasets, Prediction Model, kaggle, Naïve Bayes, radar, ground-based sensors, and satellites.

I. INTRODUCTION

Data mining has improved organizational decision-making through insightful data analyses. Weather conditions can benefit from these accurate weather forecasts. The Seattle Weather Prediction System utilizes data mining techniques to extract valuable insights from a vast amount of weather-related data. This includes historical weather data, real-time weather observations, satellite imagery, and other relevant sources. The data is processed and analyzed using machine learning algorithms and statistical techniques to uncover hidden patterns, trends, and relationships. By applying data mining techniques, the system can identify recurring weather patterns, understand the impact of different variables on weather conditions, and make predictions based on historical data and current observations. These predictions can then be used by weather forecasters, businesses, and residents to make informed decisions and plan their activities accordingly. For example, businesses in industries such as agriculture, construction, transportation, and outdoor events can utilize the accurate weather forecasts to optimize their operations. Farmers can plan their planting and harvesting schedules based on predicted weather conditions, construction companies can schedule outdoor work efficiently, transportation companies can anticipate weather-related disruptions, and event organizers can plan outdoor activities or make contingency plans based on weather forecasts. Moreover, residents can benefit from accurate weather forecasts to plan their daily activities, such as outdoor recreation, commuting, and personal events. Knowing the expected weather conditions can help individuals make appropriate clothing choices, prepare for potential weather-related hazards, or adjust their plans to avoid unfavorable weather conditions. Overall, the application of data mining techniques in weather forecasting enhances the reliability and usefulness of weather predictions. It enables forecasters, businesses, and individuals to make informed decisions based on data-driven insights, leading to improved efficiency, safety, and overall decision-making in various sectors affected by weather conditions. Individuals can also leverage weather prediction systems to plan outdoor activities and make travel arrangements. By accessing weather forecasts specific to the Seattle region, they can decide when to go hiking, have a picnic, or participate in other recreational activities. Additionally, travelers can plan their trips based on anticipated weather conditions, ensuring a more enjoyable and



hassle-free experience. In summary, the Seattle Weather Prediction System plays a crucial role in various industries and for individuals by providing accurate and timely meteorological data. Its applications range from optimizing agricultural practices and enhancing aviation safety to facilitating efficient

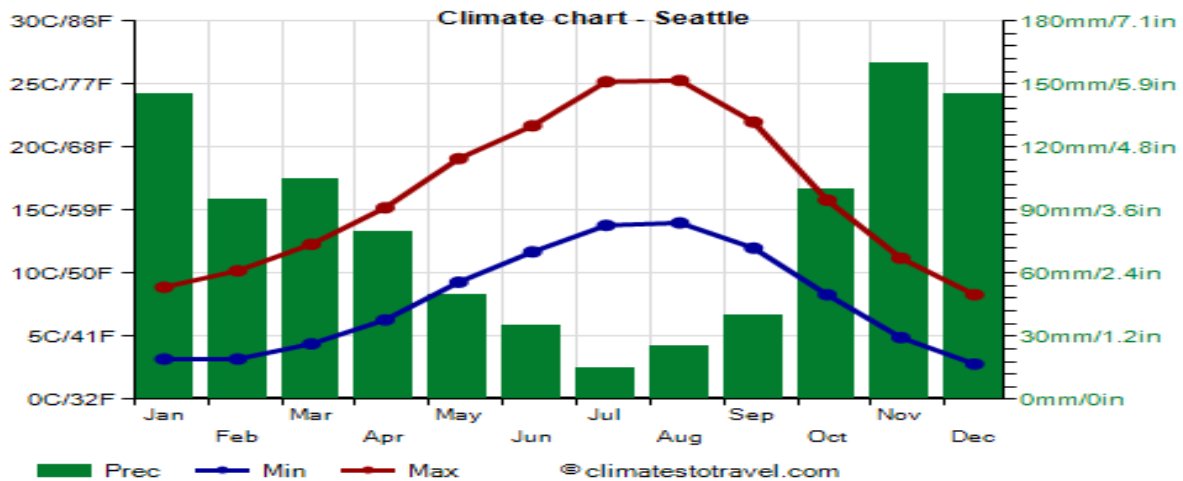


Figure 1: Seattle Climate Chart

II. PROBLEM STATEMENTS

The Seattle Weather Prediction System addresses these challenges by utilizing advanced data mining techniques to analyze a wide range of weather-related data sources. By processing historical data, real-time observations, satellite imagery, and other relevant information, the system can identify patterns and trends in Seattle's weather patterns. Through the application of machine learning algorithms and statistical analysis, the system can generate precise and trustworthy weather forecasts for the Seattle area. These forecasts take into account the unique characteristics of the region's weather patterns, including their erratic and quickly changing nature. By providing accurate and reliable weather forecasts, the system enables businesses and individuals to plan their activities more effectively. Farmers can align their planting and harvesting schedules with periods of favorable weather, optimizing their crop yields. Transportation providers can adjust their operations based on predicted weather conditions, ensuring safer and more efficient routes. Event planners can make informed decisions regarding outdoor activities, considering the likelihood of inclement weather and potential safety risks. Additionally, the Seattle Weather Prediction System plays a crucial role in supporting emergency response efforts. By providing timely and accurate forecasts, emergency management agencies can better prepare for severe weather events, allocate resources effectively, and issue timely warnings to the public, ensuring their safety. Overall, the system's utilization of data mining techniques enables it to tackle the challenges posed by Seattle's unpredictable weather patterns. By delivering precise and trustworthy forecasts, it empowers businesses, individuals, and emergency responders to make informed decisions, enhance safety measures, and optimize their operations in the face of changing weather conditions.

III. PROPOSED SOLUTION TO THE PROBLEM STATEMENTS

The Seattle Weather Prediction System utilizes a combination of data sources, including satellites, radar, and weather stations, to gather information about current weather conditions. By analyzing this data and applying sophisticated atmospheric physics and meteorological models, the system generates precise forecasts for various weather variables such as temperature, precipitation, and wind speed. The accurate weather forecasts provided by the system play a vital role in helping businesses, organizations, and individuals overcome the challenges posed by Seattle's erratic weather patterns. By



having access to reliable weather information, they can plan their activities more effectively, make informed decisions, and ensure the safety of their operations.

For businesses in sectors such as agriculture, transportation, and event planning, the precise weather forecasts enable them to optimize their operations. Farmers can determine the ideal time for planting and harvesting based on forecasted weather conditions. Transportation companies can adjust routes and schedules to account for inclement weather, reducing disruptions and improving efficiency. Event organizers can make informed decisions about whether to proceed with outdoor activities or make contingency plans based on the forecasted weather. Moreover, individuals in the Seattle area can rely on accurate weather forecasts to plan their daily activities, such as outdoor recreation, commuting, and personal events. By knowing the expected weather conditions in advance, they can dress appropriately, prepare for potential weather-related hazards, and make informed choices to ensure their safety and well-being. Overall, the Seattle Weather Prediction System plays a crucial role in assisting businesses, organizations, and individuals in navigating the challenges posed by Seattle's changing weather patterns. By offering accurate and reliable weather forecasts, the system enables them to make proactive decisions, enhance safety measures, and optimize their activities in the face of Seattle's unpredictable weather conditions.

IV. LITERATURE REVIEW

weather forecasting is indeed a complex and challenging task due to the dynamic and chaotic nature of weather systems. Weather is a continuous and multidimensional process influenced by various factors such as temperature, humidity, pressure, wind patterns, and more. Accurately predicting weather conditions requires analyzing large amounts of data and applying sophisticated mathematical models. Meteorological services worldwide, including those in India, face the responsibility of providing reliable weather forecasts to support a range of activities, from agriculture and transportation to disaster management and daily planning for individuals. These organizations rely on a combination of observational data, such as weather stations, satellites, and radars, as well as numerical weather prediction models. Time series data manipulation is an essential technique in weather forecasting. By analyzing historical weather data over a period of time, meteorologists can identify patterns and trends that help in predicting future weather conditions. Time series analysis involves statistical methods and algorithms to analyze and interpret the data, enabling forecasters to make informed predictions about future weather patterns. In recent years, advancements in technology, such as improved satellite imagery, higher-resolution weather models, and increased computing power, have enhanced the accuracy of weather forecasting. Additionally, machine learning and artificial intelligence techniques are being employed to improve forecast models by learning from historical data and making predictions based on patterns and correlations. Collaboration between meteorological organizations worldwide is crucial in sharing data, expertise, and best practices to improve weather forecasting capabilities globally. This cooperation allows for better understanding and prediction of weather phenomena that can have far-reaching impacts on various aspects of society and the economy.

Folorunsho Olaiya ,Department of Computer & Information Systems, Achievers University, Adesesan Barnabas Adeyemo University of Ibadan, Ibadan, Nigeria: "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies"[1].

The paper investigates the use of data mining techniques, specifically Artificial Neural Network and Decision Tree algorithms, for weather forecasting in the city of Ibadan, Nigeria. The study focuses on predicting maximum temperature, rainfall, evaporation, and wind speed using meteorological data collected between 2000 and 2009.

To begin with, a data model was developed to represent the meteorological data, and this model was used to train the classifier algorithms. The performance of these algorithms was evaluated using standard performance metrics, allowing for a comparison between them. The algorithm that yielded the best results was selected to generate classification rules for the mean weather variables. Additionally, a predictive Neural Network model was developed for the weather prediction program. The results obtained from this model were then compared with the actual weather data for the predicted periods. By comparing the predicted values with the observed values, the researchers could assess the accuracy and reliability of the



predictive model. Based on the results, the study concludes that with a sufficient amount of case data, data mining techniques can be effectively utilized for weather forecasting and climate change studies. It implies that by employing Artificial Neural Networks and Decision Trees, valuable insights can be gained from the collected meteorological data, aiding in predicting various weather variables and understanding climate patterns.

H.A. Issad, R. Aoudjit, and J. J. P. C. Rodrigues, “A comprehensive review of Data Mining techniques in smart agriculture[2].

The accurate and effective weather forecasting models are crucial for providing early warnings and mitigating the impacts of weather on various aspects of human livelihood. While significant progress has been made in weather forecasting, it remains a complex and challenging task due to the inherent complexity of atmospheric processes.

Weather forecasting involves analyzing vast amounts of data from various sources, such as weather satellites, weather stations, buoys, radars, and weather balloons. This data is used to create numerical models that simulate the behavior of the atmosphere and predict future weather conditions. However, there are several factors that contribute to the challenges in achieving the most accurate and effective forecasting models:

Complexity of atmospheric processes: The Earth's atmosphere is a dynamic system with intricate interactions between different variables, such as temperature, humidity, pressure, wind patterns, and cloud formation. Modeling these processes accurately requires comprehensive understanding and precise representation in the forecasting models.

Data limitations and uncertainties: Weather data collection is subject to limitations and uncertainties. Gaps in the data, errors in measurements, and variations in the quality and availability of observations can impact the accuracy of forecasts. Improving data collection methods and enhancing data quality are ongoing challenges in weather forecasting.

Preeti khare 2021, International Journal of Emerging Trends in Engineering Research:”A Survey of Weather Forecasting based on Machine Learning and Deep Learning Techniques”[3]

Weather forecasting is indeed a critical concern in today's era, as it enables us to make informed decisions and take preventive measures against disasters. Various techniques have been employed to forecast future weather, including statistical analysis, machine learning, and deep learning techniques. These methods aim to improve the accuracy and reliability of weather predictions. Statistical analysis is a traditional approach to weather forecasting that involves analyzing historical weather data and identifying patterns and correlations. By applying statistical models, meteorologists can make predictions based on past observations. However, this method may have limitations in capturing complex relationships and non-linear patterns present in weather data. Machine learning techniques have gained significant popularity in weather forecasting due to their ability to handle large and complex datasets. These methods involve training models on historical weather data and using the learned patterns to make predictions. Various machine learning algorithms, such as decision trees, random forests, support vector machines, and neural networks, can be applied to weather forecasting tasks. Machine learning models can incorporate multiple input variables, such as temperature, humidity, wind speed, and atmospheric pressure, to generate forecasts. Deep learning techniques, a subset of machine learning, have shown great promise in weather prediction. Deep learning models, particularly convolution neural networks (CNNs) and recurrent neural networks (RNNs), can effectively capture spatial and temporal dependencies in weather data. CNNs are well-suited for image-based weather forecasting tasks, where weather radar or satellite images are used as input. RNNs, on the other hand, are suitable for sequential data, such as time series data collected from weather stations. The accurate prediction of short-term weather conditions, such as hourly or daily forecasts, is crucial for various applications, including disaster management strategies and smart city programs. Timely and accurate forecasts can help authorities and individuals take proactive measures to mitigate the impact of severe weather events. By leveraging machine learning and deep learning techniques, forecast models can continuously improve their accuracy and provide valuable insights for decision-making.

Ismaila Oshodi, Department of Computing, Bournemouth University, Bournemouth England S5423662@bournemouth.ac.uk:”Machine learning-based algorithms for weather forecasting”[4]

Machine learning-based weather forecasting models have gained popularity in recent years due to their potential to improve the accuracy and efficiency of weather predictions. The traditional methods, as you mentioned, rely on complex mathematical equations and require significant computational power. In contrast, machine learning algorithms can analyze large datasets and learn patterns from historical weather data to make predictions.

Using classifier algorithms such as Random Forest, Decision Tree, Gaussian Naïve Bayes, and Gradient Boosting Classifier for weather forecasting is a novel approach. These algorithms are known for their ability to handle classification tasks



effectively. It's impressive that the Gaussian Naive Bayes algorithm demonstrated the best performance in your research, achieving a predictive accuracy of 84.153%.

Accurate weather forecasts are indeed crucial for various sectors like agriculture, transportation, tourism, and industry. Timely and reliable weather information helps in making informed decisions related to crop planting, harvesting, transportation scheduling, tourist activities, and industrial operations. Having a machine learning-based model that can provide accurate forecasts can greatly benefit these sectors and contribute to the overall economy.

It's worth mentioning that while your research focused on the city of Seattle for a specific period, the performance of the model may vary when applied to different locations and timeframes. Factors such as local climate patterns, topography, and the availability of quality data can affect the model's accuracy. Therefore, it's important to validate and fine-tune the model for specific regions before implementing it on a larger scale. Overall, the utilization of machine learning algorithms for weather forecasting holds great promise, and further research and development in this area can lead to significant advancements in accuracy, efficiency, and reliability of weather predictions, benefiting various industries and society as a whole.

Elia Georgiana Petre Universitatea Petrol-Gaze din Ploiești, Bd. București 39, Ploiești, Catedra de Informatică, "A Decision Tree for Weather Prediction "[5]

In paper mentioned presents a small application of CART (Classification and Regression Trees) for weather prediction using data collected in Hong Kong between 2002 and 2005. CART is a decision support tool commonly used due to its simplicity in understanding and interpretation. To build the decision tree, the authors utilized Weka, a free data mining software available under the GNU General Public License. Weka offers various algorithms and tools for data analysis and modeling. In this case, the authors used CART to construct the decision tree based on the values of specific variables in the data set. The decision tree generated through this process represents a graphical structure composed of nodes and branches. Each node represents a decision or a test on a particular variable, while the branches represent the outcomes or possible paths based on the variable's value. The tree structure aids in predicting whether conditions based on the given variables, in this case, weather-related predictions.

No.	Paper Title	Author Name	Key Points	Remark
1	Application of Data Mining Techniques in Weather Prediction and Climate Change Studies	Folorunsho Olaiya, Achievers University, Adesesan Barnabas Adeyemo University of Ibadan	The paper investigates the use of data mining techniques, specifically Artificial Neural Network and Decision Tree algorithms, for weather forecasting in the city of Ibadan, Nigeria.	This work is important to climatic change studies because the variation in weather conditions in term of temperature, rainfall and wind speed can be studied using these data mining techniques.
2	A comprehensive review of Data Mining techniques in smart agriculture	H.A. Issad, R. Aoudjit, and J. J. P. C. Rodrigues	The aim of this paper is to review ongoing studies and research on smart agriculture using the recent practice of Data Mining, to solve a variety of agricultural problems.	Smart management consists of collecting, transmitting, selecting and analyzing data
3	A Survey of Weather Forecasting based on Machine Learning and Deep Learning	Preeti khare 2021	Several techniques that have been used to forecast future weather are statistical analysis, machine learning, and deep learning techniques.	Required more methods based on Machine Learning, and Deep Learning Techniques used to predict



	Techniques			weather and discusses the applications, types, and problems of the forecasting.
4	Machine learning-based algorithms for weather forecasting	Ismaila Oshodi	Weather forecast is made by collecting quantitative, atmospheric data about the past and current state of the atmosphere	These traditional processes utilize expensive, complex physical and computational power to produce forecasts, which can be inaccurate and have various catastrophic impacts on society.
5	A Decision Tree for Weather Prediction	Elia Georgiana Petre,Bd. București 39	There are presented the decision tree, the results and the statistical information about the data used to generate the decision model.	They required the increase the precision in building the decision tree and the weather prediction based on it.

III.METHODOLOGY OF PROPOSED SURVEY

Here we are describing a technology called "Expression Glass" that utilizes machine vision for face or eye recognition, specifically analyzing pattern recognition and muscle variations to identify expressions such as interest or confusion. The glass incorporates electrical sensors to capture the necessary image data.This type of technology is related to the field of computer vision, which focuses on enabling machines to understand and interpret visual information. By using machine learning algorithms, computer vision systems can analyze patterns and features in images or video streams, allowing them to recognize and interpret human expressions.

Expression Glass, as appears to be a specific implementation or product that leverages machine vision for expression recognition. It likely uses a combination of image sensors, image processing algorithms, and machine learning techniques to detect and analyze facial or eye movements and infer the user's emotional state.It's important to note that my response is based on the information provided, and there may be variations or specific details about Expression Glass that are not covered here. Additionally, as an AI model, my knowledge is based on data available up until September 2021, so there may have been advancements or updates in this field since then.

A.DATA SOURCE

We have utilized Kaggle platform and resources to develop a weather prediction model based on the Seattle Weather Prediction datasets. Using Jupyter Notebooks in a no-setup environment along with access to free GPUs can be quite beneficial for data analysis and model development.

We have five features in the model and one target value, which is the overall weather. It's common in machine learning to use multiple features to predict a target variable. By training your model on the provided datasets, you aimed to achieve high accuracy in weather prediction.Accuracy is an important metric when evaluating the performance of a predictive model. It measures the proportion of correct predictions made by the model. However, it's worth noting that the accuracy of a weather prediction model can vary based on several factors, such as the quality and quantity of the training data, the chosen algorithm, and the features used.



To further enhance the accuracy of your model, you might consider exploring additional features that could potentially improve the predictions. Additionally, evaluating the model's performance using other metrics such as precision, recall, or F1 score can provide a more comprehensive understanding of its effectiveness.

Remember that weather prediction is a complex task influenced by various factors, so achieving a high level of accuracy can be challenging.

Datasets Includes:-

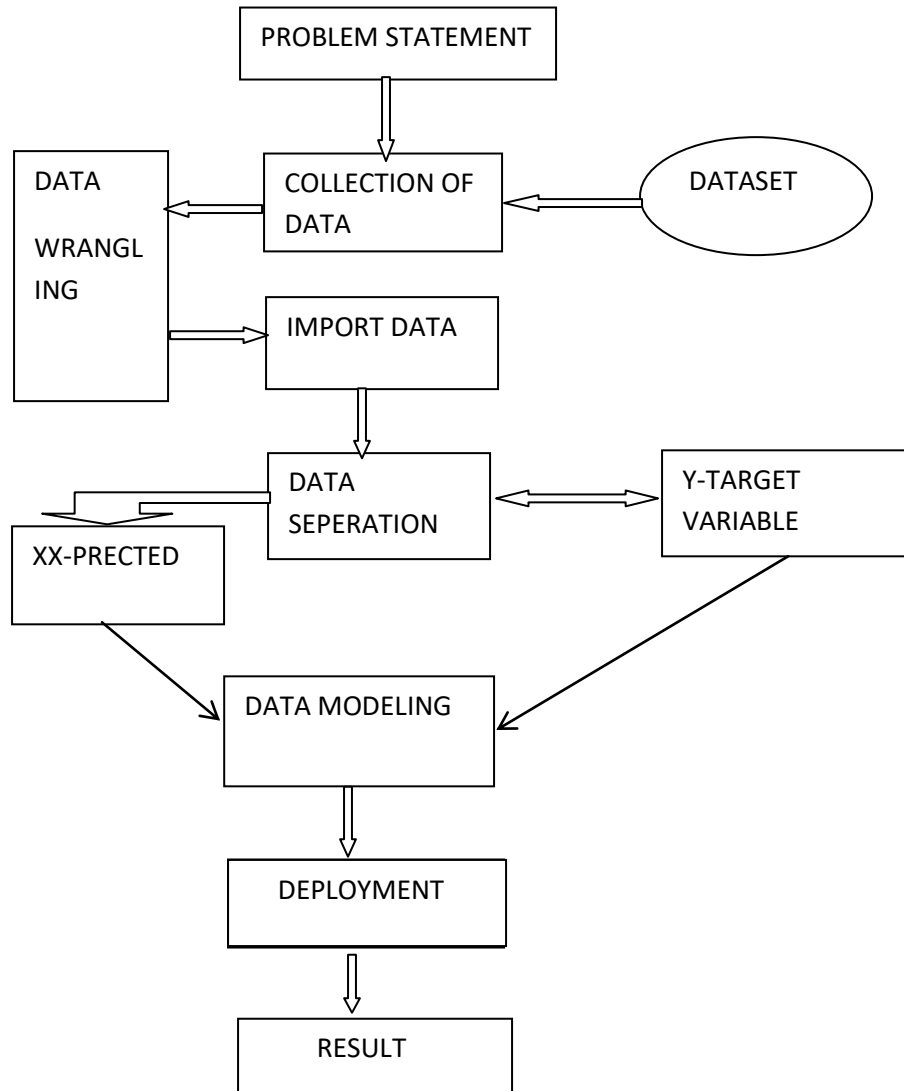
List of input and output variables

variable	Weather Parameters	Types
X1	DAY(INT)	INPUT
X2	MONTH(INT)	INPUT
X3	PRCP(FLOAT)	INPUT
X4	TMAX(INT)	INPUT
X5	TMIN(INT)	INPUT
Y1	RAINFALL(INT)	TARGET

Many researchers globally have been working on weather prediction using historical data and machine learning models in recent years to prevent weather prediction unreliability. In this article, the study used classification algorithms for the weather prediction model to classify data for the target column used to predict rain in Seattle weather data.



FRAMEWORK MODEL FOR WEATHER FORECAST



A.DESCRPTION OF ALGORITHM USED IN MODEL ACCORDING TO THE ACCURACY:-

1.NAIVE BAYES:

Naive Bayes is a probabilistic machine learning algorithm used for classification tasks. It is based on Bayes' theorem, which is a fundamental principle in probability theory. Bayes' theorem calculates the conditional probability of an event based on prior knowledge of related conditions. In the context of classification, Naive Bayes calculates the probability of a certain outcome (or class) given a set of input variables (or features). It assumes that the features are conditionally independent of each other, meaning that the presence or absence of one feature does not affect the probability of the presence or absence of another feature. This assumption is what makes Naive Bayes "naive" because it oversimplifies the relationships between features. In reality, features may be dependent on each other to some extent. Despite its naive assumptions, Naive Bayes is widely used and often performs well in practice, particularly in text classification tasks. Naive Bayes models are relatively simple, computationally efficient, and require less training data compared to other more complex algorithms. In scenarios



where the independence assumption is violated or when there are strong dependencies among features, other algorithms such as logistic regression, decision trees, or support vector machines might be more suitable

- It's Accuracy in the model is 84.15%

```

- Naive Bayes

from sklearn.naive_bayes import GaussianNB
classifier = GaussianNB()
classifier.fit(X_train, y_train)

GaussianNB()

[ ] y_pred = classifier.predict(X_test)

[ ] cm = confusion_matrix(y_test, y_pred)
print(cm)

[[ 0  0  0  0 11]
 [ 0  0  0  0 31]
 [ 0  0 141  2 12]
 [ 0  0  2  4  0]
 [ 0  0  0  0 163]]

[ ] acc4 = accuracy_score(y_test, y_pred)
print(f"Accuracy score : {acc4}")

Accuracy score : 0.8415300546448088
    
```

2. XGBOOST:-

XGBoost (Extreme Gradient Boosting) is indeed a popular machine learning algorithm used for regression and classification tasks. It is an implementation of gradient boosting, a technique that combines multiple weak learners (usually decision trees) in an ensemble. These weak learners are trained sequentially, with each subsequent model attempting to correct the errors made by the previous models. XGBoost is highly regarded for its performance and speed, making it suitable for handling large datasets with a high number of features. It utilizes parallel processing and tree pruning techniques to optimize training time and memory usage.

The algorithm has gained significant popularity in various industries for predictive modeling tasks. It has found applications in finance, advertising, healthcare, and other domains where accurate predictions are crucial. XGBoost has also been successful in machine learning competitions, such as those hosted on Kaggle, and has won several of them.

- However, the accuracy percentage you provided is 82.79%.



▾ XGBoost

```
[ ] from xgboost import XGBClassifier
    classifier = XGBClassifier()
    classifier.fit(X_train, y_train)
```

```
XGBClassifier(objective='multi:softprob')
```

```
▶ y_pred = classifier.predict(X_test)
  cm = confusion_matrix(y_test, y_pred)
  print(cm)
  acc7 = accuracy_score(y_test, y_pred)
```

```
⋮ [[ 0  0  0  0 11]
   [ 0  1  0  0 30]
   [ 0  1 143  0 11]
   [ 0  0  5  1  0]
   [ 0  5  0  0 158]]
```

```
[ ] print(acc7)
```

```
0.8278688524590164
```

3.RANDOM FOREST:-

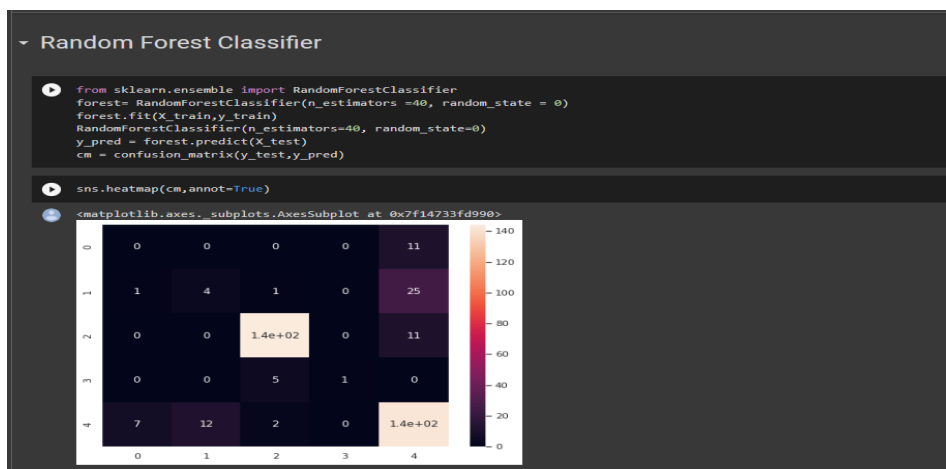
Random Forest is a popular machine learning algorithm that belongs to the ensemble learning family. It combines multiple decision trees to make predictions. Here are some key points about Random Forest:

Ensemble of Decision Trees: Random Forest creates an ensemble by constructing multiple decision trees during the training phase. Each decision tree is built using a randomly selected subset of features and observations from the training data.

Random Feature Subsets: To introduce randomness and diversify the individual trees, Random Forest randomly selects subsets of features at each node of the decision tree. This helps to reduce correlation between trees and promotes better generalization.

Bootstrap Aggregation (Bagging): Random Forest employs a technique called bagging, which involves training each decision tree on a randomly sampled subset of the training data with replacement. This means that some observations may be repeated, while others may be left out, creating diverse training sets for each tree.

- The Accuracy percentage of this algorithm in the model is **79.51%**.





```
[ ] from sklearn.metrics import classification_report
print(classification_report(y_test,y_pred))
```

	precision	recall	f1-score	support
0	0.00	0.00	0.00	11
1	0.25	0.13	0.17	31
2	0.95	0.93	0.94	155
3	1.00	0.17	0.29	6
4	0.75	0.87	0.81	163
accuracy			0.80	366
macro avg	0.59	0.42	0.44	366
weighted avg	0.77	0.80	0.78	366

```
[ ] acc6 = forest.score(X_test,y_test)
print(acc6)
```

0.7950819672131147

4.SVM(SUPPORT VECTOR MACHINE):-

Support Vector Machines, is indeed a machine learning algorithm used for classification and regression analysis. Here are some additional details about SVM:

Classification and Regression: SVM can be used for both classification and regression tasks. In classification, the algorithm assigns data points to different classes based on their features. In regression, SVM predicts a continuous target variable.

Hyperplane and Separation: SVM finds the best hyperplane in the feature space to separate data points belonging to different classes. The hyperplane is chosen to maximize the margin or distance between the closest data points of different classes, which helps improve the algorithm's generalization ability.

Kernel Functions: SVM has the flexibility to use different kernel functions to transform the data into a higher-dimensional feature space. The transformation allows for better separation of classes in cases where the data is not linearly separable in the original feature space. Common kernel functions include linear, polynomial, Gaussian (RBF), and sigmoid.

Handling Complex Data Sets: SVM is known for its effectiveness in handling complex data sets with high-dimensional feature spaces. It can handle cases where the number of features is much larger than the number of samples. SVM is also less prone to over fitting compared to other algorithms, thanks to the regularization parameter.

Applications: SVM has found applications in various fields. In finance, it can be used for tasks such as credit scoring and stock market analysis. In bio informatics, SVM is used for protein classification, gene expression analysis, and DNA classification. Additionally, SVM has been widely employed in image recognition tasks, including object detection and face recognition.

- The Accuracy percentage of this algorithm in the model is **79.51%**.



```

SVM

[ ] from sklearn.svm import SVC
classifier = SVC(kernel = 'linear', random_state = 0)
classifier.fit(X_train, y_train)

SVC(kernel='linear', random_state=0)

[ ] y_pred = classifier.predict(X_test)

[ ] cm = confusion_matrix(y_test, y_pred)
print(cm)
acc2 = accuracy_score(y_test, y_pred)

[[ 0  0  0  0 11]
 [ 0  0  0  0 31]
 [ 0  0 126  0 29]
 [ 0  0  4  2  0]
 [ 0  0  0  0 163]]

[ ] print(f"Accuracy score: {acc2}")

Accuracy score: 0.7950819672131147

```

5.LOGISTIC REGRESSION:-

Logistic regression is specifically designed for situations where the dependent variable is categorical, typically binary (e.g., yes/no, true/false, success/failure). However, it can also be extended to handle more than two categories using techniques like multinomial logistic regression or ordinal logistic regression. The goal of logistic regression is to model the relationship between the independent variables (also known as predictors or features) and the probability of a certain outcome occurring. Unlike linear regression, which predicts continuous numeric values, logistic regression models the log-odds (also known as logit) of the outcome variable. The log-odds are then transformed using the logistic function (also known as the sigmoid function) to obtain the predicted probabilities, which are bounded between 0 and 1. Logistic regression estimates the coefficients (also known as weights or parameters) for each independent variable, indicating the direction and magnitude of their influence on the probability of the outcome. These coefficients can be interpreted as the change in the log-odds or odds ratio for a one-unit increase in the corresponding independent variable, assuming all other variables are held constant. Logistic regression is widely used in various domains, including healthcare, marketing, finance, social sciences, and more. It is particularly popular in machine learning due to its simplicity, interpretability, and good performance in many classification tasks. While logistic regression is a powerful tool, it has assumptions that should be met, such as the linearity of the log-odds and the absence of multicollinearity among the independent variables.

- It's Accuracy in the model is **76.23%**



and the process continues until a stopping criterion is met, such as reaching a maximum tree depth or a minimum number of samples in each partition.

Once the tree is constructed, prediction is performed by traversing the tree from the root node to a leaf node based on the values of the input variables for a given data point. At each decision node, the algorithm evaluates a condition on one of the input variables, determining which branch to follow. Finally, when a leaf node is reached, the predicted output value is typically calculated as the average of the target variable values in the training samples that belong to that leaf node.

- The Accuracy percentage of this algorithm in the model is **72.40%**.

```

Decision Tree Classification

[ ] from sklearn.tree import DecisionTreeClassifier
classifier = DecisionTreeClassifier(criterion = 'entropy', random_state = 0)
classifier.fit(X_train, y_train)

DecisionTreeClassifier(criterion='entropy', random_state=0)

▶ y_pred = classifier.predict(X_test)

[ ] y_pred

array([[4, 0, 2, 4, 4, 2, 2, 2, 4, 2, 0, 2, 4, 1, 0, 4, 1, 2, 2, 2, 2, 2, 2,
2, 4, 4, 0, 4, 2, 0, 4, 2, 4, 3, 2, 2, 2, 0, 2, 0, 2, 4, 2, 2, 2, 2, 2, 2,
4, 0, 4, 4, 4, 4, 4, 4, 4, 2, 2, 4, 4, 3, 4, 2, 4, 4, 2, 4, 4, 1, 1, 2,
4, 1, 4, 2, 2, 3, 2, 0, 1, 4, 2, 0, 2, 2, 4, 4, 2, 2, 4, 4, 2, 2, 4, 4, 2,
4, 2, 4, 1, 2, 3, 4, 0, 4, 2, 2, 4, 2, 2, 2, 2, 2, 4, 4,
2, 4, 2, 4, 4, 4, 4, 4, 2, 2, 4, 2, 2, 1, 2, 4, 4, 0, 4, 2, 4, 2,
4, 2, 2, 2, 4, 4, 4, 2, 3, 4, 4, 4, 2, 2, 4, 4, 4, 1, 4, 2,
1, 2, 0, 4, 2, 4, 4, 0, 2, 1, 2, 2, 4, 2, 2, 2, 1, 1, 2, 4, 2, 2,
4, 4, 4, 4, 4, 2, 1, 2, 1, 4, 2, 2, 4, 4, 4, 1, 4, 2, 2, 4, 0,
2, 4, 2, 2, 1, 4, 2, 2, 2, 4, 1, 1, 3, 2, 2, 2, 1, 4, 1, 2, 2, 2,
1, 4, 4, 2, 4, 4, 2, 4, 2, 0, 0, 2, 2, 2, 2, 1, 2, 4, 2, 2, 4, 2,
4, 2, 0, 2, 4, 0, 4, 4, 4, 2, 3, 4, 4, 4, 4, 4, 2, 1, 4, 4,
1, 2, 4, 2, 4, 1, 4, 4, 2, 2, 0, 1, 2, 4, 2, 4, 2, 2, 2, 0, 0, 4,
2, 4, 2, 2, 1, 2, 4, 2, 1, 2, 2, 4, 2, 4, 2, 4, 4, 2, 4, 4, 2, 4, 2,
2, 4, 2, 0, 2, 2, 4, 1, 2, 2, 2, 0, 2, 1, 2, 2, 4, 4, 2, 2, 2, 4,
4, 2, 0, 4, 2, 4, 4, 4, 2, 2, 4, 4, 4, 2, 4, 4, 2, 2, 1, 2, 4, 4,
4, 4, 2, 4, 4, 2, 4, 4, 4, 2, 1, 2, 2, 4]])

[ ] cm = confusion_matrix(y_test, y_pred)
print(cm)

[[ 1  2  1  0  7]
 [ 6  8  1  0 16]
 [ 0  1 140  4 10]
 [ 0  0  3  3  0]
 [18 23  9  0 113]]
    
```

```

[ ] sns.heatmap(cm,annot=True)

<matplotlib.axes._subplots.AxesSubplot at 0x7f1473121750>

[[ 1  2  1  0  7]
 [ 6  8  1  0 16]
 [ 0  1 140  4 10]
 [ 0  0  3  3  0]
 [18 23  9  0 113]]

[ ] acc5 = accuracy_score(y_test, y_pred)
print(f"Accuracy score: {acc5}")

Accuracy score: 0.7240437158469946
    
```



7.KNN:-

KNN (k-Nearest Neighbors) is indeed a machine learning algorithm used for classification and regression analysis. Here's a more detailed explanation of how KNN works:

Training Phase: During the training phase, the algorithm memorizes the entire training dataset, which consists of labeled data points with their corresponding class labels or target values.

Distance Calculation: When a new unlabeled test data point is given, KNN calculates the distance between that test point and every point in the training dataset. The most commonly used distance metric is the Euclidean distance, although other distance metrics like Manhattan distance or Minkowski distance can also be used.

Finding the Neighbors: After calculating distances, KNN selects the k nearest neighbors from the training dataset based on the smallest distances. The value of k is a user-defined parameter and determines the number of neighbors to consider.

Classification or Regression: In the case of classification, KNN assigns the class label that is most frequent among the k nearest neighbors to the test data point. For regression, KNN predicts the average of the target values of the k nearest neighbors

- The Accuracy percentage of this algorithm in the model is **67.60%**.

```

- KNN Model

[ ] from sklearn.neighbors import KNeighborsClassifier
    classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
    classifier.fit(X_train, y_train)

KNeighborsClassifier()

[ ] y_pred = classifier.predict(X_test)

cm = confusion_matrix(y_test, y_pred)
print(cm)

[[ 1  1  3  0  6]
 [ 1  4  5  0 21]
 [ 0  3 127  0 25]
 [ 0  0  3  1  2]
 [ 5 17 26  0 115]]

[ ] acc3 = accuracy_score(y_test, y_pred)
    print(f"Accuracy score: {acc3}")

Accuracy score: 0.6775956284153005
    
```

Every model has their different evaluation matrix. The evaluation metrics allow for a quick assessment of the model's performance. All the performance of the evaluation matrix is on ,the accuracy score, precision score, recall score, and F1-score evaluation metrics were used in this experiment.

IV.CONCLUSION

The Naive Bayes algorithm-based Seattle weather forecasting system has been successful in predicting the weather and providing accurate classifications of various weather conditions. The Naive Bayes algorithm is known for its simplicity and effectiveness in classification tasks, so it's not surprising that it has performed well in this context. Adding more input variables to the system can certainly enhance its accuracy and robustness. Weather prediction is a complex task that depends on multiple factors such as temperature, humidity, wind speed, precipitation, and atmospheric pressure. By incorporating more relevant variables into the system, it can capture a broader range of influences on the weather, leading to more accurate predictions. Additionally,



incorporating different machine learning techniques can also contribute to improving the system's performance. Different algorithms have their strengths and weaknesses, and combining multiple algorithms through ensemble methods or using more advanced techniques like deep learning can potentially provide even better results. These techniques can help capture more complex patterns and relationships in the weather data, leading to more accurate predictions. The Accuracy of the same was **84.15%**

REFERENCES

- [1]Folorunsho Olaiya ,Department of Computer & Information Systems, Achievers University, Adesesan Barnabas Adeyemo University of Ibadan, Ibadan, Nigeria: "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies"
- [2] H. A. Issad, R. Aoudjit, and J. J. P. C. Rodrigues, "A comprehensive review of Data Mining techniques in smart agriculture," *Eng. Agric. Environ. Food*, vol. 12, no. 4, pp. 511–525, Oc
- [3]Preeti khare 2021, International Journal of Emerging Trends in Engineering Research: "A Survey of Weather Forecasting based on Machine Learning and Deep Learning Techniques"
- [4]Ismaila Oshodi ,Department of Computing,Bournemouth University ,Bournemouth England S5423662@bournemouth.ac.uk: "Machine learning-based algorithms for weather forecasting"
- [5]Elia Georgiana Petre Universitatea Petrol-Gaze din Ploiești, Bd. București 39, Ploiești, Catedra de Informatică, "A Decision Tree for Weather Prediction "



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor
7.54

ISSN

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com