[Project Report]

WEATHER PREDICTION

Vishal Thakur
vt8001369@gmail.com | JUNE __, 2022

PREDICTING THE WEATHER WITH ANN

PROJECT REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF

BACHELOR OF TECHNOLOGY

In

(Computer Science Engineering)

Submitted by

Vishal Thakur (1954302008)

Abhishek Thakur (1954302046)

Prachi Aggarwal (1954302064)

Arya (GF/2020/6098)

Vinayak (1954302106)

YOGANANDA SCHOOL OF AI, COMPUTERS

AND DATA SCIENCES

SHOOLINI UNIVERSITY OF BIOTECHNOLOGY AND MANAGEMENT SCIENCES

SOLAN, H.P., INDIA

JULY, 2022

DECLARATION BY THE CANDIDATE

I hereby declare that the project report entitled "**Predicting Weather with ANN**" submitted in partial fulfilment for the award of degree of **Bachelor of Technology** to Shoolini University of Biotechnology and Management Sciences, Solan (H.P.) is original research work carried out by me under the guidance and supervision of Mrs. Sonia. No part of this thesis has been submitted for any other degree or diploma to this or any other university.

VISHAL THAKUR (1954302008)

PRACHI AGGARWAL (1954302064)

ABHISHEK THAKUR (1954302046)

ARYA (GF/2020/6098)

VINAYAK (1954302106)

PLACE: SOLAN

DATE:

Shoolini University

Shoolini University, Solan-173229

CERTIFICATE 1

This is to certify that the project report entitled "**Predicting Weather with ANN**" submitted in partial fulfilment for the award of the degree of **Bachelor of Technology** to Shoolini University of Biotechnology and Management Sciences, Solan (H.P.) is original research work carried out by **VISHAL, ABHISHEK, PRACHI, ARYA, VINAYAK** under my guidance and supervision. No part of this report has been submitted for any other degree or diploma to this or any other university.

The assistance and help received during the course of investigation has been duly acknowledged.

(Name and Signature of Research Guide)

Countersigned By:

Head of School, Yogananda School of AI,

Computers and Data Sciences

Place:

Date………

CONTENTS

ACKNOWLEDGEMENT

At the onset, I'd like to start by thanking Mr. Anurag Rana. He always has been so cooperative. Without their cooperation, the project work/ training report couldn't have been what it evolved to be.

Word of thanks to Mr. Pankaj Vaidya (Head, Computer Science Engineering Department, Shoolini University) and the distinguished faculty of Computer Science and Engineering Department, Shoolini University for their supportive and helping custom. I'd like to thank my colleagues, who helped me in the write-ups at times and suggested a few ideas.

Above all, thanks to almighty for always being around us and helping without asking.

VISHAL THAKUR

ABHISHEK THAKUR

PRACHI AGGARWAL

ARYA

VINAYAK

LIST OF ABBREVIATIONS AND SYMBOLS

ANN	Artificial Neural Network
BP	Back Propagation
MLP	Multi-Layer Perceptron
SVM	Support Vector Machine

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Accurate weather forecasting is very important because agricultural and industrial sector are based on it. We are presenting weather predictions using Artificial Neural Network and Back Propagation Algorithm. In this paper, we have evaluated the machine learning techniques to predict weather with much accuracy. During this research process we have used following parameters to predict weather: temperature, rainfall, evaporation, sunshine, wind speed, wind direction, cloud, humidity and size of dataset. This research aims to compare the performance of some machine learning algorithms for predicting weather using weather data. From the collected weather data which contains some weather attributes, which are most relevant to weather prediction. In this paper, various Machine Learning Techniques have explored which includes MLP, SVM. The experimental results show that SVM algorithm has good level of accuracy than other algorithms.

Keywords→ Neural Network, BP Algorithm, Feed Forward neural network, MLP, SVM, ANN, Data pre-processing, Weather Prediction.

CHAPTER-1

INTRODUCTION

1.1 INTRODUCTION

The application of science and technology that predicts the state of atmosphere at any given particular time period is known as Weather forecasting. There is a many different methods to weather forecast. Weather forecast notices are important because they can be used to prevent destruction of life and environment. The weather forecasting methods used in the ancient time usually implied pattern recognition i.e., they usually rely on observing patterns of events. For example, it is found that the following day has brought fair weather; if the preceding day sunset is particularly red. However, all of the predictions prove not to be reliable.

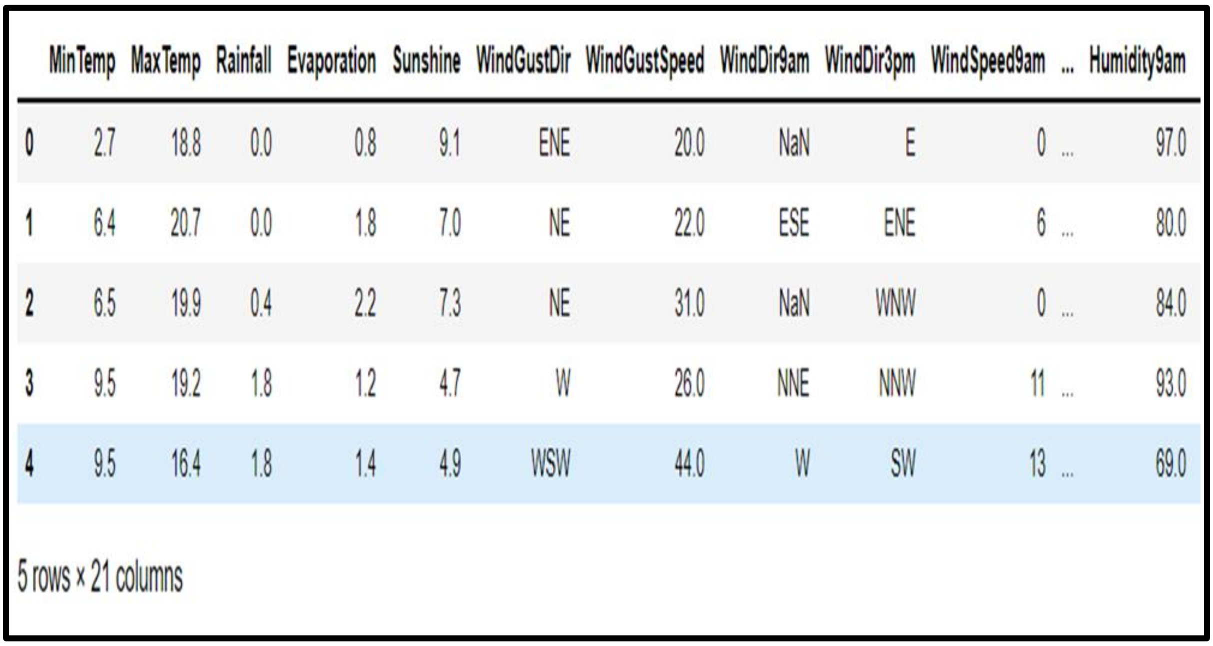
1.2 DATA

Weather forecasting is simply the prediction of future weather based on different parameters of the past like temperature, humidity, dew, wind speed and direction, precipitation, Haze and contents of air, Solar and terrestrial radiation etc. Once the data is taken, it is trained. The more parameters considered, the higher the accuracy. This project can help many people finding the weather of tomorrow.

The project simply uses temperature, dew, pressure and humidity for training the data. The main goal is to find out that how proficient our recommended soft computing Artificial Neural Network and Back Propagation Algorithm are functioning. This project will help a person to foresee the Minimum Temperature, Maximum Temperature and Pressure of the coming day.

This project uses the data of 9-10 years (2008-2017). Artificial Neural Network provides a strategy for solving various nonlinear problems that are inconvenient to solve by conventional methods. A connection can be established with Artificial Neural Networks between the inputs and the outputs without providing the internal processing involved. Hence, it can be concluded that the above-mentioned characteristics of Artificial Neural Network are appropriate for weather forecasting under inspection.

Dataset contain Wind speed, Min-Max Temperature, Humidity, Pressure etc.



	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	...	Humidity9am
0	2.7	18.8	0.0	0.8	9.1	ENE	20.0	NaN	E	0	...	97.0
1	6.4	20.7	0.0	1.8	7.0	NE	22.0	ESE	ENE	6	...	80.0
2	6.5	19.9	0.4	2.2	7.3	NE	31.0	NaN	WNW	0	...	84.0
3	9.5	19.2	1.8	1.2	4.7	W	26.0	NNE	NNW	11	...	93.0
4	9.5	16.4	1.8	1.4	4.9	WSW	44.0	W	SW	13	...	69.0

5 rows × 21 columns

Fig-1: Dataset

1.3 Why Weather Prediction

The ultimate goal of weather forecasting is to protect human lives and property, improve health, safety, and economic prosperity.

Weather prediction positively impacts people's lives in various ways:

- Weather Forecasting benefits tourism
- Weather Forecasting improves transportation safety
- Weather forecasting beneficial to farmers

CHAPTER-2

LITEATURE REVIEW

2.1 Review of Literature

In past years, many researchers work on weather prediction using different techniques. Some are explained in this section. In this research paper comparative study on weather prediction using ML Techniques data. Researcher analysis on different Machine Learning Algorithms. Firstly, describes weather prediction has many different problems. Even the simplest weather predictions are not perfect. Prediction of forecast varies from one to two degrees of the actual temperature. Although this accuracy of weather prediction is not bad, as predictions are made for further in time. Also, sometimes accuracy of weather prediction can be even worse. Furthermore, weather prediction in some areas where the climate is not consistence, is off by even more. Machine Learning Algorithms and many classifiers' names Naive Bayes Bernoulli, Logistic Regression, Gaussian, support vector machine are uses for evaluate more accurate output.

A number of projects and works have been done in the field of forecasting of temperature and pressure. This will help us interpret our project in a more effective way, a few of those works are discussed below:

MohsenHayati& Zahra Mohebi uses Artificial Neural Network in forecasting the temperature of the coming day. They divided the data in 4 sections, each section representing a season and each season had its own separate network. MLP was used to train the network and the data of 10 years, i.e., 1996-2006 was considered. The error varied between 0-2 MSE in the result.

Dr.S. SanthoshBaboo& I. KadarShereef examined the prime algorithms to train uses Artificial Neural Network and used the BPN for forecasting. Year weather information was used to train Artificial Neural Network and test which included Temperature, Due Point, Humidity, SLP, Visibility & wind speed etc. The minimum error recorded according to research was 0.0079 and the maximum error was 1.2916 RMSE.

From the above studies it can be drawn close to that implementing soft computing is one of the most efficient ways for small scale weather prediction. According to the studies accurate and well-grounded data is what weather prediction by Artificial Neural Network determines. In ANN the number of variables chosen plays a very important role. The more the variables the more efficient would be the result.

The number of data or information which is gathered for training a model has a very important role in the prediction reliability. The Training set is the most significant unit of our project; it is the set of the input and output provided by Artificial Neural Network. The effective the training set, the effective the result. The Artificial Neural Network consists of artificial neurons which have a processing node which are connected to other nodes, it also has various layers known as the input layer, middle layer and the output layer. The architecture of the Artificial Neural Network model depends upon its use.

Another research paper titled 'Issues with weather prediction' discussed the major problems with weather prediction. Even the simplest weather prediction is not perfect. The one-day forecast typically falls within two degrees of the actual temperature. Although this accuracy isn't bad, as predictions are made for further in time. For example, in a place like New England where temperatures have a great variance the temperature prediction are more inaccurate than a place like the tropics.

Another research paper titled 'Current weather prediction' used numerical methods to stimulate what is most likely going to happen based on known state of the atmosphere. For example, if a forecaster is looking at three different numerical models, and two model predict that a storm is going to hit a certain place, the forecaster would most likely predict that the storm is going to hit the area. These numerical models work well and are being tweaked all the time, but they still have errors because some of the equations used by the models aren't precise.

CHAPTER-3

METHODOLOGY

3.1 METHODOLOGY

In any Prediction accuracy is very important. The input parameters for a weather forecasting model are different type of data required different types of methods and need to be handled accordingly.

Weather forecasting can be done more accurately using ANN. Because daily weather data has multiple parameters such as temperature, humidity, rainfall amount, cloud distance and size, wind speed and direction, etc. All these parameters are non-linear, but they required to be processed together to determine temperature, rainfall, humidity or weather status for the future day. Such type of applications needs complex models and can able to produce the required result by generating the patterns on its own by performing self-learning using the training data given to the model.

To develop an ANN model for weather forecasting, region selection for input data and parameters is necessary. The input data is to be taken from a specific area on which the model is trained and tested so that the model is able to generate accurate results. The number of input data given to model also helps to improve accuracy of the model by giving the results with a high degree of similarity between predicted and actual output data. The available data may be noisy thus, data should be cleaned. Similarly, it has to be normalized because, all the parameters are of different units and normalization will help the input and output parameters to correlate with each other. The data should be divided in training and testing samples in proper proportion so that the results can be predicted, tested and validated properly. Structure of the NN model also has a great impact on generation of accurate results. The multilayer ANN helps in predicting nonlinear data more efficiently. The activation function will be different for different layers of NN as per need.

3.2 SOFTWARE REQUIREMENTS

The software used in our project are:

Python 3.7: Python is an interpreted, high level, general programming language. Its formatting is visually uncluttered, and it often uses English keywords where other languages use punctuation. It provides a vast library for data mining and predictions.

Jupiter Notebook/ VS-code/ PyCharm: It is an open-source cross-platform integrated development environment (IDE) for scientific programming in the Python language. VS-code integrates with a number of prominent packages as well as another open-source software.

NumPy: NumPy was used for building the front-end part of the system.
Pandas: Pandas was used for the data preprocessing and statistical analysis of data.
Matplotlib: Matplotlib was used for the graphical representation of our prediction.

3.3 FUNCTIONAL REQUIREMENTS

- The system must provide the predicted weather.
- The system must have an easy-to-use interface for using the system for all the users.
- The admin must be able to update/modify the Dataset.
- The Dataset of the weather must be available for the system.

3.4 DATA COLLECTION AND PRE-PROCESSING

Data mining is a technique that changes raw data into a comprehensible format. Raw data (real world data) is always incomplete and cannot be sent through a model. Data mining process steps have been applied to pre-process the data and clean the collected raw weather data Understanding how the data is collected, stored, transformed, reported, and used is essential for the data mining Process.

3.4.1 Data collection

For predict the weather we have collected weather data. For the prediction model, we have used weather data. In raw weather data maximum temperature, minimum temperature (in degree Celsius), humidity, rainfall, evaporation, sunshine, wind gust, wind direction (9am), wind direction (3pm), wind speed, Air pressure (9am), Air pressure (3pm), cloud (9am), cloud (3pm) and temperature (9am), temperature (3pm) above all are the parameters. For weather prediction, we have used the Average temperature, Average Humidity, Average air pressure, Average wind and Events features. We have ignored less relevant features in the dataset for better model computation and prediction.

3.4.2. Data Pre-processing and Data Cleaning

The main challenge in weather prediction is poor data quality and selection. For this reason, we have used pre-process data carefully to obtain accurate and correct prediction results. In this phase unwanted data or noise is removed from the collected data set which is done by removing the unwanted attributes and keeping the most relevant attributes that help in better prediction. Another major issue is to be rectified the missing values in the collected data set. Missing values in the data set is filled by using various techniques.

Data mining is the process for extracting the useful data from dataset that will give us clean valuable dataset for model computation and better prediction. Most of the data mining algorithms would require data to be structured in a tabular format with records in rows and attributes in columns.

3.5 THE STEPS INVOLVED IN PREPROCESSING ARE

The data we have collected has many unwanted attributes which will not be needed in our project. Hence, we use the attributes which we need only.

3.5.1 Normalization

The data we collected from internet should be first normalized. Normalization refers to rescaling real valued numeric attributes into the rage or 0 and 1. After the data are filtered it is then normalized.

3.5.2 Machine Learning

Training a model is the process of iteratively improving your prediction equation by looping through the dataset multiple times, each time updating the weight and bias values in the direction indicated by the slope of the cost function (gradient). Training is complete when we reach an acceptable error threshold, or when subsequent training iterations fail to reduce our cost.

CHAPTER-4
ALGORITHMS

4.1 Back-Propagation Approach

The back propagation algorithm is used in layered feed-forward ANNs. It uses supervised learning, which means the model trains itself with the use of target output. For every set of input data, the target output is provided. The neural network model processes the input data with random values for weights and suitable activation function using one or more hidden layer in between and then produces the predicted output. This predicted output is then compared with the target output provided for same input dataset. Thus, error is calculated by subtracting predicted output from target output. Using this error, the weights are adjusted and again the entire process is repeated for multiple epochs until the error is minimal or in acceptable range. We start the training with random weights, and the goal is to adjust them so that the error will be minimal.

Input data is then pre-processed and cleaned. That means it is checked with any outlier and that is removed, missing values are entered, and data is checked if it is in the given range for the given parameter. Later ANN is designed with number of input and output nodes, hidden layers, activation function, and maximum number of epochs, weights, bias, goal and learning function. Neural network is trained with seventy percentages of the input data. Where the model is trained using this observed data to forecast the weather, followed by testing done using remaining thirty percentages of input data. Then the mean squared error and accuracy is calculated for the model by comparing the output of testing with target output. This model generates output in terms of minimum and maximum temperature of the day, relative humidity and rainfall.

Seventy percentages of the dataset will be used for training and the other thirty % of the dataset will be used for testing and validation. Hidden layers are required for processing nonlinear data. Better results can be achieved with high accuracy when learning rate is smaller but its performance is slower. Activation functions are applied on each neuron to get the output of neuron on a given input in the neural network. The sigmoid function is a special case of logistic function which has a sigmoid curve. The sigmoid transfer function can be used for hidden layers and for the output layer the linear transfer function can be used.

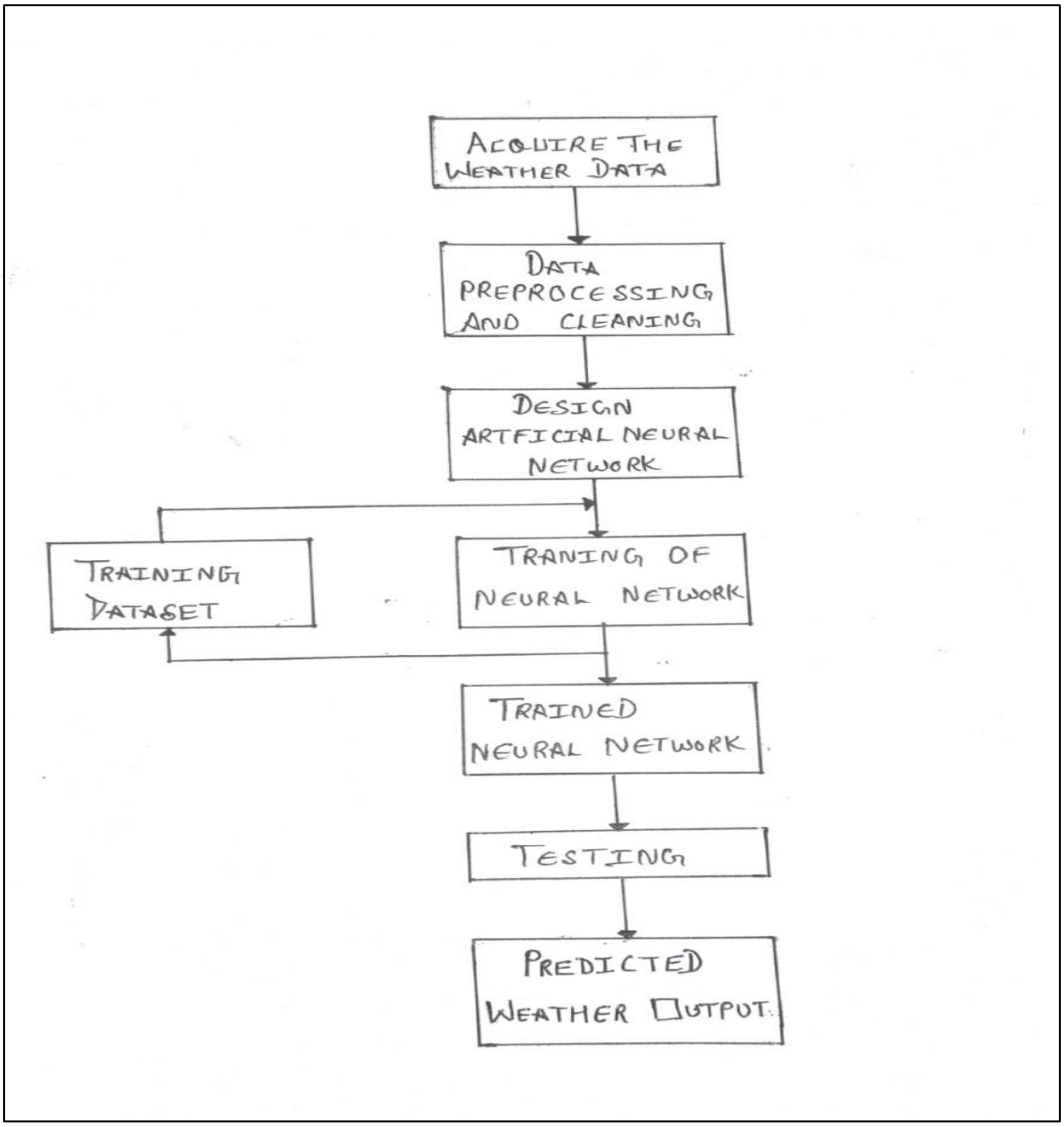


Fig-2: Back Propagation Approach

4.2 Classifier:

In data science, a classifier is a type of machine learning algorithm used to assign a class label to a data input. An example is an image recognition classifier to label an image (e.g., “car,” “truck,” or “person”).

4.2.1 MLP

We use MLP Classifier. MLP Classifier stands for Multi-layer Perceptron Classifier. Unlike other classification algorithms such as Support Vectors or Naive Bayes Classifier, MLP Classifier relies on an underlying Neural Network perform the task of classification.

To train a MLP network, the data should always be scaled because it is very sensitive to it. Our accuracy was found to be around 89%.

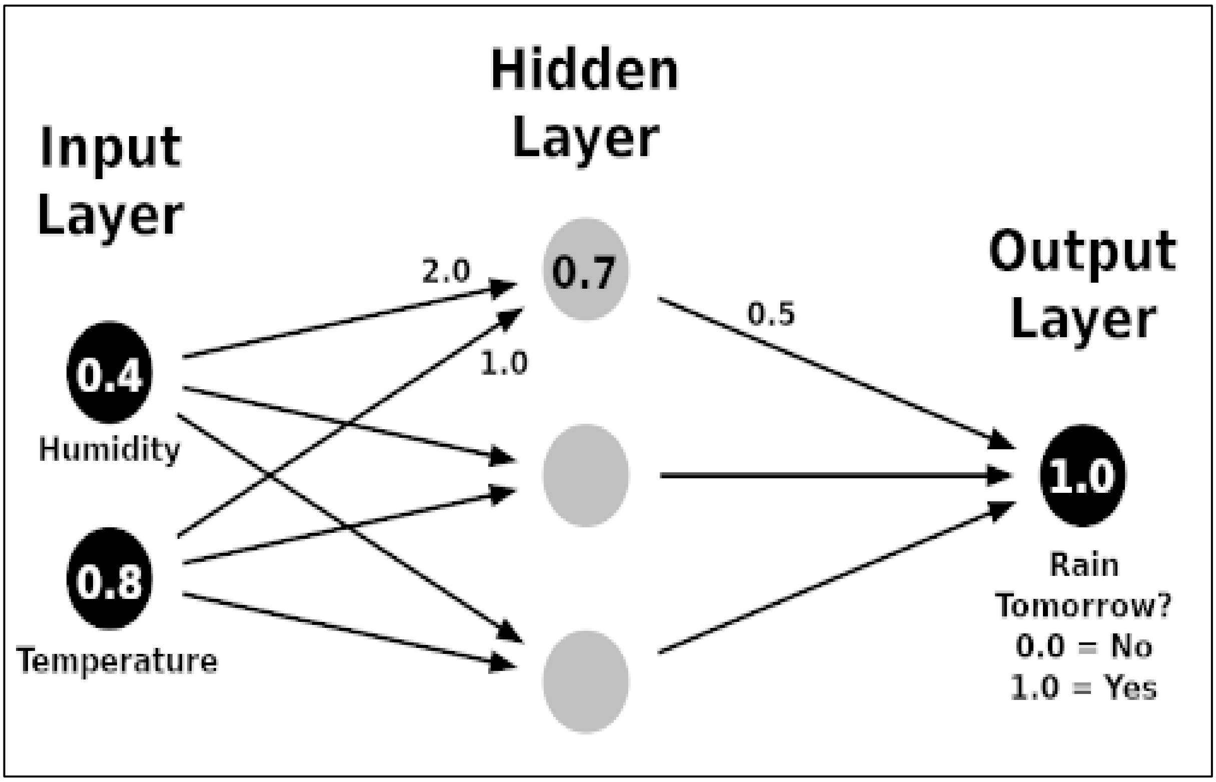


Fig-3: MLP Network

[Source: https://www.coursera.org/learn/predicting-weather-artificial-neural-networks]

4.2.2 SVM

We also decided to use another classifier for to check best suited classifier. So, we use SVM Classifier stands for Support Vector Machine. SVM is a supervised machine learning algorithm which can be used for classification or regression problems. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible outputs.

SVM finds the best way to classify the data based on the position in relation to a border between positive class and negative class. This border is known as the hyperplane which maximize the distance between data points from different classes.

Our accuracy was found to be around 90%.

4.2.3 KNN

KNN makes predictions using the dataset. Probabilities are made for new instance (x) by searching through the data set for the K most similar instances and predict the output variable for those K instances.

Our accuracy was found to be around 88%.

4.2.4 GaussianNB

Gaussian Naive Bayes algorithm is a particular type of NB algorithm. Naive Bayes Algorithm used when the features have continuous values. After completing the data preprocessing implement machine learning algorithm on it. We have built a Gaussian NB classifier. The classifier is trained using training data. After building a Gaussian NB classifier, our model is ready to make predictions using predict () method with test set features as parameters.

Our accuracy was found to be around 86%.

TABLE-1: Accuracy Comparison:

MODEL	ACCURACY (%)
MLP	0.8928
KNN	0.8878
SVC / SVM	0.9049
GaussianNB	0.8698

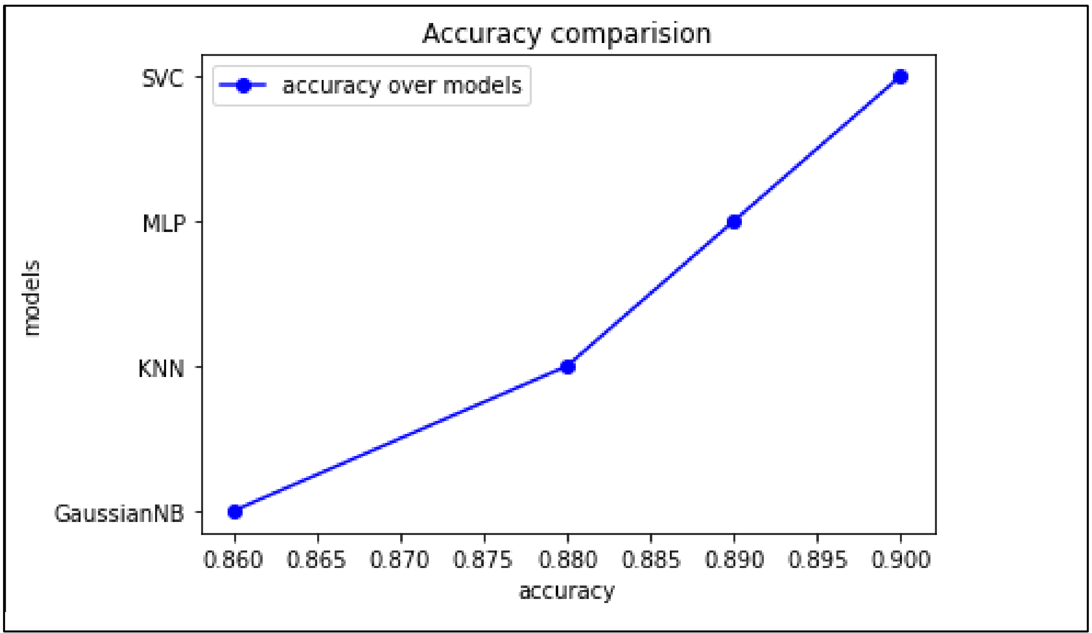


Fig-4: Accuracy score

TABLE-2: Precision and Recall value comparison:

MODEL	PRECESSION (0 / 1)	RECALL (0 / 1)
MLP	0.93 / 0.76	0.94 / 0.71
KNN	0.91 / 0.80	0.96 / 0.62
SVC / SVM	0.91 / 0.86	0.97 / 0.65
GaussianNB	0.95 / 0.65	0.88 / 0.82

CHAPTER-5

SOURCE CODE

Predicting the Weather with Neural Networks

In [1]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.preprocessing import StandardScaler
from sklearn.neural_network import MLPClassifier
from sklearn.model_selection import GridSearchCV
from sklearn.preprocessing import LabelEncoder
from sklearn.naive_bayes import GaussianNB
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
```

In [63]:
```python
df = pd.read_csv("weatherData.csv")
df.head(10)

```

Out[63]:

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	...	Humidity9am	Humidity3p
0	2.7	18.8	0.0	0.8	9.1	ENE	20.0	NaN	E	0	...	97.0	53
1	6.4	20.7	0.0	1.8	7.0	NE	22.0	ESE	ENE	6	...	80.0	39
2	6.5	19.9	0.4	2.2	7.3	NE	31.0	NaN	WNW	0	...	84.0	71
3	9.5	19.2	1.8	1.2	4.7	W	26.0	NNE	NNW	11	...	93.0	73
4	9.5	16.4	1.8	1.4	4.9	WSW	44.0	W	SW	13	...	69.0	57
5	0.7	15.9	6.8	2.4	9.3	NNE	24.0	ENE	NE	4	...	86.0	41
6	0.7	18.3	0.0	0.8	9.3	N	37.0	NE	NNE	15	...	72.0	36
7	3.2	20.4	0.0	1.4	6.9	NNW	24.0	NE	N	9	...	58.0	42
8	9.8	19.5	8.0	1.2	2.5	ESE	31.0	NaN	ESE	0	...	97.0	64

In [64]:
```python
wind_attributes = ['WindGustDir', 'WindDir9am', 'WindDir3pm']

# Cardinal direction to radians
dirs = ['N','NNE','NE','ENE','E','ESE','SE','SSE','S','SSW','SW','WSW','W','WNW','NW','NNW']
angles = np.arange(0.0, 2.0*np.pi, 2.0*np.pi / 16.0)
wind_angles = dict(zip(dirs, angles))
print(wind_angles)
for var in wind_attributes:
    df[var] = df[var].map(wind_angles)
    df[var + '_cos'] = np.cos(df[var])
    df[var + '_sin'] = np.sin(df[var])
    df = df.drop(columns=var)
df.head()

```

```
{'N': 0.0, 'NNE': 0.39269908169872414, 'NE': 0.7853981633974483, 'ENE': 1.1780972450961724, 'E': 1.5707963267948966, 'ESE': 1.9634954084936207, 'SE': 2.356194490192345, 'SSE': 2.748893571891069, 'S': 3.141592653589793, 'SSW': 3.5342917352885173, 'SW': 3.9269908169872414, 'WSW': 4.319689898685965, 'W': 4.71238898038469, 'WNW': 5.105088062083414, 'NW': 5.497787143782138, 'NNW': 5.890486225480862}
```

Out[64]:

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustSpeed	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	...	Temp9am	Temp3pm
0	2.7	18.8	0.0	0.8	9.1	20.0	0	7.0	97.0	53.0	...	8.5	18.1
1	6.4	20.7	0.0	1.8	7.0	22.0	6	9.0	80.0	39.0	...	11.1	19.7
2	6.5	19.9	0.4	2.2	7.3	31.0	0	4.0	84.0	71.0	...	12.1	17.7
3	9.5	19.2	1.8	1.2	4.7	26.0	11	6.0	93.0	73.0	...	13.2	17.7
4	9.5	16.4	1.8	1.4	4.9	44.0	13	17.0	69.0	57.0	...	15.9	16.0

5 rows × 24 columns

In [65]:
```python
df = df.dropna()

```

In [66]:
```
bools = ['RainToday', 'RainTomorrow']
for var in bools:
    df[var] = df[var].map({
        'Yes': 1,
        'No': 0
    })
df.head()
```

Out[66]:

	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustSpeed	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	...	Temp9am	Temp3pm
1	6.4	20.7	0.0	1.8	7.0	22.0	6	9.0	80.0	39.0	...	11.1	19.7
3	9.5	19.2	1.8	1.2	4.7	26.0	11	6.0	93.0	73.0	...	13.2	17.7
4	9.5	16.4	1.8	1.4	4.9	44.0	13	17.0	69.0	57.0	...	15.9	16.0
5	0.7	15.9	6.8	2.4	9.3	24.0	4	7.0	86.0	41.0	...	6.9	15.5
6	0.7	18.3	0.0	0.8	9.3	37.0	15	13.0	72.0	36.0	...	8.7	17.9

5 rows × 24 columns

In [67]:
```
y = df["RainTomorrow"]
X = df.drop(columns="RainTomorrow")
```

In [68]:
```
X_train, X_test, y_train, y_test = train_test_split(
    X,
    y,
    test_size=0.33,
    random_state=0
)
print('X_train', X_train.shape)
print('X_test', X_test.shape)
```

```
X_train (2026, 23)
X_test (999, 23)
```

In [81]:
```
scaler = StandardScaler()
scaler.fit(X_train)
X_train = scaler.transform(X_train)
X_test = scaler.transform(X_test)
```

In [82]:
```
print(X_train.shape)
```

```
(2026, 23)
```

In [84]:
```
print(y_train.shape)
```

```
(2026,)
```

In [86]:
```
print(X_test.shape)
```

```
(999, 23)
```

In [85]:
```
print(y_test.shape)
```

```
(999,)
```

GaussianNB

In [87]:
```
classifier=GaussianNB()
classifier.fit(X_train, y_train)
```

Out[87]:
```
GaussianNB()
```

In [88]:
```
classifier.score(X_train, y_train)
```

Out[88]:
```
0.8529121421520237
```

In [89]:
```
y_pred = classifier.predict(X_test)
print(accuracy_score(y_test, y_pred))
```

```
0.8698698698698699
```

MLP

In [97]:
```
classifier = MLPClassifier(
    hidden_layer_sizes=(50,50),
    max_iter=500,
    random_state=0
)
classifier.fit(X_train, y_train)
```

Out[97]:
```
MLPClassifier(hidden_layer_sizes=(50, 50), max_iter=500, random_state=0)
```

In [98]:
```
y_pred = classifier.predict(X_test)
print(accuracy_score(y_test, y_pred))
```

```
0.8928928928928929
```

In [ ]:

KNN

In [102]:
```
knn= KNeighborsClassifier()
knn.fit(X_train, y_train)
y_pred = knn.predict(X_test)
```

In [103]:
```
print(accuracy_score(y_test, y_pred))
```

```
0.8878878878878879
```

In [ ]:

SVC

In [105]:
```
svc = SVC()
svc.fit(X_train, y_train)
y_pred = svc.predict(X_test)
```

In [106]:
```
print(accuracy_score(y_test, y_pred))
```

```
0.9049049049049049
```

In [107]:
```
from sklearn.metrics import classification_report
print(classification_report(y_test,y_pred))
```

```
              precision    recall  f1-score   support

           0       0.91      0.97      0.94       792
           1       0.86      0.65      0.74       207

    accuracy                           0.90       999
   macro avg       0.89      0.81      0.84       999
weighted avg       0.90      0.90      0.90       999
```

In [ ]:

In [78]:
```
parameters = {
    'hidden_layer_sizes': (
        (2,), (10,), (50,50),
    )
}
nn = MLPClassifier(max_iter=2000, random_state=0)
gridsearch = GridSearchCV(nn, parameters, cv=3)
gridsearch.fit(X_train, y_train)
```

Out[78]:
```
GridSearchCV(cv=3, estimator=MLPClassifier(max_iter=2000, random_state=0),
             param_grid={'hidden_layer_sizes': ((2,), (10,), (50, 50))})
```

In [79]:
```
print(gridsearch.cv_results_['params'])
print(gridsearch.cv_results_['mean_test_score'])
```

```
[{'hidden_layer_sizes': (2,)}, {'hidden_layer_sizes': (10,)}, {'hidden_layer_sizes': (50, 50)}]
[0.79812477 0.89832712 0.88646431]
```

In [80]:
```
best_nn = gridsearch.best_estimator_
y_pred = best_nn.predict(X_test)
print(accuracy_score(y_test, y_pred))
```

```
0.8958958958958959
```

CHAPTER-6

LIMITATIONS & FUTURE SCOPE

6.1 LIMITATIONS

A tiny disturbance in one layer, even one as tiny as a butterfly flapping its wings, can have a domino effect, affecting the other layers and snowballing into radically different weather patterns. All that variation and uncertainty is why there's a limit to how far out we can meaningfully predict the weather.

6.1.1 MLP The perceptron can only learn simple problems. It can place a hyperplane in pattern space and move the plane until the error is reduced. Unfortunately, this is only useful if the problem is linearly separable.

6.1.2 SVM algorithm is not suitable for large data sets. SVM does not perform very well when the data set has more noise i.e., target classes are overlapping. In cases where the number of features for each data point exceeds the number of training data samples, the SVM will underperform. The problem with hard margin SVM is that if the dataset is not linearly separable then it becomes impossible to classify the data (that is, no separating hyperplane is found) and the presence of noise or outliers greatly affect the margin.

6.2 FUTURE SCOPE

The accuracy from weather predicting model using ANN and Back-propagation Algorithm is more than another Statistical Model. An extension to this Technology can be done using any of the other technique instead of Data mining and Different Algorithm.

CHAPTER-7

SUMMARY & CONCLUSION

The different methods for weather forecasting are reviewed. ANN with back propagation is recommended for weather Forecasting. ANN with back propagation uses an iterative process of training where, it repeatedly compares the observed output with targeted output and calculates the error. This error is used to read just the values of weights and bias to get an even better output. Hence this method tries to minimize the error. Thus, Artificial Neural network with Back propagation algorithm seems to be most appropriate method for forecasting weather accurately.

The weather prediction done using BP algorithm and Naïve Bayes algorithm are very essential for improving the future performance for the people. For predicting the weather, the linear regression algorithm and Naïve Bayes algorithm was applied to the datasets of the weather. We made a model to predict the weather using some selected input variables collected from Google. The problem with current weather scenario is that we are not able to prepare our self and not able to do some important works. So, for knowing the weather scenario at high accuracy considering every factor that affects in the weather scenario, this model is created.

CHAPTER-8

REFERENCES

1. S.S. De, University of Kolkata, Artificial Neural Network Based Prediction of Max. & Min. Temperature in the Summer-Monsoon month over India, Applied Physics Research, Vol.1, No.2, Nov-2009.
2. Mohsen Hayati& Zahra Mohebi, Temperature Forecasting Based on Neural Network Approach, World Applied Science Journal 2(6) 613-620, 2007, ISSN 818-4952 ©IDOSI Publications 2007.
3. Dr.S. SanthoshBaboo& I. KadarShereef, an efficient weather forecasting system using Artificial Neural network, International Journal of Environment Science & Development, W1.1, No.4, October-2010, ISSN: 2010-0264
4. Jabani B and Priyanka Sebastian. (2014). Analysis on The Weather Forecasting and Techniques.
   Samenow and Frirz. (2015). Issues with weather prediction.
   Gould & Bryan. (2017). Current weather prediction.
5. Y. Radhika& M. Shashi, Atmospheric Temperature prediction using Support Vector Machine., International Journal of Computer Theory & Engineering, Vol.1. No.1.April 2009 1793-8209
6. P. Lynch, (2008) “The Origins of Computer Weather Prediction and Climatic Modeling", Journal based on Computational Physics, vol. 227, 3431-3444
7. Cohen, J., Cohen P., West, S.G., & Aiken, L.S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences. (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum Associates