A Thesis/Project/Dissertation Report

on

Weather Prediction Using ML

Submitted in partial fulfillment of the
requirement for the award of the degreeof

Master of Computer Applications

GALGOTIAS UNIVERSITY
(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of
Name of Supervisor: Mr. Christian
Designation

Submitted By
Ayushman Rout
20SCSE1280031

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING /
DEPARTMENT OF COMPUTERAPPLICATION
GALGOTIAS UNIVERSITY, GREATER NOIDA

INDIA
MONTH, YEAR

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"CAPS…."** in partial fulfillment of the requirements for the award of the ~~Name of Degree~~ submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of month, Year to Month and Year, under the supervision of Name… Designation, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Supervisor Name

Mr.Christian

Designation

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of Deepanshu Agrawal:20scse1280020, Ritik Kumar:20scse1280030, Rahul Kumar:20scse1280009 has been held on Weather Prediction Using ML and his/her work is recommended for the award of B.Tech in Artificial Intelligence and Data Science.

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: November, 2013

Place: Greater Noida

Abstract

Paragraph-1 Existing Problem- 5-8 Lines
Paragraph-2 Proposed Solution- 6-8 Lines
Paragraph-3 Tools and Technology Used- 3-5 lines
Paragraph -4 Results and output- 3-5 lines
Paragraph-5 Conclusion and Future Scope – 2-4 Lines

Table of Contents

Table For Student

S.no	Name	Email	Phone Number
1	Ayushman	ayushman.20scse1280031@galgotiasuniversity.edu.in	

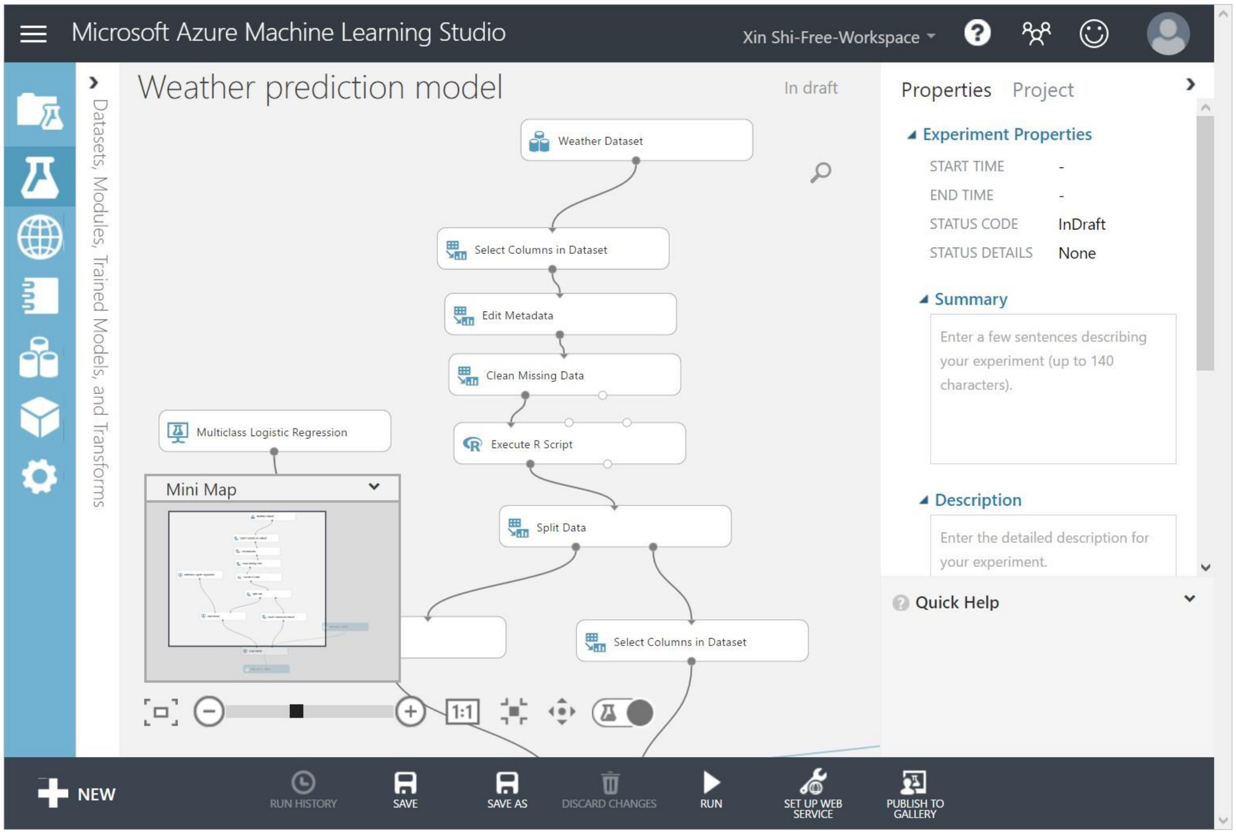
Table for Faculty

S.No	Name	Email	Phone no.	Sign
1	Mr. Christian			

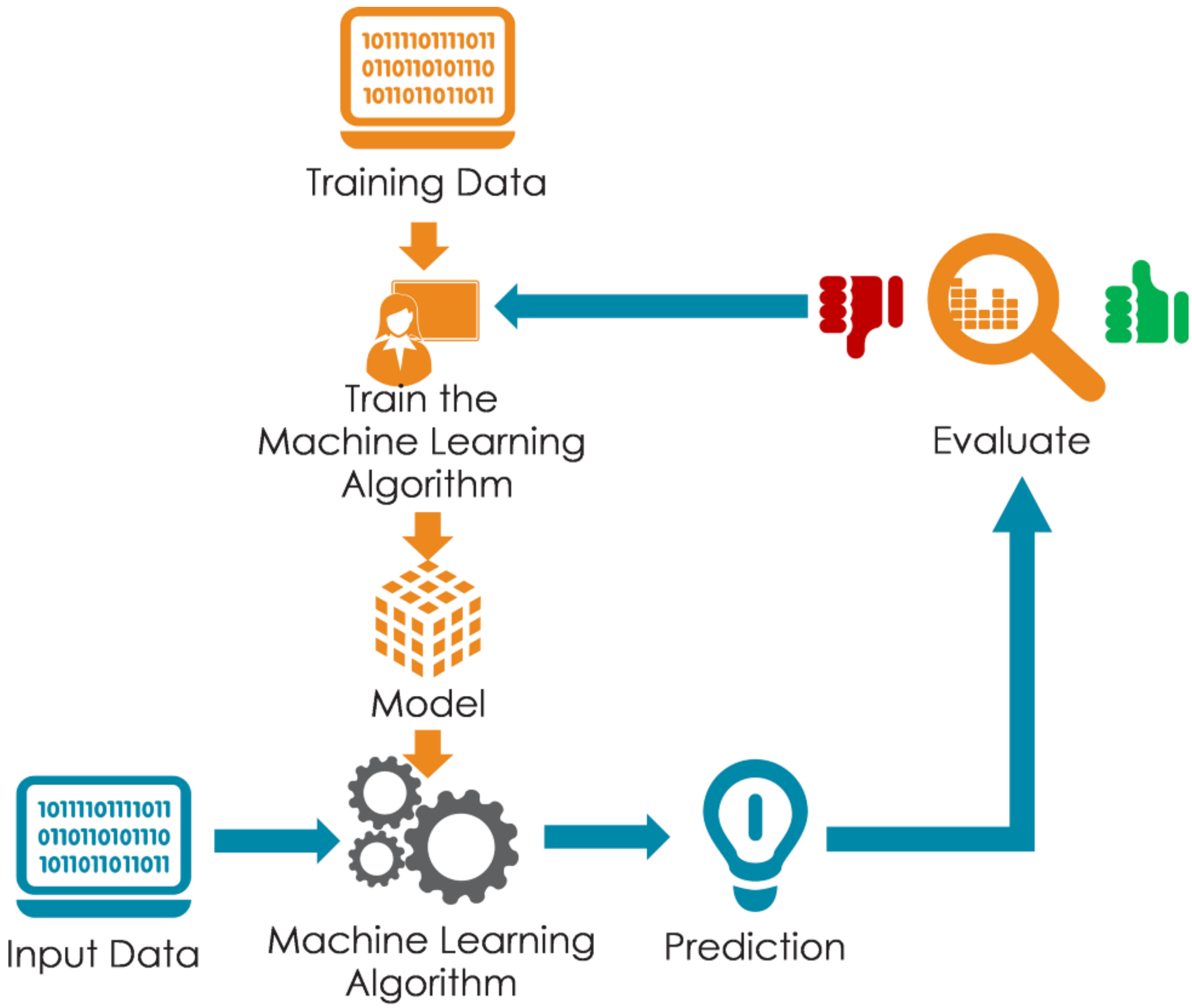
Acronyms

B.Tech.	Bachelor of Technology
M.Tech.	Master of Technology
BCA	Bachelor of Computer Applications
MCA	Master of Computer Applications
B.Sc. (CS)	Bachelor of Science in Computer Science
M.Sc. (CS)	Master of Science in Computer Science
SCSE	School of Computing Science and Engineering

UML Diagram


Microsoft Azure Machine Learning Studio
Xin Shi-Free-Workspace
Weather prediction model
In draft
Datasets, Modules, Trained Models, and Transforms
Weather Dataset
Select Columns in Dataset
Edit Metadata
Clean Missing Data
Multiclass Logistic Regression
Execute R Script
Split Data
Select Columns in Dataset
Mini Map
Properties
Project
Experiment Properties
START TIME -
END TIME -
STATUS CODE InDraft
STATUS DETAILS None
Summary
Enter a few sentences describing your experiment (up to 140 characters).
Description
Enter the detailed description for your experiment.
Quick Help
NEW
RUN HISTORY
SAVE
SAVE AS
DISCARD CHANGES
RUN
SET UP WEB SERVICE
PUBLISH TO GALLERY

Data Flow Diagram



Abstract

Climate conducts a completely critical function in many key production sectors, e.g. farming. Climate change with high charging these days, which is why old weather forecasts are getting closer and less powerful and continue to be annoying. Miles are therefore very important to decorate and modify the weather forecast model. those predictions affect the country's financial system and people's lives. A system of information and statistics analysis algorithms has been used that includes a wooded area used for weather forecasting. The weather is one of the highest natural barriers in all parts of our lives in the world, we need to look at the weather including temperature, rain, humidity and other protection. remarkable. The purpose of our artwork is to format effective weather forecasts. Earth's climate will change over a long period of time and also what kind of impact it will have on the lives of future generations. Our predictive nature of end-of-life climates offers an excellent desire to provide information as a way to allow stadium insurers to make an informed wish for the future of the world. Our approach greatly enhances the model in a positive way to govern the state of staff inconsistencies and inequalities and performs its function of accurately predicting the weather.

CHAPTER-1

1.1 Introduction

Weather forecasting is the task of predicting the state of the atmosphere at a future time and a specified location. Traditionally, this has been done through physical simulations in which the atmosphere is modeled as a fluid. The present state of the atmosphere is sampled, and the future state is computed by numerically solving the equations of fluid dynamics and thermodynamics. However, the system of ordinary differential equations that govern this physical model is unstable under perturbations, and uncertainties in the initial measurements of the atmospheric conditions and an incomplete understanding of complex atmospheric processes restrict the extent of accurate weather forecasting to a ten day period, beyond which weather forecasts are significantly unreliable. Machine learning, on the contrary, is relatively robust to perturbations and doesn't require a complete understanding of the physical processes that govern the atmosphere. Therefore, machine learning may represent a viable alternative to physical models in weather forecasting. Two machine learning algorithms were implemented: linear regression and a variation of functional regression. A corpus of historical weather data for Stanford, CA was obtained and used to train these algorithms. The input to these algorithms was the weather data of the past two days, which include the maximum temperature, minimum temperature, mean humidity, mean atmospheric pressure, and weather classification for each day. The output was then the maximum and minimum temperatures for each of the next seven days.

1.2 Formulation of Problem

Predictability inaccuracies are due to the prevailing weather conditions, the high calculation power required to solve atmospheric calculations, the error involved in estimating the initial conditions, and an incomplete understanding of atmospheric processes. Therefore, the predictions are less accurate as the difference between the current time and the time the forecast is made (the range of the forecast) increases. The use of ensembles and a harmonious model helps to minimize error and select the possible outcome.

There are various ways to end climate use. Weather warnings are important predictions because they are used to protect health and property. Temperatures based on temperatures and rainfall are important for agriculture, so for traders in the middle of the commodity markets.

CHAPTER-2

Literature Survey

There are many research papers that have been published related to predicting the weather [9]. A paper was published on 'The Weather Forecast Using Data Mining Research Based on Cloud Computing' This paper proposes a modern method to develop a service oriented architecture for the weather information systems which forecast weather using these data mining techniques. This can be carried out by using Artificial Neural Network and Decision tree Algorithms and meteorological data collected in Specific time. Algorithm has presented the best results to generate classification rules for the mean weather variables. The results showed that these data mining techniques can be enough for weather forecasting [9]. Another paper was published on 'Analysis on The Weather Forecasting and Techniques' where they decided that artificial neural network and concept of fuzzy logic provides a best solution and prediction comparatively [10]. They decided to take temperature, humidity, pressure, wind and various other attributes into consideration [3].

Another research paper titled 'Issues with weather prediction' discussed the major problems with weather prediction [11]. 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 [4]. 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 [12]. 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.

CHAPTER-3

Working on Project

Related works included many different and interesting techniques to try to perform weather forecasts. While much of current forecasting technology involves simulations based on physics and differential equations, many new approaches from artificial intelligence used mainly machine learning techniques, mostly neural networks while some drew on probabilistic models such as Bayesian networks. Out of the three papers on machine learning for weather prediction we examined, two of them used neural networks while one used support vector machines. Neural networks seem to be the popular machine learning model choice for weather forecasting because of the ability to capture the non-linear dependencies of past weather trends and future weather conditions, unlike the linear regression and functional regression models that we used. This provides the advantage of not assuming simple linear dependencies of all features over our models. Of the two neural network approaches, one [3] used a hybrid model that used neural networks to model the physics behind weather forecasting while the other [4] applied learning more directly to predicting weather conditions. Similarly, the approach using support vector machines [6] also applied the classifier directly for weather prediction but was more limited in scope than the neural network approaches. Other approaches for weather forecasting included using Bayesian networks. One interesting model [2] used Bayesian networks to model and make weather predictions but used a machine learning algorithm to find the most optimal Bayesian networks and parameters which was quite computationally expensive because of the large amount of different dependencies but performed very well. Another approach [1] focused on a more specify

to minimize is
J(θ) = 1 2 Xm i=1 khθ(x (i)) − y (i) k 2 1

where m is the number of training examples. Letting $X \in R^{m\times 9}$ be defined such that $X_{ij} = x^{(i)}_j$ and $Y \in R^{m\times 14}$ be defined such that $Y_{ij} = y^{(i)}_j$, the value of θ that minimizes the cost in equation 1 is

$$\theta = (X^T X)^{-1} X^T Y. \quad \text{---- } 2$$

The second algorithm that was used was a variation of functional regression, which searches for historical weather patterns that are most similar to the current weather patterns, then predicts the weather based upon these historical patterns. Given a sequence of nine consecutive days, define its spectrum f as follows. Let $f(1), f(2) \in R^5$ be the feature vectors for the first day and the second day, respectively. For i in the range 3 to 9, let $f(i) \in R^2$ be a vector containing the maximum temperature and the minimum temperature for the i-th day in the sequence. Then define a metric on the space of spectra

$$d(f_1, f_2) = \sum_{j=1}^{2} \left[w_1 1[f_1(j)_1 \neq f_2(j)_1] + \sum_{k=2}^{5} w_k \left(f_1(j)_k - f_2(j)_k\right)^2 \right] \quad \text{----} 3$$

where w is a weight vector that assigns weights to each feature. Since the first feature is the weather classification and the difference between classifications is meaningless, the squared difference has been replaced by an indicator function of whether the classifications are different. Define a kernel

$$ker(t) = \max\{1 - t, 0\} \quad 4$$

and let $neigh_k(f)$ denote the k indices $i \in \{1, \ldots, m\}$ of the k spectra in the training set that are the closest to f with respect to the metric d. That is,

$$d(f^{(i)}, f) < d(f^{(j)}, f) \quad 5$$

for all $i \in neigh_k(f)$ and $j \notin neigh_k(f)$, and $|neigh_k(f)| = k$. Furthermore, define h = max i∈{1,...,m} d(f (i) , f---5

Then, given the values f(1), f(2) of the first two days of a spectrum f, the remainder of the spectrum f(i) for i in the range 3 to 9 can be predicted as
ˆf(i) = P j∈neighk(f) ker(d(f (j) , f)/h)f (j) (i) P j∈neighk(f) ker(d(f (j) , f)/h) .---6

initial estimates of w1 = 20 and w4 = w5 = 0. w4 = 0 was the optimum weight of the mean humidity presumably since humidity correlates poorly with the maximum temperature and the minimum temperature, and humidity would be a more useful determinant of precipitation. w5 = 0 turned out to be the optimum weight of the mean atmospheric pressure since there were only small deviations in the atmospheric pressure which did not appear to be correlated with the maximum temperature and the minimum temperature. With this in mind, the mean humidity and the mean atmospheric pressure were removed as features. The hyperparameter of the number of neighbours k was then chosen in a similar manner, with an exhaustive grid search over both constant values and values proportional to the data set size. Values of k in the range 5-50 in increments of 5 and values of k proportional to the data set size with proportionality constant in the range 0.05-0.50 in increments of 0.05 were considered. Taking k proportional to the data set size greatly outperformed taking k to be constant, and the optimum proportionality constant was 0.10. w1 and k were then fine-tuned together with one final exhaustive grid search, taking w1 from the range 15-25 in increments of 1, and the proportionality constant of k from the range 0.05-0.15 in increments of 0.05. This yielded a final value of w1 = 18 and k = 0.095|D|, where |D| is the number of data points.

Methods

The first algorithm that was used was linear regression, which seeks to predict the high and low temperatures as a linear combination of the features. Since linear regression cannot be used with classification data, this algorithm did not use the weather classification of each day. As a result, only eight features were used: the maximum temperature, minimum temperature, mean humidity, and mean atmospheric pressure for each of the past two days. Therefore, for the i-th pair of consecutive days, x (i) ∈ R 9 is a nine-dimensional feature vector, where x0 = 1 is defined as the intercept term. There are 14 quantities to be predicted for each pair of consecutive days: the high and low temperatures for each of the next seven days. Let y (i) ∈ R 14 denote the 14-dimensional vector that contains these quantities for the i-th pair of consecutive days. The prediction of y (i) given x (i) is hθ(x (i)) = θ T x, where θ ∈ R 9×14. The cost function that linear regression seeks.
The error of the estimator ˆf is defined to be Error = X 9 i=3 k ˆf(i) − f(i)k 2 . (8)

A more useful error that will be used in lieu of this is the root mean square (rms) error, which is defined to be

Errorrms = vuutX 9 i=3 k ˆf(i) − f(i)k 2 14 , (9)

and provides the standard deviation of the individual error terms.

Research Methodology

There are two types of category in Machine Learning: supervised learning and unsupervised learning. In this work we have carried out research on supervised learning. Classification is a supervised learning approach which is based on training sample set. Machine Learning tool

is used to build predictive models. We have implemented four classifications which are experimentally implemented and compared against each other. These Classification algorithms are Naive Bayes Bernoulli, Logistic Regression, Naive Bayes Gaussian and KNN. The methodology consists following stages for each study period data of weather parameters which are (i) Computation of descriptive statistics.
(ii) Development of weather forecasting models and comparison of their predictive ability. (iii) Identification of precise and reliable weather forecasting model.

Naive Bayes Bernoulli Algorithm

Naive Bayes classifier gives more accurate results when we use it for textual data analysis. Bayes approach is a method to classify events based on occurrence probability or not happening [6]. Naive Bayes shows proper results using native attribute when it receives primitive practice. Bayes' theorem: -

Probability of B occurring given evidence A has already occurred

Probability of A occurring

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Probability of A occurring given evidence B has already occurred

Probability of B occurring

CODE


```
{
  "nbformat": 4,
  "nbformat_minor": 0,
  "metadata": {
    "colab": {
      "name": "Untitled1.ipynb",
      "provenance": [],
      "collapsed_sections": []
    },
    "kernelspec": {
      "name": "python3",
      "display_name": "Python 3"
    },
    "language_info": {
      "name": "python"
    }
  },
  "cells": [
    {
      "cell_type": "code",
      "metadata": {
        "id": "1MA02_D_KaI7"
      },
      "source": [
        "import pandas as pd\n",
        "from sklearn.preprocessing import LabelEncoder\n",
        "from sklearn.naive_bayes import GaussianNB"
      ],
      "execution_count": 34,
      "outputs": []
    },
    {
      "cell_type": "code",
```

```
    "metadata": {
     "colab": {
      "base_uri": "https://localhost:8080/",
      "height": 480
     },
     "id": "NUx0pLQvPy9o",
     "outputId": "58756338-c07f-4807-bdf1-2e4e1bea31bd"
    },
    "source": [
     "#Reading CSV files\n",
     "df=pd.read_csv('Weather prediction using machine learning.CSV')\n",
     "df"
    ],
    "execution_count": 2,
    "outputs": [
     {
      "output_type": "execute_result",
      "data": {
       "text/html": [
        "<div>\n",
        "<style scoped>\n",
        "    .dataframe tbody tr th:only-of-type {\n",
        "    vertical-align: middle;\n",
        "    }\n",
        "\n",
        "    .dataframe tbody tr th {\n",
        "    vertical-align: top;\n",
        "    }\n",
        "\n",
        "    .dataframe thead th {\n",
        "    text-align: right;\n",
        "    }\n",
        "</style>\n",
        "<table border=\"1\" class=\"dataframe\">\n",
```

```
"  <thead>\n",
"    <tr style=\"text-align: right;\">\n",
"    <th></th>\n",
"      <th>Outlook</th>\n",
"      <th>Temp</th>\n",
"      <th>Humidity</th>\n",
"      <th>Windy</th>\n",
"    <th>Play</th>\n", "
    </tr>\n",
"  </thead>\n",
"   <tbody>\n",
"  <tr>\n",
"      <th>0</th>\n",
"      <td>Rainy</td>\n",
"      <td>Hot</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"      <td>no</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>1</th>\n",
"      <td>Rainy</td>\n",
"      <td>Hot</td>\n",
"      <td>High</td>\n",
"      <td>t</td>\n",
"      <td>no</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>2</th>\n",
"      <td>Overcast</td>\n",
"      <td>Hot</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
```

```
"    <tr>\n",
"      <th>3</th>\n",
"      <td>Sunny</td>\n",
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
     </tr>\n",
"    <tr>\n",
"      <th>4</th>\n",
"      <td>Sunny</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
     </tr>\n",
"    <tr>\n",
"      <th>5</th>\n",
"      <td>Sunny</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>t</td>\n",
"      <td>no</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>6</th>\n",
"      <td>Overcast</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>t</td>\n",
"    <td>yes</td>\n", "
     </tr>\n",
"    <tr>\n",
"      <th>7</th>\n",
"      <td>Rainy</td>\n",
```

```
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"  <td>f</td>\n",
"      <td>no</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>8</th>\n",
"      <td>Rainy</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
"    <tr>\n",
"      <th>9</th>\n",
"      <td>Sunny</td>\n",
"      <td>Mild</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
"    <tr>\n",
"      <th>10</th>\n",
"      <td>Rainy</td>\n",
"      <td>Mild</td>\n",
"      <td>Normal</td>\n",
"      <td>t</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
"    <tr>\n",
"      <th>11</th>\n",
"      <td>Overcast</td>\n",
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"      <td>t</td>\n",
```

```
"    <td>yes</td>\n", "
     </tr>\n",
"    <tr>\n",
"      <th>12</th>\n",
"      <td>Overcast</td>\n",
"      <td>Hot</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
     </tr>\n",
"    <tr>\n",
"      <th>13</th>\n",
"      <td>Sunny</td>\n",
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"      <td>t</td>\n",
"      <td>no</td>\n",
"    </tr>\n",
"  </tbody>\n",
"</table>\n",
"</div>"
],
"text/plain": [
"     Outlook Temp Humidity Windy Play\n",
"0   Rainy     Hot    High    f   no\n",
"1      Rainy  Hot    High    t   no\n",
"2   Overcast  Hot    High    f yes\n",
"3      Sunny Mild    High    f yes\n",
"4      Sunny Cool  Normal    f yes\n",
"5      Sunny Cool  Normal    t   no\n",
"6   Overcast Cool  Normal    t yes\n",
"7      Rainy Mild    High    f   no\n",
"8      Rainy Cool  Normal    f yes\n",
"9      Sunny Mild  Normal    f yes\n",
"10     Rainy Mild  Normal    t yes\n",
```

```
        "11  Overcast  Mild     High     t yes\n",
        "12  Overcast   Hot   Normal     f yes\n",
        "13     Sunny  Mild     High     t   no"
       ]
      },
      "metadata": {},
      "execution_count": 2
    }
  ]
},
{
  "cell_type": "code",
  "metadata": {
    "id": "y4C8SQ30Qm-4"
  },
  "source": [
    "#Encoding the strings to Numericals\n",
    "Numerics=LabelEncoder()"
  ],
  "execution_count": 13,
  "outputs": []
},
{
  "cell_type": "code",
  "metadata": {
    "colab": {
      "base_uri": "https://localhost:8080/"
    },
    "id": "LK3hCPscQtIy",
    "outputId": "9569059a-a652-4225-ab18-561d955d9f35"
  },
  "source": [
    "#Dropping the target variable and make it is as newframe\n",
    "inputs=df.drop('Play',axis='columns')\n",
    "target = df['Play']\n",
```

"target"
],
"execution_count": 10,
"outputs": [
{
"output_type": "execute_result",
"data": {
"text/plain": [
"0   no\n",
"1   no\n",
"2   yes\n",
"3   yes\n",
"4   yes\n",
"5   no\n",
"6   yes\n",
"7   no\n",
"8   yes\n",
"9   yes\n",
"10   yes\n",
"11   yes\n",
"12   yes\n",
"13   no\n",
"Name: Play, dtype: object"
]
},
"metadata": {},
"execution_count": 10
}
]
},
{
"cell_type": "code",
"metadata": {
"colab": {
"base_uri": "https://localhost:8080/",

"height": 480
},
"id": "1t4fycSfQ1RI",
"outputId": "da50dc75-0669-472d-c657-57b7096cb404"
},
"source": [
"#Creating the new dataframe\n",
"inputs['Outlook_n']=Numerics.fit_transform(inputs['Outlook'])\n"
,
"inputs['Temp_n']=Numerics.fit_transform(inputs['Temp'])\n",
"inputs['Humidity_n']=Numerics.fit_transform(inputs['Humidity'])
\n",
"inputs['Windy_n']=Numerics.fit_transform(inputs['Windy'])\n",
"inputs"
],
"execution_count": 17,
"outputs": [
{
"output_type": "execute_result",
"data": {
"text/html": [
"<div>\n",
"<style scoped>\n",
"   .dataframe tbody tr th:only-of-type {\n",
"   vertical-align: middle;\n",
"   }\n",
"\n",
"   .dataframe tbody tr th {\n",
"   vertical-align: top;\n",
"   }\n",
"\n",

```
"    .dataframe thead th {\n",
"    text-align: right;\n",
"    }\n",
"</style>\n",
"<table border=\"1\" class=\"dataframe\">\n",
"  <thead>\n",
"    <tr style=\"text-align: right;\">\n",
"    <th></th>\n",
"      <th>Outlook</th>\n",
"      <th>Temp</th>\n",
"      <th>Humidity</th>\n",
"      <th>Windy</th>\n",
"      <th>Outlook_n</th>\n",
"      <th>Temp_n</th>\n",
"      <th>Humidity_n</th>\n",
"      <th>Windy_n</th>\n",
"    </tr>\n",
"  </thead>\n",
"   <tbody>\n",
"  <tr>\n",
"      <th>0</th>\n",
"      <td>Rainy</td>\n",
"      <td>Hot</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"  </tr>\n",
"  <tr>\n",
"  <th>1</th>\n",
"      <td>Rainy</td>\n",
"  <td>Hot</td>\n",
"      <td>High</td>\n",
```

```
"      <td>t</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>2</th>\n",
"      <td>Overcast</td>\n",
"      <td>Hot</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>3</th>\n",
"      <td>Sunny</td>\n",
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>4</th>\n",
"      <td>Sunny</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"      <td>2</td>\n",
```

```
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>5</th>\n",
"      <td>Sunny</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>t</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>6</th>\n",
"      <td>Overcast</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>t</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>7</th>\n",
"      <td>Rainy</td>\n",
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
```

```
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>8</th>\n",
"      <td>Rainy</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"  </tr>\n",
"  <tr>\n",
"  <th>9</th>\n",
"      <td>Sunny</td>\n",
"      <td>Mild</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"  </tr>\n",
"  <tr>\n",
"  <th>10</th>\n",
"      <td>Rainy</td>\n",
"      <td>Mild</td>\n",
"      <td>Normal</td>\n",
"      <td>t</td>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"  </tr>\n",
```

```
"    <tr>\n",
"      <th>11</th>\n",
"      <td>Overcast</td>\n",
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"      <td>t</td>\n",
"      <td>0</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"  </tr>\n",
"  <tr>\n",
"  <th>12</th>\n",
"      <td>Overcast</td>\n",
"      <td>Hot</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"  </tr>\n",
"  <tr>\n",
"  <th>13</th>\n",
"      <td>Sunny</td>\n",
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"      <td>t</td>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"  </tr>\n",
"  </tbody>\n",
"</table>\n",
```

"</div>"
],
"text/plain": [
"   Outlook   Temp Humidity Windy   Outlook_n Temp_n
Humidity_n Windy_n\n",
"0   Rainy   Hot   High   f   1   1   0   0\n",
"1   Rainy   Hot   High   t   1   1   0   1\n",
"2   Overcast   Hot   High   f   0   1   0   0\n",
"3   Sunny Mild   High   f   2   2   0   0\n",
"4   Sunny   Cool   Normal   f   2   0   1
0\n",
"5   Sunny   Cool   Normal   t   2   0   1
1\n",
"6   Overcast   Cool   Normal   t   0   0   1
1\n",
"7   Rainy Mild   High   f   1   2   0   0\n",
"8   Rainy   Cool   Normal   f   1   0   1
0\n",
"9   Sunny   Mild   Normal   f   2   2   1
0\n",
"10   Rainy   Mild   Normal   t   1   2   1
1\n",
"11   Overcast   Mild   High   t   0   2   0
1\n",
"12   Overcast   Hot   Normal   f   0   1   1
0\n",
"13   Sunny Mild   High   t   2   2   0   1"
]
},
"metadata": {},
"execution_count": 17
}
]
},
{

```
    "cell_type": "code",
    "metadata": {
     "colab": {
      "base_uri": "https://localhost:8080/",
      "height": 480
     },
     "id": "Ewqqy6OvQ5tB",
     "outputId": "b02175fd-6445-4783-917a-8e847a43b436"
    },
    "source": [
     "#Dropping the string values\n",
"inputs_n=inputs.drop(['Outlook','Temp','Humidity','Windy'],axis
='columns')\n",
     "inputs_n"
    ],
    "execution_count": 18,
    "outputs": [
     {
      "output_type": "execute_result",
      "data": {
       "text/html": [
        "<div>\n",
        "<style scoped>\n",
        "    .dataframe tbody tr th:only-of-type {\n",
        "    vertical-align: middle;\n",
        "    }\n",
        "\n",
        "    .dataframe tbody tr th {\n",
        "    vertical-align: top;\n",
        "    }\n",
        "\n",
        "    .dataframe thead th {\n",
        "    text-align: right;\n",
        "    }\n",
```

```
"</style>\n",
"<table border=\"1\" class=\"dataframe\">\n",
"  <thead>\n",
"    <tr style=\"text-align: right;\">\n",
"    <th></th>\n",
"      <th>Outlook_n</th>\n",
"      <th>Temp_n</th>\n",
"      <th>Humidity_n</th>\n",
"      <th>Windy_n</th>\n",
"    </tr>\n",
"  </thead>\n",
"   <tbody>\n",
"  <tr>\n",
"      <th>0</th>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"  </tr>\n",
"  <tr>\n",
"      <th>1</th>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"  </tr>\n",
"  <tr>\n",
"      <th>2</th>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"  </tr>\n",
"  <tr>\n",
"  <th>3</th>\n",
```

```
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>4</th>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>5</th>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>6</th>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>7</th>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"    <th>8</th>\n",
```

```
"      <td>1</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>9</th>\n",
"      <td>2</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>10</th>\n",
"      <td>1</td>\n",
"      <td>2</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>11</th>\n",
"      <td>0</td>\n",
"      <td>2</td>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>12</th>\n",
"      <td>0</td>\n",
"      <td>1</td>\n",
"      <td>1</td>\n",
"      <td>0</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>13</th>\n",
```

```
       "     <td>2</td>\n",
       "     <td>2</td>\n",
       "     <td>0</td>\n",
       "     <td>1</td>\n",
       "  </tr>\n",
       "  </tbody>\n",
       "</table>\n",
       "</div>"
      ],
      "text/plain": [
       "    Outlook_n  Temp_n  Humidity_n  Windy_n\n",
       "0           1       1           0        0\n",
       "1           1       1           0        1\n",
       "2           0       1           0        0\n",
       "3           2       2           0        0\n",
       "4           2       0           1        0\n",
       "5           2       0           1        1\n",
       "6           0       0           1        1\n",
       "7           1       2           0        0\n",
       "8           1       0           1        0\n",
       "9           2       2           1        0\n",
       "10          1       2           1        1\n",
       "11          0       2           0        1\n",
       "12          0       1           1        0\n",
       "13          2       2           0        1"
      ]
     },
     "metadata": {},
     "execution_count": 18
    }
   ]
  },
  {
   "cell_type": "code",
   "metadata": {
```

```
    "colab": {
      "base_uri": "https://localhost:8080/"
    },
    "id": "py9BEivwVQNj",
    "outputId": "43d4f456-8139-43a2-b7e1-1a2382c766bb"
   },
   "source": [
    "#Applying the Gaussian naivebayes\n",
    "Classifier=GaussianNB()\n",
    "Classifier.fit(inputs_n,target)"
   ],
   "execution_count": 22,
   "outputs": [
    {
     "output_type": "execute_result",
     "data": {
      "text/plain": [
       "GaussianNB(priors=None, var_smoothing=1e-09)"
      ]
     },
     "metadata": {},
     "execution_count": 22
    }
   ]
  },
  {
   "cell_type": "code",
   "metadata": {
    "colab": {
     "base_uri": "https://localhost:8080/"
    },
    "id": "kLNNhW-3VlnL",
    "outputId": "18cb8e37-efe2-4b72-bf27-d0d882e0d72a"
   },
   "source": [
```

"GaussianNB()"
],
"execution_count": 21,
"outputs": [
{
"output_type": "execute_result",
"data": {
"text/plain": [
"GaussianNB(priors=None, var_smoothing=1e-09)"
]
},
"metadata": {},
"execution_count": 21
}
]
},
{
"cell_type": "code",
"metadata": {
"colab": {
"base_uri": "https://localhost:8080/"
},
"id": "DvxvTiLHVooC",
"outputId": "24e79fbe-ea86-4274-9cd5-ccf57f4ff806"
},
"source": [
"#85% accuracy\n",
"Classifier.score(inputs_n,target)"
],
"execution_count": 23,
"outputs": [
{
"output_type": "execute_result",
"data": {
"text/plain": [

```
        "0.8571428571428571"
       ]
      },
      "metadata": {},
      "execution_count": 23
     }
    ]
   },
   {
    "cell_type": "code",
    "metadata": {
     "colab": {
      "base_uri": "https://localhost:8080/"
     },
     "id": "o5PpwqYyV22B",
     "outputId": "04c0525e-b452-40cd-ba39-96df4ca304af"
    },
    "source": [
     "0.8571428571428571"
    ],
    "execution_count": 24,
    "outputs": [
     {
      "output_type": "execute_result",
      "data": {
       "text/plain": [
        "0.8571428571428571"
       ]
      },
      "metadata": {},
      "execution_count": 24
     }
    ]
   },
   {
```

"cell_type": "code",
"metadata": {
"colab": {
"base_uri": "https://localhost:8080/"
},
"id": "z6NHbHHpV8oq",
"outputId": "6f86bd1c-c30e-4a16-acbd-91581c1ad67f"
},
"source": [
"#Predicton\n",
"Classifier.predict([[0,0,1,1]])"
],
"execution_count": 31,
"outputs": [
{
"output_type": "execute_result",
"data": {
"text/plain": [
"array(['yes'], dtype='<U3')"
]
},
"metadata": {},
"execution_count": 31
}
]
},
{
"cell_type": "code",
"metadata": {
"colab": {
"base_uri": "https://localhost:8080/",
"height": 480
},
"id": "WqXcAzyTQ9Si",
"outputId": "a1a30bea-b306-47e2-88bd-4471e510afe0"

```
},
"source": [
 "df"
],
"execution_count": 29,
"outputs": [
 {
  "output_type": "execute_result",
  "data": {
   "text/html": [
    "<div>\n",
    "<style scoped>\n",
    "   .dataframe tbody tr th:only-of-type {\n",
    "   vertical-align: middle;\n",
    "   }\n",
    "\n",
    "   .dataframe tbody tr th {\n",
    "   vertical-align: top;\n",
    "   }\n",
    "\n",
    "   .dataframe thead th {\n",
    "   text-align: right;\n",
    "   }\n",
    "</style>\n",
    "<table border=\"1\" class=\"dataframe\">\n",
    "  <thead>\n",
    "   <tr style=\"text-align: right;\">\n",
    "   <th></th>\n",
    "     <th>Outlook</th>\n",
    "     <th>Temp</th>\n",
    "     <th>Humidity</th>\n",
    "     <th>Windy</th>\n",
    "   <th>Play</th>\n", "
       </tr>\n",
    "  </thead>\n",
```

```
"  <tbody>\n",
"  <tr>\n",
"      <th>0</th>\n",
"      <td>Rainy</td>\n",
"      <td>Hot</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"      <td>no</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>1</th>\n",
"      <td>Rainy</td>\n",
"      <td>Hot</td>\n",
"      <td>High</td>\n",
"      <td>t</td>\n",
"      <td>no</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>2</th>\n",
"      <td>Overcast</td>\n",
"      <td>Hot</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
     </tr>\n",
"    <tr>\n",
"      <th>3</th>\n",
"      <td>Sunny</td>\n",
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
     </tr>\n",
"    <tr>\n",
"      <th>4</th>\n",
```

```
"      <td>Sunny</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
"    <tr>\n",
"      <th>5</th>\n",
"      <td>Sunny</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>t</td>\n",
"      <td>no</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>6</th>\n",
"      <td>Overcast</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
"      <td>t</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
"    <tr>\n",
"      <th>7</th>\n",
"      <td>Rainy</td>\n",
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"      <td>f</td>\n",
"      <td>no</td>\n",
"    </tr>\n",
"    <tr>\n",
"      <th>8</th>\n",
"      <td>Rainy</td>\n",
"      <td>Cool</td>\n",
"      <td>Normal</td>\n",
```

```
"      <td>f</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
"    <tr>\n",
"      <th>9</th>\n",
"      <td>Sunny</td>\n",
"      <td>Mild</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
"    <tr>\n",
"      <th>10</th>\n",
"      <td>Rainy</td>\n",
"      <td>Mild</td>\n",
"      <td>Normal</td>\n",
"      <td>t</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
"    <tr>\n",
"      <th>11</th>\n",
"      <td>Overcast</td>\n",
"      <td>Mild</td>\n",
"      <td>High</td>\n",
"      <td>t</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
"    <tr>\n",
"      <th>12</th>\n",
"      <td>Overcast</td>\n",
"      <td>Hot</td>\n",
"      <td>Normal</td>\n",
"      <td>f</td>\n",
"    <td>yes</td>\n", "
    </tr>\n",
```

```
      "    <tr>\n",
      "      <th>13</th>\n",
      "      <td>Sunny</td>\n",
      "      <td>Mild</td>\n",
      "      <td>High</td>\n",
      "      <td>t</td>\n",
      "      <td>no</td>\n",
      "    </tr>\n",
      "  </tbody>\n",
      "</table>\n",
      "</div>"
     ],
     "text/plain": [
      "     Outlook Temp Humidity Windy Play\n",
      "0   Rainy     Hot     High     f   no\n",
      "1      Rainy   Hot     High     t   no\n",
      "2   Overcast   Hot     High     f yes\n",
      "3      Sunny  Mild     High     f yes\n",
      "4      Sunny  Cool   Normal     f yes\n",
      "5      Sunny  Cool   Normal     t   no\n",
      "6   Overcast  Cool   Normal     t yes\n",
      "7      Rainy  Mild     High     f   no\n",
      "8      Rainy  Cool   Normal     f yes\n",
      "9      Sunny  Mild   Normal     f yes\n",
      "10     Rainy  Mild   Normal     t yes\n",
      "11  Overcast  Mild     High     t yes\n",
      "12  Overcast   Hot   Normal     f yes\n",
      "13     Sunny  Mild     High     t   no"
     ]
    },
    "metadata": {},
    "execution_count": 29
   }
  ]
}
```

  ]
}

Chapter-5

Conclusion And Future Works

Both linear regression and functional regression were outperformed by professional weather forecasting services, although the discrepancy in their performance decreased significantly for later days, indicating that over longer periods of time, our models may outperform professional ones. Linear regression proved to be a low bias, high variance model whereas functional regression proved to be a high bias, low variance model. Linear regression is inherently a high variance model as it is unstable to outliers, so one way to improve the linear regression model is by collection of more data. Functional regression, how-ever, was high bias, indicating that the choice of model was poor, and that its predictions cannot be improved by further collection of data. This bias could be due to the design choice to forecast weather based upon the weather of the past two days, which may be too short to capture trends in weather that functional regression requires. If the forecast were instead based upon the weather of the past four or five days, the bias of the functional regression model could likely be reduced. However, this would require much more computation time along with retraining of the weight vector w, so this will be deferred to future work.

Reference

[1] Olaiya, Folorunsho and Adesesan Barnabas Adeyemo, “Application of data mining techniques in weather prediction and climate change studies”, International Journal of Information Engineering and Electronic Business, Vol. 4, No. 1, pp. 51, 2012.

[2] Sawaitul, D. Sanjay, K. P. Wagh and P. N. Chatur, “Classification and prediction of future weather by using back propagation algorithm-an approach”, International Journal of Emerging Technology and Advanced Engineering, Vol. 2, No. 1, pp. 110-113, 2012.

[3] P. Langley, W. Iba and K. Thompson, “An analysis of Bayesian Classifiers”, in Proceedings of the Tenth National Conference on Artificial Intelligence, San Jose, CA, 1992.

[4] Kotu, Vijay and BalaDeshpande, Predictive analytics and data mining: concepts and practice with rapidminer, Morgan Kaufmann, 2014.

[5] D. Hand, H. Mannila and P. Smyth, “Principles of data mining”, MIT, 2001.

[6] A. Mccallum and K. Nigam, “A Comparison of Event Models for Naive Bayes Text Classification”, Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98)-Workshop on Learning for Text Categorization, pp. 41-48, 1998.

[7] R. O. Duda and P. E. Hart, Pattern classification and scene analysis, John Wiley and Sons, 1973.

[8] Mohita Anand Sharma1 and Jai Bhagwan Singh2” COMPARATIVE STUDY OF RAINFALL FORECASTING MODELS”, Department of Mathematics and Statistics, School of Basic and Applied Science, Shobhit University, Meerut, Uttar Pradesh, 250110, India.

[9] Razeef Mohd1 , Muheet Ahmed Butt2 and MajidZaman Baba3”Comparative Study of Rainfall Prediction Modeling Techniques”, Student, 2 Scientist-D, Department of Computer Science, 3 Scientist-D, Directorate of IT&SS, University of Kashmir, Jammu and Kashmir, India.

[10] Mark Holmstrom, Dylan Liu, Christopher Vo” Machine Learning Applied to Weather Forecasting”, Stanford University (Dated: December 15, 2016).

[11]E. B. Abrahamsen, O. M. Brastein, B. Lie∗ ”Machine Learning in Python for Weather Forecast based on Freely Available Weather Data”, Department of Electrical Engineering, Information Technology and Cybernetics University of SouthEastern Norway, N-3918 Porsgrunn.

[12]A H M Jakaria, Md Mosharaf Hossain, Mohammad Ashiqur Rahman” Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee”, Tennessee Tech University Cookeville, Tennessee.

[13]Ahmed, Bilal, “Predictive capacity of meteorological data: Will it rain tomorrow?”,Science and Information Conference (SAI), 2015, IEEE, 2015.