Weather Forecasting Using Machine Learning Algorithm

Nitin Singh
Department of Electronics and Communication Engineering Jaypee Institute of Information Technology, NOIDA, India
singhnitin0523@gmail.com

Saurabh Chaturvedi
Department of Electronics and Communication Engineering Jaypee Institute of Information Technology, NOIDA, India
saurabh.chaturvedi@jiit.ac.in

Shamim Akhter
Department of Electronics and Communication Engineering Jaypee Institute of Information Technology, NOIDA, India
shamim.akhter@jiit.ac.in

***Abstract* - The activities of many primary sectors depend on the weather for production, e.g. farming. The climate is changing at a drastic rate nowadays, which makes the old weather prediction methods less effective and more hectic. To overcome these difficulties, the improved and reliable weather prediction methods are required. These predictions affect a nation's economy and the lives of people. To develop a weather forecasting system that can be used in remote areas is the main motivation of this work. The data analytics and machine learning algorithms, such as random forest classification, are used to predict weather conditions. In this paper, a low-cost and portable solution for weather prediction is devised.**

***Keywords* - Weather forecast, machine learning, Raspberry Pi, Python, confusion matrix, sensor.**

I. INTRODUCTION

The process of predicting weather conditions for future is called weather forecasting. This paper presents the rain prediction with the use of real-time data of temperature, humidity, and pressure using various sensors. Machine learning provides a capability to the systems to learn and improve from experience without being programmed by a user [1], [2]. With the introduction of machine learning concept, there is a great ease for data analysis and prediction. Machine learning does not require to understand the physical processes that control the atmosphere but uses the past data to predict future data. Therefore, this process may be used as a weather forecasting method [3].

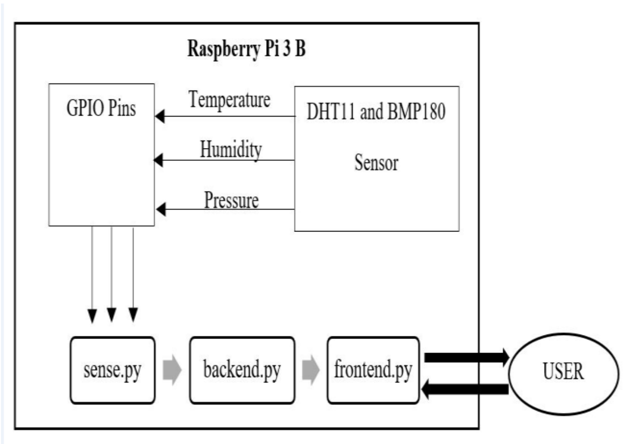
The random forest classification, a machine learning algorithm, uses an ensemble learning method in which two or more machine learning models are combined to form a single machine learning model. It operates by making multiple decision trees while training the dataset and outputs the mode of classification of the different trees.

The rest of the paper is organized as follows: Section II discusses the overview of the implemented system. Section III presents the description of the machine learning approach. Section IV explains the details of the hardware and software involved in this work. The results are presented in Section V, and the paper is concluded in Section VI.

II. OVERVIEW OF IMPLEMENTED SYSTEM

In the proposed system, an application has been developed on Raspberry Pi 3 B. This application gets the real-time data from humidity-temperature and pressure sensors to predict the possibility of rain on the present day. Fig. 1 demonstrates the block-level diagram of the system. The GPIO pins of Raspberry Pi communicate between the hardware (sensors) and software (application).

The application is made from three different modules: sense.py, backend.py, and frontend.py. The sense.py module gets the sensor value when prompted by a user. It uses `Adafruit_DHT` and `Adafruit_BMP` libraries to collect the data through GPIO pins. The backend.py module includes the machine learning model which is trained according to the random forest classification. It collects the data from sense.py as a NumPy array of size three. The data include temperature, humidity, and pressure. The system produces the result in the form of 0 or 1, where 1 points out that rain will happen and 0 points out that there will be no rain. The frontend.py module provides a graphical user interface (GUI) to the application. This part of application interacts with the user. It is connected to the sense.py and backend.py modules.



978-1-5386-9436-7/19/$31.00 ©2019 IEEE

Fig. 1. Block diagram of the system.

The snap shot of the developed GUI is shown in Fig. 2.

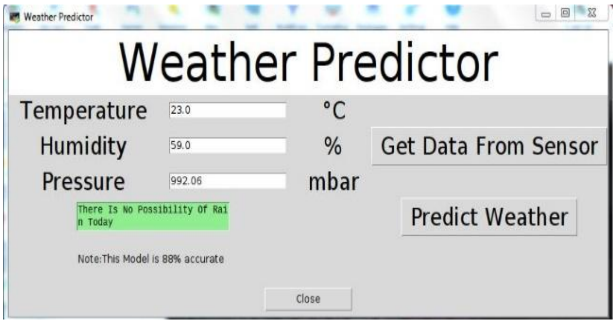


Fig. 2. GUI of the developed application.

III. MACHINE LEARNING APPROACH

The block diagram for the machine learning-based prediction system is depicted in Fig. 3.

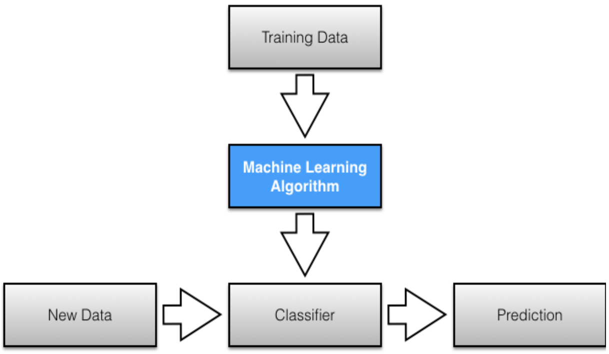


Fig. 3. Machine learning-based prediction system.

Several steps are followed for implementing the machine learning approach in this work. The steps are as follows:

A. Data Collection

The data set used for this work is available at [4]. It is Delhi's weather data for the past 20 years.

B. Data Pre-processing and Feature Extraction

All attributes of the data set are processed once, and the most prominent features of the data set are extracted to be used further.

C. Deriving Another Data set

From the primary data set, another data set was derived in this work by averaging the temperature, humidity, pressure, and rain for a given date.

D. Training and Testing the Machine Learning Model

The random forest classification model is trained with 75% of the derived data set and 25% is used as the test set for the model.

The raw data set had the details of temperature, humidity, pressure, and rain for each hour of a day. As shown in Table I, the new data set is derived by averaging the values.

TABLE I. Derived data set from raw data by averaging

	A	B	C	D	E
1	datetime	temp_avg	hum_avg	pressure_avg	rain
2	01-11-1996	22.33333	52.9166667	1011.33333	0
3	02-11-1996	22.91667	48.625	1009.83333	0
4	03-11-1996	21.79167	55.9583333	1010.5	0
5	04-11-1996	22.72222	48.0555556	1011.33333	0
6	05-11-1996	27.8	29.4	1011.8	0
7	06-11-1996	21.9375	47.875	1011.8125	0
8	07-11-1996	21.75	48	1013.41667	0
9	08-11-1996	21.43547	45.7462284	1014.70833	0
10	09-11-1996	19.90909	47.1818182	1013.86364	0
11	10-11-1996	20.45455	44.1363636	1012.18182	0
12	11-11-1996	22.08696	37.7391304	1012.47826	0
13	12-11-1996	23.45833	39.3333333	1012.20833	0
14	13-11-1996	23.4544	51.5178035	1012.43478	0
15	14-11-1996	21.88889	45.7222222	1016.88889	0
16	15-11-1996	17.61111	53.4444444	1019.05556	0
17	16-11-1996	17.56597	40.6322491	1016.18182	0
18	17-11-1996	17.45833	41.125	1017.41667	0
19	18-11-1996	18.63158	39.2105263	1017.21053	0
20	19-11-1996	19.21739	46.2173913	1016.47826	0
21	20-11-1996	19.63636	61.3636364	1016.95455	0
22	21-11-1996	17.68182	60.0909091	1016.63636	0
23	22-11-1996	14.88889	56.4444444	1015.77778	0
24	23-11-1996	16.21739	52.8695652	1015.47826	0
25	24-11-1996	16.125	51.7083333	1013.875	0
26	25-11-1996	16.90909	47.4090909	1013.90909	0
27	26-11-1996	19.41092	47.1264992	1014.52174	0
28	27-11-1996	17.52381	54.9047619	1014.28571	0
29	28-11-1996	17.31579	45.6842105	1014.47368	0

IV. HARDWARE AND SOFTWARE USED

This section deals with the details of the hardware and software used in the work.

A. Details of Used Hardware

In this work, the Raspberry Pi 3 B board, DHT11 humidity and temperature sensor, and BMP180 pressure sensor are used.

Raspberry Pi Board: Raspberry Pi is a computer on a single board. It has advanced in many aspects, including memory and connectivity of different peripherals. The Raspberry Pi 3 B is one of the latest versions of Raspberry Pi which includes inbuilt features like wireless LAN and Bluetooth [5]. The block diagram for Raspberry Pi is displayed in Fig. 4.

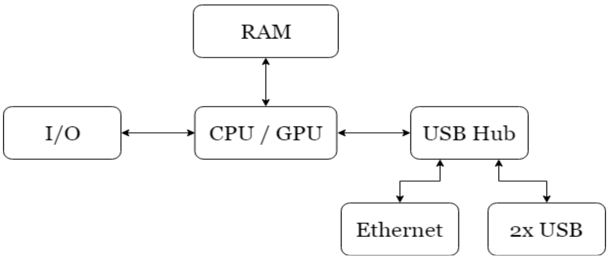


Fig. 4. Block diagram for Raspberry Pi [5].

The Raspberry Pi 3 B consists of 1.2 GHz Broadcom BCM2837 system on chip and 1 GB RAM. It has 40-pin extended GPIO for connecting sensors and other devices. There are 4 USB ports, an HDMI port, a CSI camera port for connecting a Pi camera, a DSI display port for connecting a Raspberry Pi touchscreen display, and a micro SD slot for loading the operating system and storing data. The Raspberry Pi 3 B board used is shown in Fig. 5.

Fig. 5. Raspberry Pi 3 B model [6].

For using Raspberry Pi board, the following components are required:

- Prepared operating system SD card
- USB keyboard
- Display (with HDMI, DVI, or composite input)
- Power supply

Many programming languages can be used in Raspberry Pi board, such as Python, C, C++, and Java. Python is widely used in Raspberry Pi.

DHT11 Humidity and Temperature Sensor: The DHT11 sensor has the temperature and humidity sensing capabilities. It provides a digital signal output. The sensor is integrated with an 8-bit microcontroller. This high-performance sensor comprises a resistive element and a sensing element for negative temperature coefficient measuring devices [7]. The DHT11 sensor is shown in Fig. 6.

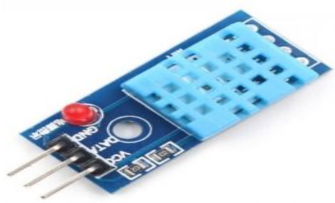


Fig. 6. DHT11 humidity and temperature sensor [8].

BMP180 Barometric Pressure Sensor: This high-precision sensor measures the barometric pressure or atmospheric pressure [9]. The sensor gives a digital output. It works on inter-integrated circuit protocol. The BMP180 sensor is illustrated in Fig. 7.

Fig. 7. BMP180 barometric pressure sensor [10].

B. Details of Used Software

The Raspbian operating system (OS), Spyder 3, and Python 3 are used in the system.

Raspbian OS: Raspbian is a computer OS for Raspberry Pi which is Debian based [11]. There are several versions of Raspbian, including Raspbian Stretch and Raspbian Jessie. It uses the Pi improved X-window environment, lightweight (PIXEL) as its main desktop environment.

Spyder 3: Spyder is an open source integrated development environment for Python, which is significantly used for machine learning [12]. Spyder integrates with many key packages in the Python stack, such as NumPy, SciPy, Matplotlib, Pandas, IPython, SymPy, and Cython.

Python 3: Python is a high-level programming language. The automatic memory management and dynamic type system are its salient features. The object-oriented, functional, imperative, and procedural programming attributes are supported by Python [13]. It has a significant number of libraries for different purposes. Python can be easily integrated with many different languages, such as SQL, MATLAB, C, and Java.

V. RESULTS

The random forest classification provides deep insights into the data dependencies. The main result expected in this work is about whether rain will happen or not on a particular day according to the sensor data. In Fig. 8, the graph shows the dependency of rain on temperature, humidity, and pressure. In this graph, 0, 1, and 2 indicate temperature, humidity, and pressure, respectively. It can be inferred from the graph that humidity is the most important factor while predicting rain followed by temperature and pressure.

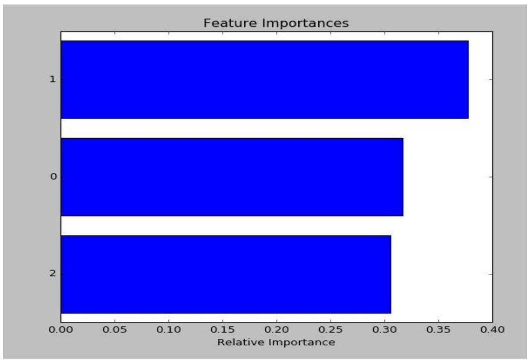


Fig. 8. Features of importance of the developed model.

In this work, out of total 7339 data sets, the data sets are divided into 75% as the training set and 25% as the test set, i.e. 5504 data for training and 1835 for testing. The splitting between the training set and test set is random.

From sklearn.metrics, an inbuilt function, `accuracy_score()`, is used to check the accuracy of this model. An accuracy of 87.90% could be achieved in this machine learning model.

The resultant confusion matrix of the test set is shown in Table II. The size of the test set is 1835, out of which the number 1491 represents the count of correct predictions that rain will not happen and 122 is the number of correct predictions that rain will happen. Therefore, total 1613 correct predictions are made out of 1835 test data.

TABLE II. Confusion matrix of test data set

Confusion matrix	Predicted: No	Predicted: Yes
Actual: No	1491	56
Actual: Yes	166	122

VI. CONCLUSION

The prime objective of this work is to develop a low-cost, reliable, and efficient weather forecasting application using the machine learning concept in Python on Raspberry Pi board.

REFERENCES

[1] N. Hasan, M. T. Uddin, and N. K. Chowdhury, "Automated weather event analysis with machine learning," in *Proc. IEEE 2016 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, 2016, pp. 1-5.

[2] L. L. Lai, H. Braun, Q. P. Zhang, Q. Wu, Y. N. Ma, W. C. Sun, and L. Yang, "Intelligent weather forecast," in *Proc. IEEE 2004 International Conference on Machine Learning and Cybernetics*, 2004, pp. 4216-4221.

[3] A. G. Salman, B. Kanigoro, and Y. Heryadi, "Weather forecasting using deep learning techniques," in *Proc. IEEE 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2015, pp. 281-285.

[4] Delhi Weather Data. [Online]. Available: https://www.kaggle.com/mahirkukreja/delhi-weather-data/home

[5] Raspberry Pi. [Online]. Available: https://en.wikipedia.org/wiki/Raspberry_Pi

[6] Raspberry Pi 3 Model B. [Online]. Available: https://images-na.ssl-images-amazon.com/images/

[7] DHT11 Temperature and Humidity Sensor. [Online]. Available: https://www.robot-r-us.com/vmchk/sensor-temp/humid/dht11-temperature-and-humidity-sensor.html

[8] DHT11 Temperature and Humidity Module. [Online]. Available: https://www.makerfabs.com/DHT11-Temperature-and-Humidity-Module.html

[9] Sensors - Pressure, Temperature and Altitude with the BMP180. [Online]. Available: https://thepihut.com/blogs/raspberry-pi-tutorials/18025084-sensors-pressure-temperature-and-altitude-with-the-bmp180

[10] BMP180 Barometric Pressure Sensor Module. [Online]. Available: https://ae01.alicdn.com/kf/HTB108MyX2JNTKJjSspoq6A6mpXai/BMP180-GY-68-GY68-3-3V-5V-BMP-180-Temperature-Pressure-Sensor-Module-Barometric-IIC-I2C.jpg

[11] Raspbian. [Online]. Available: https://en.wikipedia.org/wiki/Raspbian

[12] Spyder (software). [Online]. Available: https://en.wikipedia.org/wiki/Spyder_(software)

[13] Python (programming language). [Online]. Available: https://en.wikipedia.org/wiki/Python_(programming_language)