RAINFALL PREDICTION

A Project Report submitted in the partial fulfilment of the requirements for the award of the degree

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by

N.Bhanu Sravya (19471A0539)
P.Rafiya (19471A0545)
K.V.S.Abhigna (19471A0528)

Under the esteemed guidance of

Ms.A.Thanuja M.Tech
Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

NARASARAOPETA ENGINEERING COLLEGE:NARASARAOPETA

(AUTONOMOUS)

Accredited by NAAC with A+ Grade and NBA under Cycle -1 Approved by AICTE, New Delhi, Permanently Affiliated to JNTUK, Kakinada

KOTAPPAKONDA ROAD, YALAMANDA VILLAGE, NARASARAOPET-522601

2022-2023

NARASARAOPETA ENGINEERING COLLEGE: NARASARAOPETA
(AUTONOMOUS)
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the main project entitled "RAINFALL PREDICTION" is a bonafide Work done by **N.Bhanu Sravya (19471A0539), P.Rafiya (19471A0545), K.V.S.Abhigna (19471A0528)** in partial fulfilment of the requirements for the award of the degree of **BACHELOR OF TECHNOLOGY** in the Department of **COMPUTER SCIENCE AND ENGINEERING** during 2022-2023.

PROJECT GUIDE
A.Thanuja M.Tech
Asst. Prof.

PROJECT CO-ORDINATOR
Mrs. M. Sireesha M.Tech.,(Ph.D)
Assoc. Prof.

HEAD OF THE DEPARTMENT
Dr. S. N. TirumalaRao **M.Tech, Ph.D.**

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We wish to express our thanks to carious personalities who are responsible for the completion of the project. We are extremely thankful to our beloved chairperson **Mr. M.V.Koteswara Rao**, B.sc who took keen interest on us in every effort throughout this course. We owe out gratitude to our principal **Dr.M.Sreenivasa Kumar**, **M.Tech, Ph.D(UK),MISTE,FIE(1)** his kind attention and valuable guidance throughout thecourse.

We express our deep felt gratitude to **Dr.S.N.Tirumala Rao**, M.Tech, Ph.D. H.O.D,CSE department and our guide **Ms.A.Thanuja**, M.Tech, Assistant Professor of CSE department whose valuable guidance and unstinting encouragement enable us to accomplish our project successfully in time.

We extend our sincere thanks to **Mrs.M.Sireesha** **M.Tech, (Ph.D)** Associate Professor and Coordinator of the project for extending his encouragement. Their profound knowledge and willingness have been a constant source of inspiration for us throughoutthis project work.

We extend our sincere thanks to all other teaching and non-teaching staff to department for their cooperation and encouragement during our B.Tech degree. we have no words to acknowledge the warm affection, constant inspiration and encouragement that we receive from our parents.

We affectionately acknowledge the encouragement received from our friends and those who involved in giving valuable suggestions had clarifying out doubts, which had really helped us in successfully completing our project.

By

N.Bhanu Sravya **(19471A0539)**

P.Rafiya **(19471A0545)**

K.V.S.Abhigna **(19471A0528)**

ABSTRACT

Rainfall prediction is one of the challenging and uncertain tasks which has a significant impact on human society. Timely and accurate predictions can help to proactively reduce human and financial loss. This study presents a set of experiments which involve the use of prevalent machine learning techniques to build models to predict whether it is going to rain tomorrow or not based on weather data for that particular day in major cities of Australia. This comparative study is conducted concentrating on three aspects: modeling inputs, modeling methods, and pre-processing techniques. The results provide a comparison of various evaluation metrics of these machine learning techniques and their reliability to predict the rainfall by analyzing the weather data. Logistic Regression, Random Forest Classifier, Gradient Boosting, KNN, Decision Tree, Adaboost Classifier, and Catboost Classifier are the algorithms used to make the predictions.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

INSTITUTE VISION AND MISSION

INSTITUTION VISION

To emerge as a Centre of excellence in technical education with a blend of effective student centric teaching learning practices as well as research for the transformation of lives and community,

INSTITUTION MISSION

M1: Provide the best class infra-structure to explore the field of engineering and research

M2: Build a passionate and a determined team of faculty with student centric teaching, imbibing experiential, innovative skills

M3: Imbibe lifelong learning skills, entrepreneurial skills and ethical values in students for addressing societal problems

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VISION OF THE DEPARTMENT

To become a centre of excellence in nurturing the quality Computer Science & Engineering professionals embedded with software knowledge, aptitude for research and ethical values to cater to the needs of industry and society.

MISSION OF THE DEPARTMENT

The department of Computer Science and Engineering is committed to

M1: Mould the students to become Software Professionals, Researchers and Entrepreneurs by providing advanced laboratories.

M2: Impart high quality professional training to get expertize in modern software tools and technologies to cater to the real time requirements of the Industry.

M3: Inculcate team work and lifelong learning among students with a sense of societal and ethical responsibilities.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Program Specific Outcomes (PSO's)

PSO1: Apply mathematical and scientific skills in numerous areas of Computer Science and Engineering to design and develop software-based systems.

PSO2: Acquaint module knowledge on emerging trends of the modern era in Computer Science and Engineering

PSO3: Promote novel applications that meet the needs of entrepreneur, environmental and social issues.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Program Educational Objectives (PEO's)

The graduates of the programme are able to:

PEO1: Apply the knowledge of Mathematics, Science and Engineering fundamentals to identify and solve Computer Science and Engineering problems.

PEO2: Use various software tools and technologies to solve problems related to academia, industry and society.

PEO3: Work with ethical and moral values in the multi-disciplinary teams and can communicate effectively among team members with continuous learning.

PEO4: Pursue higher studies and develop their career in software industry.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Program Outcomes

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Project Course Outcomes (CO'S):

CO425.1: Analyze the System of Examinations and identify the problem.

CO425.2: Identify and classify the requirements.

CO425.3: Review the Related Literature

CO425.4: Design and Modularize the project

CO425.5: Construct, Integrate, Test and Implement the Project.

CO425.6: Prepare the project Documentation and present the Report using appropriate method.

Course Outcomes – Program Outcomes mapping

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C425.1		✓											✓		
C425.2	✓		✓		✓								✓		
C425.3				✓		✓	✓	✓					✓		
C425.4			✓			✓	✓	✓					✓	✓	
C425.5					✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
C425.6									✓	✓	✓		✓	✓	

Course Outcomes – Program Outcome correlation

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
C425.1	2	3											2		
C425.2			2		3								2		
C425.3				2		2	3	3					2		
C425.4			2			1	1	2					3	2	
C425.5					3	3	3	2	3	2	2	1	3	2	1
C425.6									3	2	1		2	3	

Note: The values in the above table represent the level of correlation between CO's and PO's:

1. **Low level**

2. **Medium level**

3. **High level**

Project mapping with various courses of Curriculum with Attained PO's:

Name of the course from which principles are appliedin this project	Description of the device	Attained PO
C3.2.4, C3.2.5	Gathering the requirements and defining the problem, plan to develop a customer segmentation	PO1, PO3
CC4.2.5	Each and every requirement is critically analyzed, the process model is identified and divided into four modules	PO2, PO3
CC4.2.5	Logical design is done by using the unified modelling language which involves individual team work	PO3, PO5, PO9
CC4.2.5	Each and every module is tested, integrated, and evaluated in our project	PO1, PO5
CC4.2.5	Documentation is done by all our four members in the form of a group	PO10
CC4.2.5	Each and every phase of the work ingroup is presented periodically	PO10, PO11
CC4.2.5	Implementation is done and theproject will be handled by the mall customers it is developed to the prediction of customers in the market.	PO4, PO7
CC4.2.8 CC4.2.	The physical design includes the website to check the future data of the customer.	PO5, PO6

INDEX

LIST OF FIGURES

1. INTRODUCTION

1.1 Introduction

Rainfall prediction remains a serious concern and has attracted the attention of governments, industries, risk management entities, as well as the scientific community. Rainfall is a climatic factor that affects many human activities like agricultural production, construction, power generation, forestry and tourism, among others. To this extent, rainfall prediction is essential since this variable is the one with the highest correlation with adverse natural events such as landslides, flooding, mass movements and avalanches. These incidents have affected society for years. Therefore, having an appropriate approach for rainfall prediction makes it possible to take preventive and mitigation measures for these natural phenomena .

To solve this uncertainty, we used various machine learning techniques and models to make accurate and timely predictions. These paper aims to provide end to end machine learning life cycle right from Data preprocessing to implementing models to evaluating them. Data Preprocessing steps include imputing missing values, feature transformation, encoding categorical features, feature scaling and feature selection. We implemented models such as Logistic Regression, Decision Tree, K Nearest Neighbour, Rule-based and Ensembles. For evaluation purpose, we used Accuracy, Precision, Recall, F-Score and Area Under Curve as evaluation metrics. For our experiments, we train our classifiers using Australian weather data gathered from various weather stations in Australia.

1.2 Existing System

One of the biggest issues with customer segmentation is data quality. Inaccurate data in source systems will usually result in poor grouping. For example, customers who are individuals, attributes like age, gender, and marital status are frequently used. If these attributes are not maintained properly, the segments will be inaccurate and as a result, the information will likely be less useful

Disadvantages:

1.Data Quality issues arises.

2.Data management of the customers.

3.Computation process is more.

1.3 Proposed System

By using this system, we need to analyze the data and after analysis of data by classifying customers with features annual income and total spending ,we got clusters of customers and with formed clusters marketing team form strategies for customers specific recommendation to make value of it and making quick decisions.

Advantages:

1. Generates accurate and efficient results.
2. Computation time is greatly reduced.
3. Data replica.
4. Data cleaning and data repository.

1.4 System Requirements

1.4.1 Hardware Requirements

- System Type: intel®core™i5-7500UCPU@1.03gh
- Cache memory: 4 Megabyte(MB)
- RAM: 4 Gigabyte (GB)
- Hard Disk: 1 Terabyte (TB)

1.4.2 Software Requirements

- Operating System: Windows 10 Home, 64-bit Operating System
- Coding Language: Python
- Python distribution: Anaconda(pycharm), Google Colab

2.LITERATURE SURVEY

2.1 Machine Learning

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions inthe future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actionsaccordingly.

Over the years, the commercial world has become more competitive, as organizations such as these have to meet the needs and desires of their customers, attract new customers and thus improve their businesses. The task of identifying and meeting the needs and requirements of every customer in the business is very difficult. This is because customers can vary according to their needs, wants, demographics, size, taste and taste, features etc. As it is, it is a bad practice to treat all customers equally in business. This challenge has adopted the concept of customer segmentation or market segmentation, where consumers are divided into subgroups or segments, where members of each subcategory exhibit similar market behaviours or characteristics. Accordingly, customer segmentation is the process of dividing the market into indigenous groups.

Data collection is the process of collecting and measuring information against targeted changes in an established system, which enables one to answer relevant questions and evaluate the results. Data collection is part of research in all fields of study including physical and social sciences, humanities and business. The purpose of all data collection is to obtain quality evidence that leads the analysis to construct concrete and misleading answers to the questions presented. We collected data from the UCI machine learning repository.

CatBoost is an algorithm for gradient boosting on decision trees. It is developed by Yandex researchers and engineers, and is used for search, recommendation systems, personal assistant, self-driving cars, weather prediction and many other tasks at Yandex and in other companies, including CERN, Cloudflare, Careem taxi. It is in open-source and can be used by anyone. Gradient boosting algorithms often have a tendency to overfit. Since ensembles work iteratively building upon the base learners over the same dataset, it affects the generalization capability of the model.By using this algorithm we achieved an accuracy of 86%.

2.2 Some machine learning methods

Machine learning algorithms are often categorized as supervised and unsupervised.

- **Supervised machine learning algorithms** can apply what has been learned in thepast to new data using labelled examples to predict future events. Starting from the analysisof a known training dataset, the learning algorithm produces an inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.
- **unsupervised machine learning algorithms** are used when the information used to train is neither classified nor labelled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabelled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasetsto describe hidden structures from unlabelled data.
- **Reinforcement machine learning algorithms** is a learning method that interacts with its environment by producing actions and discovers errors or rewards. search and delayed reward are the most relevant characteristics of reinforcement learning. This method allows machines and software agents to automatically determine the ideal behaviors within a specific context in order to maximize its performance. Simple reward feedback is required for the agent to learn which action is best. This is known as the reinforcement signal.

2.3 Applications of machine learning

1. Real-time chatbot agents
2. Decision support
3. Customer recommendation engines
4. Dynamic pricing tactics
5. Rainfall Prediction
6. Fraud Detection
7. Text parsing
8. Image Classification
9. Improving Cyber security
10. Detecting Spam

3. SYSTEM ANALYSIS

3.1 System Architecture

Initially we will see the dataset and then we will perform exploratory data analysis which deals with the missing data, duplicates values and null values. And then we will deploy our algorithm Catboost Classifier which is supervised learning in machine learning.

CatBoost is based on gradient boosted decision trees. During training, a set of decision trees is built consecutively. Each successive tree is built with reduced loss compared to the previous trees.

3.2. Importance of machine learning in Rainfall Prediction

Machine learning has become increasingly important in the field of rainfall prediction due to its ability to analyze large amounts of data and identify patterns that may not be apparent to humans. Here are some reasons why machine learning is important in rainfall prediction:

Improved accuracy: Machine learning algorithms can analyze multiple data sources simultaneously, including historical rainfall patterns, atmospheric conditions, and weather satellite images. This enables them to identify patterns and relationships that may not be apparent to humans, resulting in more accurate predictions.

Faster predictions: Machine learning algorithms can process large amounts of data quickly, allowing them to make predictions in real-time. This is particularly important in the case of extreme weather events, where timely predictions can help to prevent damage and save lives.

Adaptability: Machine learning algorithms can adapt to changing conditions and update their predictions accordingly. For example, they can incorporate new data sources or adjust their predictions based on feedback from sensors or other devices.

Scalability: Machine learning algorithms can be trained on large datasets and can handle multiple variables simultaneously. This makes them suitable for predicting rainfall patterns over large geographic areas or for long periods of time.

3.3 Implementation of machine learning using Python

Python is a popular programming language. It was created in 1991 by Guido van Rossum.

The most recent major version of Python is Python 3. However, Python 2, although not being updated with anything other than security updates, is still quite popular.

It is possible to write Python in an Integrated Development Environment, such as Thonny, PyCharm, NetBeans or Eclipse, Anaconda which are particularly useful when managing larger collections of Python files.

Python was designed for its readability. Python uses new lines to complete a command, as opposed to other programming languages which often use semicolons or parentheses.

Python relies on indentation, using whitespace, to define scope; such as the scope of loops, functions and classes. Other programming languages often use curly-brackets for this purpose.

In the older days, people used to perform Machine Learning tasks manually by coding all the algorithms and mathematical and statistical formula. This made the process time consuming, tedious and inefficient. But in the modern days, it is become very much easy and efficient compared to the olden days by various python libraries, frameworks, and modules. Today, Python is one of the most popular programming languages for this task and it has replaced many languages in the industry, one of the reasons is its vast collection of libraries. Python libraries that used in Machine Learning are:

1. Scikit learn

2. NumPy

3. Pandas

4.Matplotlib

5.Seaborn

1. Scikit-learn

It is a free Python machine learning software, sometimes known as sklearn. It is meant to interact with the Python numerical and scientific libraries NumPy and SciPy, and features support vector machines, random forests, gradient boosting, k-means, and DBSCAN, among other classification, regression, and clustering algorithms.

2. NumPy

NumPy is a general-purpose array-processing package. It provides a high performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python. It contains various features including these important ones: A powerful N-dimensional array object

- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multidimensional container of generic data. Arbitrary data-types can be defined using NumPy which allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

3. Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyse. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

4. Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy

things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery. For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object-oriented interface or via a set of functions familiar to MATLAB users.

5. Seaborn

Seaborn is a library for making statistical graphics in python. It builds on top of matplotlib and integrates closely with pandas' data structures. Seaborn helps you explore and understand your data .Its plotting functions operate on data frames and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots. Its dataset-oriented, declarative API lets you focus on what the different elements of your plots mean, rather than on the details of how to draw them.

3.4 Scope of the project

The scope of a project on rainfall prediction using machine learning may include the following:

Problem Statement: A clear description of the problem you are trying to solve, including the objective of the project and the expected outcomes.

Data Collection: A description of the data that will be used for the project. This should include the sources of the data, the size of the dataset, and any preprocessing that will be necessary to prepare the data for analysis.

Exploratory Data Analysis: An analysis of the data to understand the relationship between the variables and the distribution of the data. This should include visualizations and statistical summaries.

Feature Engineering: The selection and engineering of relevant features from the raw data that can be used to build the machine learning models. This includes both domain knowledge-based and data-driven feature selection.

Model Building: The development and implementation of machine learning models to predict rainfall. This could include regression models, neural networks, decision trees, or other appropriate algorithms.

Model Evaluation: The evaluation of the performance of the machine learning models. This should include metrics such as accuracy, precision, recall, and F1-score, and the use of appropriate validation techniques such as cross-validation.

Model Deployment: The deployment of the machine learning models in a production environment, which may involve building a web interface or integrating the models into an existing system.

Conclusion and Future Work: A summary of the results obtained from the project and recommendations for future work, including any limitations or areas for improvement.

Documentation: A well-documented codebase that explains the different parts of the code and how to use them. This should include installation instructions, configuration information, and examples of how to use the code to make predictions.

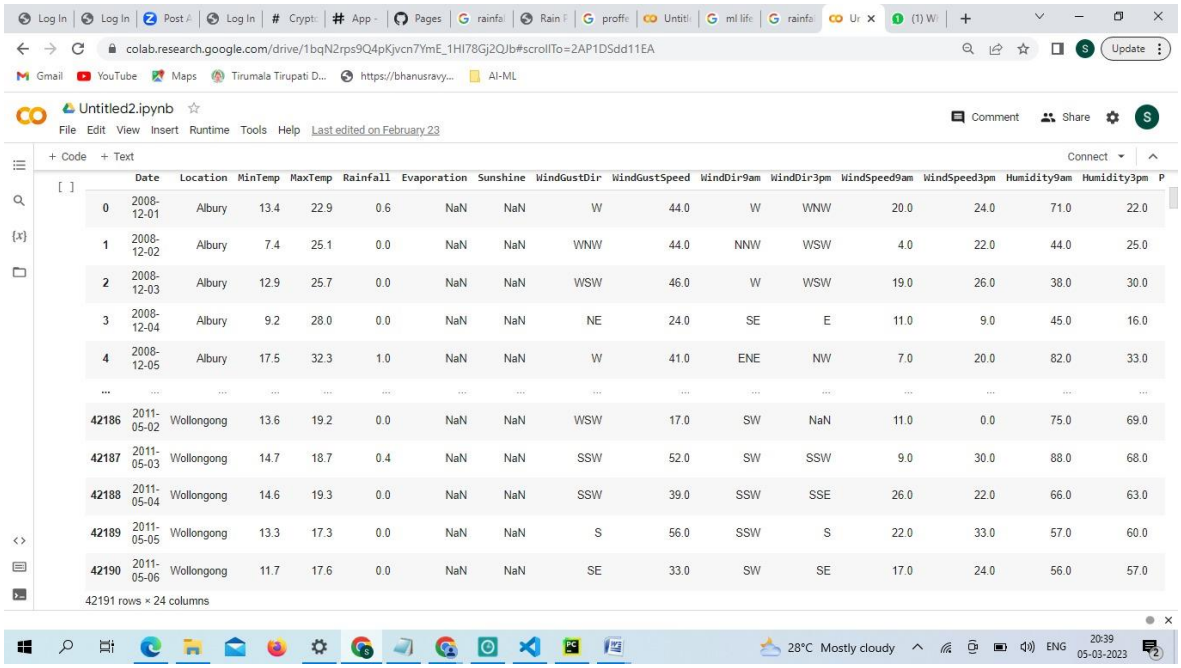
3.5 Analysis

An analysis of the data to understand the relationships between the variables, the distribution of the data, and the presence of missing values. EDA can help to identify trends, patterns, and anomalies in the data that can guide the feature engineering process. The selection and engineering of relevant features from the raw data that can be used to build the machine learning models. This includes both domain knowledge-based and data-driven feature selection. The feature engineering process can involve techniques such as scaling, normalization, and transformation of the data.

The data includes the following features:

'Date', 'Location', 'MinTemp', 'MaxTemp', 'Rainfall', 'Evaporation', 'Sunshine', 'WindGustDir', 'WindGustSpeed', 'WindDir9am', 'WindDir3pm', 'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm', 'Pressure9am', 'Pressure3pm', 'Cloud9am', 'Cloud3pm', 'Temp9am', 'Temp3pm', 'RainToday', 'RISK_MM','RainTomorrow'

Data Set:



Untitled2.ipynb

File Edit View Insert Runtime Tools Help Last edited on February 23

+ Code + Text

	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm
0	2008-12-01	Albury	13.4	22.9	0.6	NaN	NaN	W	44.0	W	WNW	20.0	24.0	71.0	22.0
1	2008-12-02	Albury	7.4	25.1	0.0	NaN	NaN	WNW	44.0	NNW	WSW	4.0	22.0	44.0	25.0
2	2008-12-03	Albury	12.9	25.7	0.0	NaN	NaN	WSW	46.0	W	WSW	19.0	26.0	38.0	30.0
3	2008-12-04	Albury	9.2	28.0	0.0	NaN	NaN	NE	24.0	SE	E	11.0	9.0	45.0	16.0
4	2008-12-05	Albury	17.5	32.3	1.0	NaN	NaN	W	41.0	ENE	NW	7.0	20.0	82.0	33.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
42186	2011-05-02	Wollongong	13.6	19.2	0.0	NaN	NaN	WSW	17.0	SW	NaN	11.0	0.0	75.0	69.0
42187	2011-05-03	Wollongong	14.7	18.7	0.4	NaN	NaN	SSW	52.0	SW	SSW	9.0	30.0	88.0	68.0
42188	2011-05-04	Wollongong	14.6	19.3	0.0	NaN	NaN	SSW	39.0	SSW	SSE	26.0	22.0	66.0	63.0
42189	2011-05-05	Wollongong	13.3	17.3	0.0	NaN	NaN	S	56.0	SSW	S	22.0	33.0	57.0	60.0
42190	2011-05-06	Wollongong	11.7	17.6	0.0	NaN	NaN	SE	33.0	SW	SE	17.0	24.0	56.0	57.0

42191 rows × 24 columns

Fig:3.1 DataSet

3.6 Data Pre-processing

Before feeding data to an algorithm, we have to apply transformations to our data which is referred as pre-processing. By performing pre-processing, the raw data which is not feasible for analysis is converted into clean data. In-order to achieve better results using a model in Machine Learning, data format has to be in a proper manner. The data should be in a particular format for different algorithms. For example, if we consider Random Forest algorithm it does not support null values. So that those null values have to be managed using raw data.

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

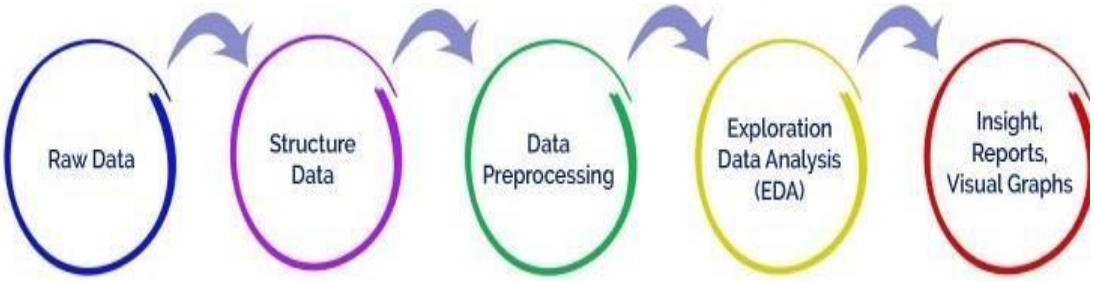


Fig: 3.2 Data Pre-processing

Need of Data Pre-processing: For achieving better results from the applied model in Machine Learning projects the format of the data has to be in a proper manner. Some specified Machine Learning model needs information in a specified format. For example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null values have to be managed from the original raw data set. Another aspect is that data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set, and best out of them is chosen.

3.6.1 Missing values

Filling missing values is one of the pre-processing techniques. The missing values in the dataset is represented as '?' but it a non-standard missing value and it has to be converted into a standard missing value Nan. So that pandas can detect the missing values.

In our dataset, there are no missing values.

3.6.2 Correlation

We can find dependency between two attributes p and q using Correlation coefficient method using the formula.

$rp= \sum(pi-p)(qi-q)/n\sigma p\sigma q$

$q=\sum(pi\ qi)-np\ q/\ n\sigma p\sigma q$

n is the total number of patterns, pi and qi are respective values of p and q attributes in patterns i, p and q are respective mean values of p and q attributes, σp , σq are respective standard deviations values of p and q attributes. Generally, $-1 \leq rp,q \leq +1$. If $rp,q < 0$, then p and q are negatively correlated. If $rp,q = 0$, then p and q are independent attributes and there is no correlation between them. If $rp,q > 0$, then p and q are positively correlated. We can drop the attributes that are having correlation coefficient value as 0 as it indicates that the variables are independent with respect to the prediction attribute. Fig:3.8.2 is the correlation

heat map. After applying correlation, the attributes are PR interval , QRS duration , QT interval , QTc interval, P wave , T wave , QRS wave and problem .

Correlation heatmaps are a type of plot that visualize the strength of relationships between numerical variables. Correlation plots are used to understand which variables are related to each other and the strength of this relationship. A correlation plot typically contains a number of numerical variables, with each variable represented by a column. Correlation heatmaps can be used to find potential relationships between variables and to understand the strength of these relationships. The color-coding of the cells makes it easy to identify relationships between variables at a glance. Correlation heatmaps can be used to find both linear and nonlinear relationships between variables. Here is the Python code which can be used to draw a **correlation heatmap** for the **housing data set** representing the correlation between different variables including predictor and response variables.

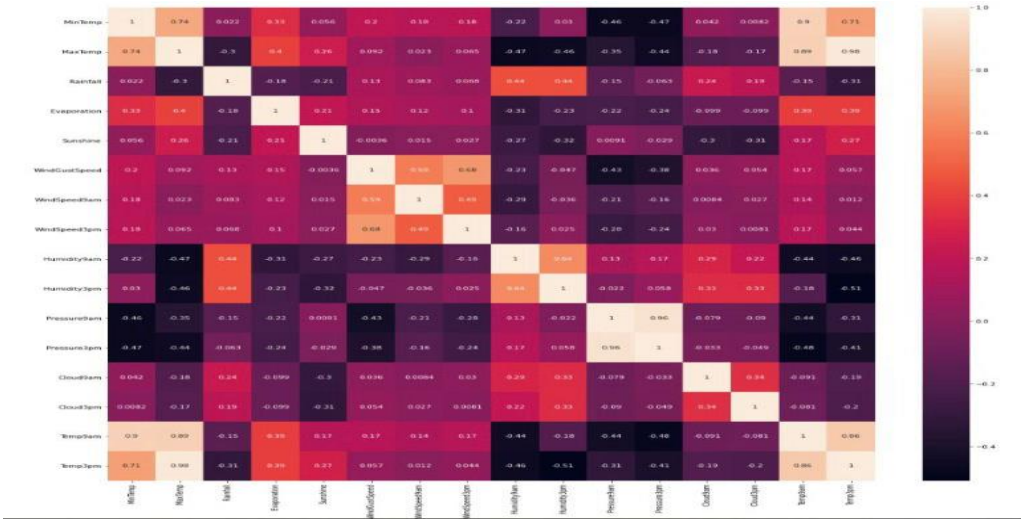


Fig: 3.3 Correlation

3.6.3 Principal Component Analysis

Principal Component Analysis(PCA) is an unsupervised learning algorithm that is used for the dimensionality reduction in machine learning. It is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new transformed features are called the Principal Components. It is one of the popular tools that is used for exploratory data analysis and predictive modeling. It is a technique to draw strong patterns from the given dataset by reducing the variances.

3.6.4 Visualize of Data

Plotting

Data Visualization is the process of presenting data in the form of graphs or charts. It helps to understand large and complex amounts of data very easily. It allows the decision-makers to make decisions very efficiently and also allows them in identifying new trends and patterns very easily. It is also used in high-level data analysis for Machine Learning and Exploratory Data Analysis (EDA). Data visualization can be done with various tools like Tableau, Power BI, Python.

Types of plots

Box Plot

Box plot gives statistical information about the distribution of numeric data divided into different groups. It is useful for detecting outliers within each group. A Box Plot is also known as Whisker plot is created to display the summary of the set of data values having properties like minimum, first quartile, median, third quartile and maximum. In the box plot, a box is created from the first quartile to the third quartile, a vertical line is also there which goes through the box at the median. X-axis denotes the data to be plotted while the Y-axis shows the frequency distribution.

Scatter Plot

The scatter plots are preferred while comparing the data variables to determine the relationship between dependant and independent variables. The data is displayed as a collection of points, each having the value of one variable which determines the position on the horizontal axis and the value of other variable determines the position on the vertical axis. Scatter plots are used to plot data points on horizontal and vertical axis in the attempt to show how much one variable is affected by another. Each row in the data table is represented by a marker the position depends on its values in the columns set on the X and Y axes.

Visualizing the dataset according to various features using different types of plots. First, visualize the gender using the bar plot. Figure shows that the female count is higher than the male count.

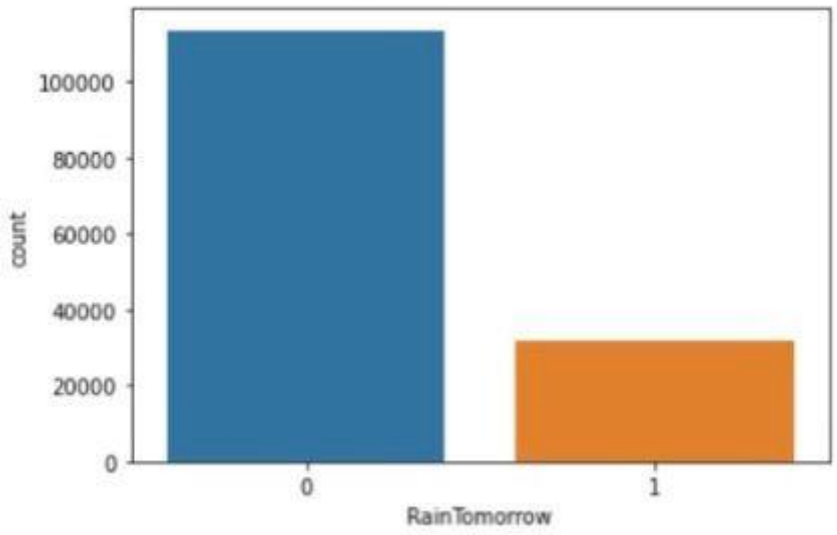


Fig: 3.4 RainTomorrow

BOX PLOTS FOR ATTRIBUTES:

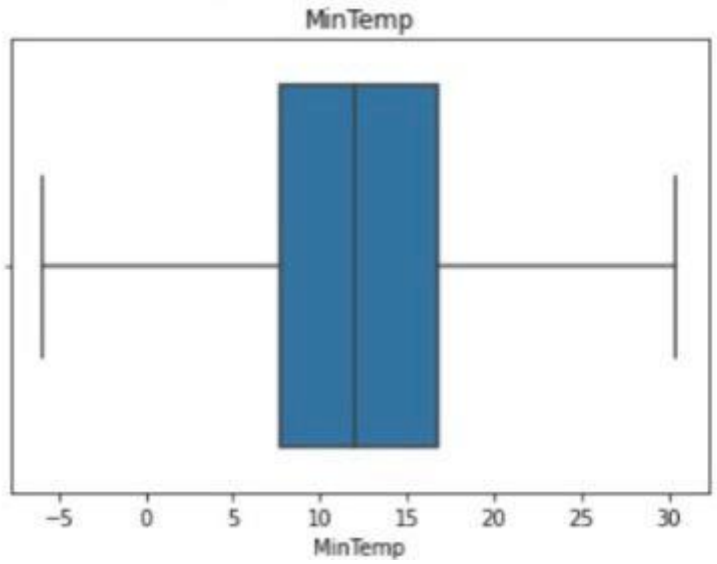


Fig:3.5 Box plot of MinTemp

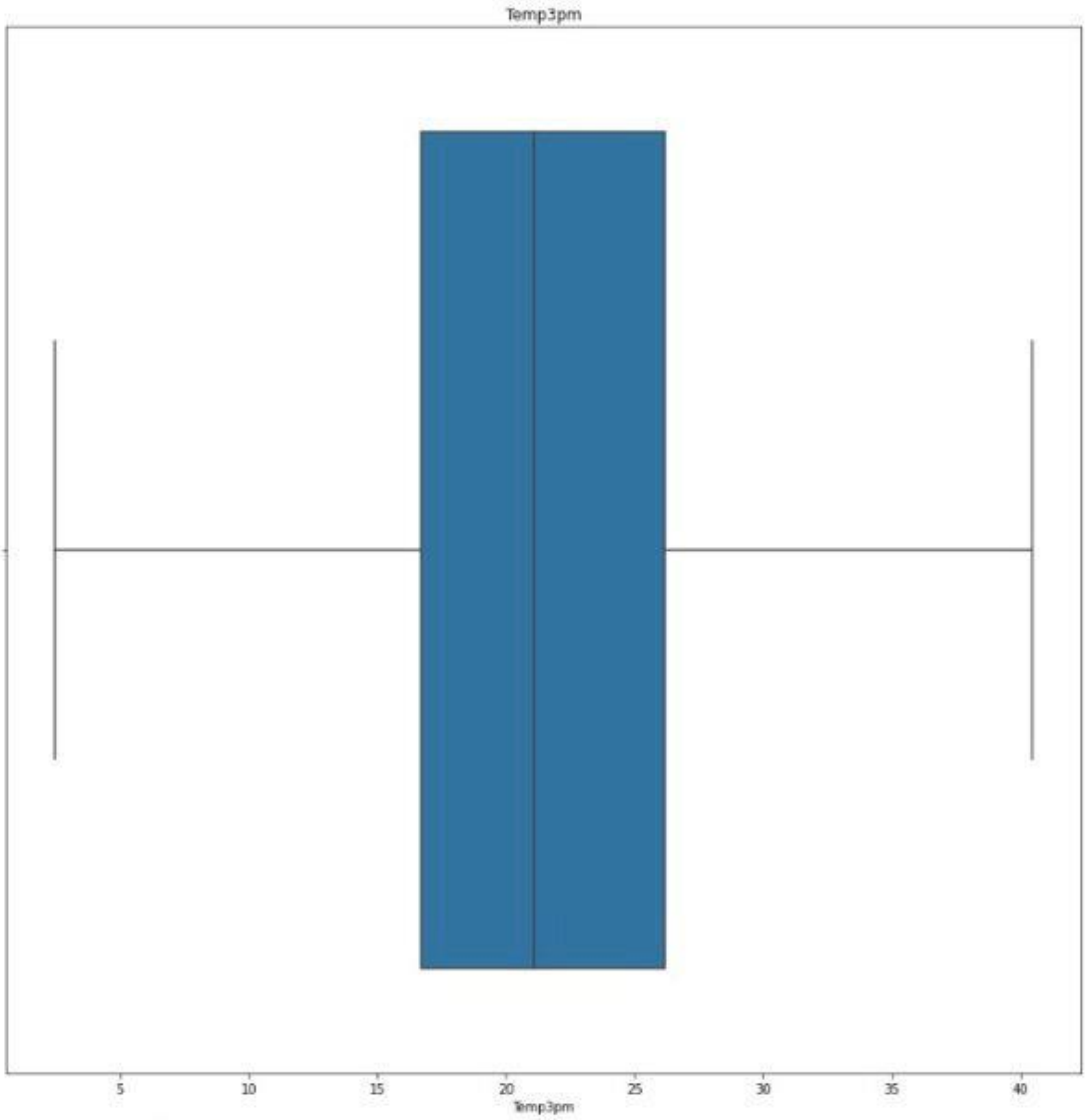


Fig:3.6 Box plot of Temp3pm

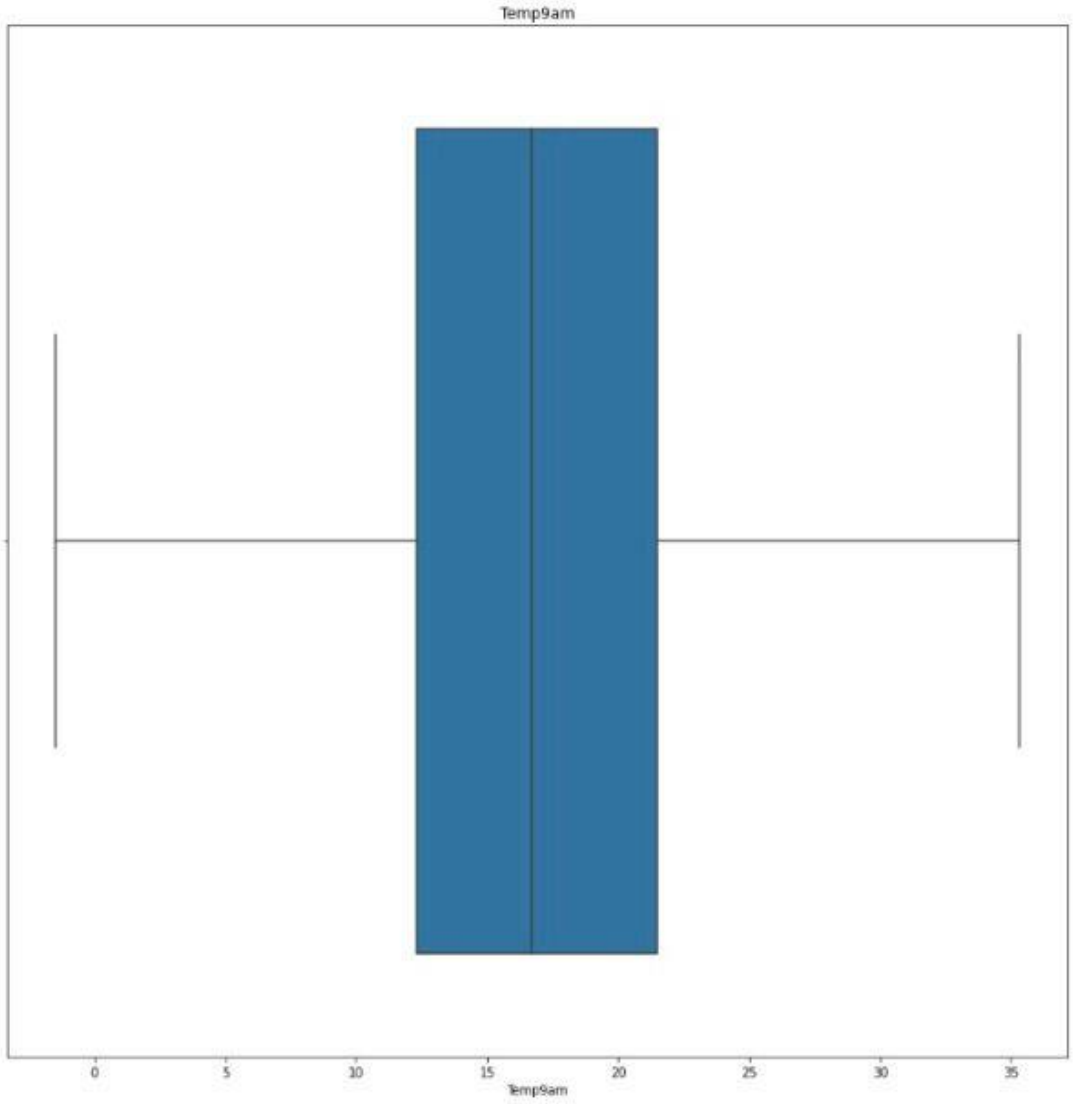


Fig:3.7 Box plot of Temp9am

HISTOGRAM PLOTS FOR ATTRIBUTES:

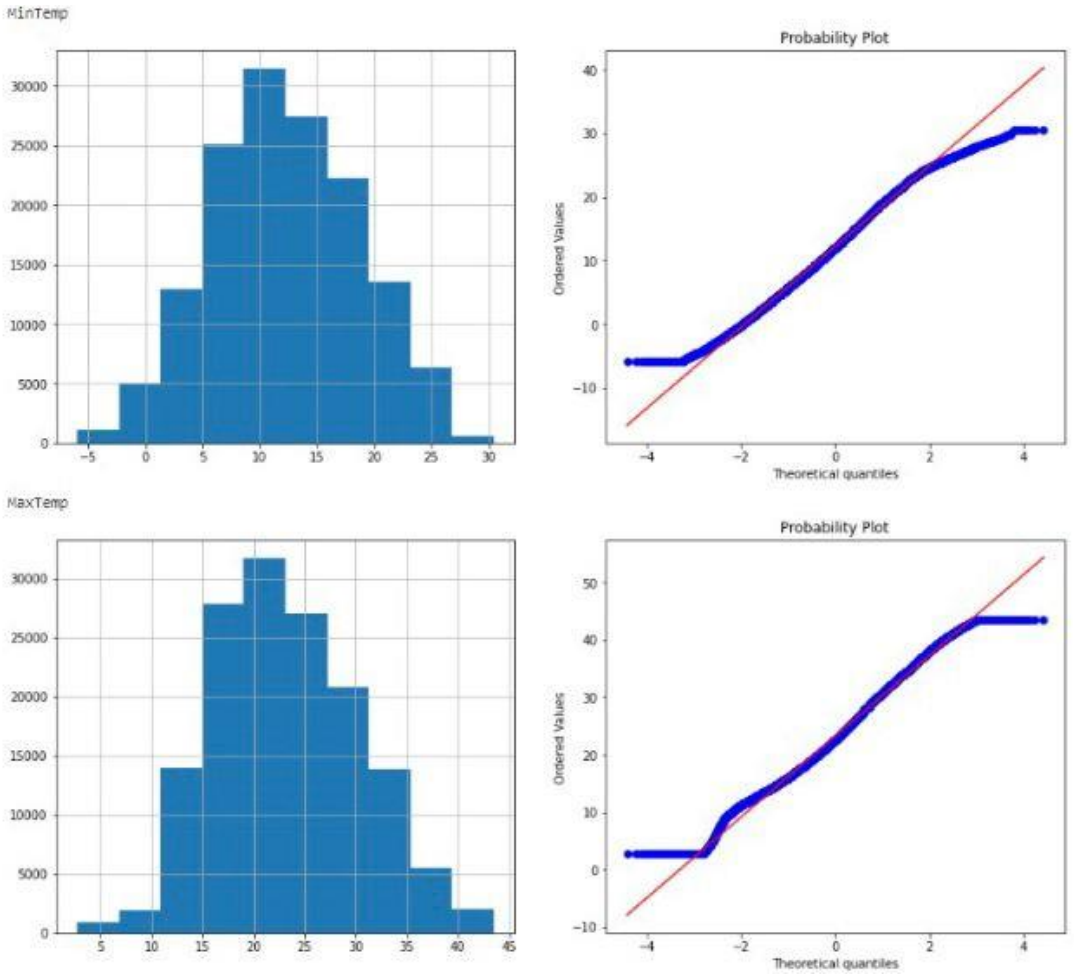


Fig: 3.8 Histogram of MinTemp and MaxTemp

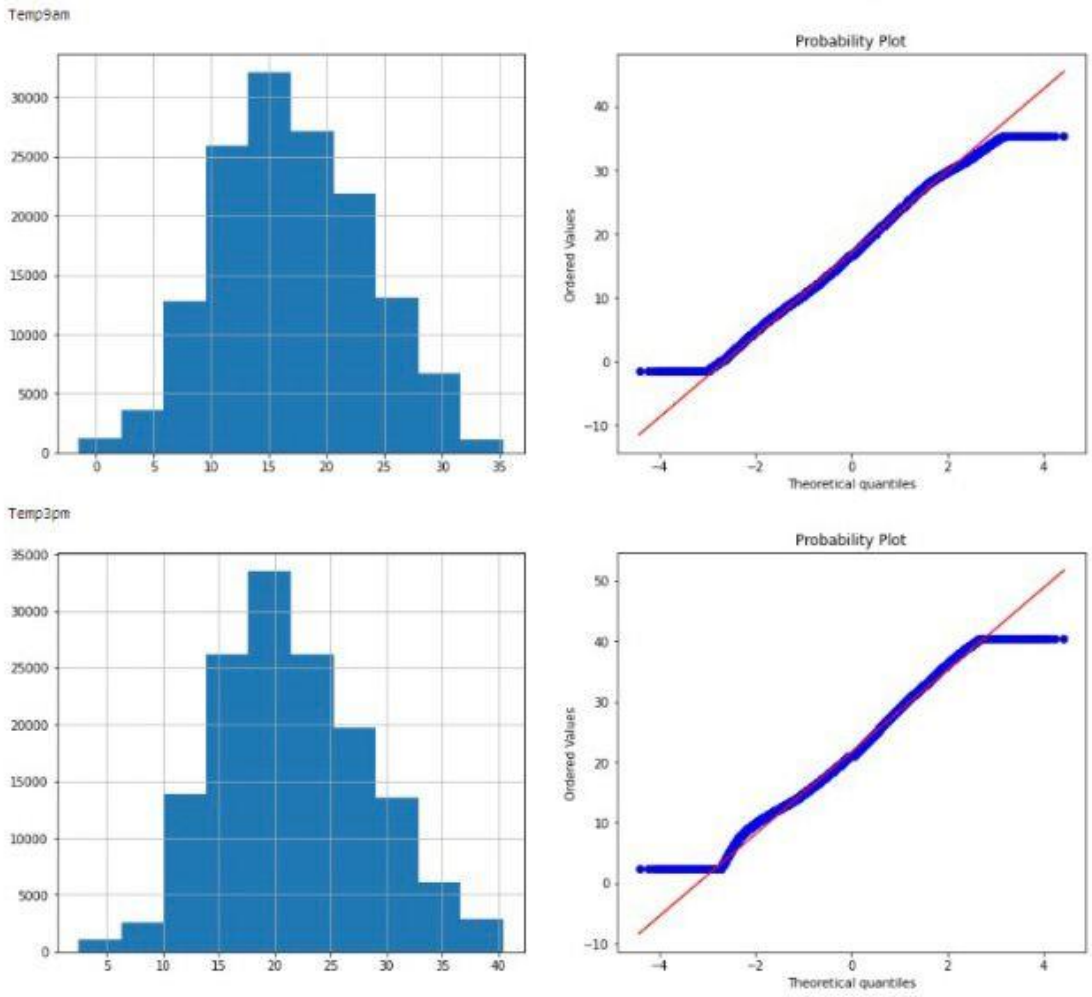


Fig: 3.9 Histogram of Temp9am and Temp3pm

3.7 Classification

3.7.1 Machine Learning Algorithms for Classification

Research on data mining has led to the formulation of several data mining algorithms. These algorithms can be directly used on a dataset for creating some models or to draw vital conclusions and inferences from that dataset. Some popular data mining algorithms are Random forest, Decision tree, Gaussian Naïve Bayes, Support vector machine etc.

1. **Decision Tree:**

   Decision Tree Analysis is a general, predictive modelling tool that has applications spanning a number of different areas. In general, decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. The decision rules are generally in form of if-then-else statements. The deeper the tree, the more complex the rules and fitter the model.

2. **K-Nearest Neighbour:**

   This Algorithm is a non-parametric and lazy learning algorithm.Non-parametric means there is no assumption for underlying data distribution.In other words, the model structure is determined from the dataset. Lazy algorithm means it does not need any training data points for model generation.All training data used in the testing phase. KNN performs better with a lower number of features than a large number of features. We can say that when the number of features increases than it requires more data. Increase in dimension also leads to the problem of overfitting. However, we have performed feature selection which helps to reduce dimension and hence KNN looks a good candidate for our problem.

3. **Logistic Regreesion:**

   This Algorithm is a classification algorithm used to predict a binary outcome (1 / 0, Yes / No, True / False) given a set of independent variables. To represent binary / categorical outcome, we use dummy variables. We can also think of logistic regression as a special case of linear regression when the outcome variable is categorical, where we are using log of odds as dependent variable. In simple words,

it predicts the probability of occurrence of an event by fitting data to a logit function. Hence, this makes Logistic Regression a better fit as ours is a binary classification problem.

4. **Random Forest:**

   This Algorithm is a supervised ensemble learning algorithm. Ensemble means that it takes a bunch of weak learners and have them work together to form one strong predictor. Here, we have a collection of decision trees, known as Forest.To classify a new object based on attributes, each tree gives a classification and we say the tree votes for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

5. **AdaBoost:**

   This Algorithm fits a sequence of weak learners on different weighted training data. It starts by predicting original data set and gives equal weight to each observation.If prediction is incorrect using the first learner, then it gives higher weight to observations which have been predicted incorrectly. Being an iterative process,it continues to add learner(s) until a limit is reached in the number of models or accuracy

6. **Gradient Boosting:**

   Here, many models are trained sequentially. Each new model gradually minimizes the loss function ($y = ax + b + e$, where e is the error term) of the whole system using Gradient Descent method. The learning method consecutively fits new models to give a more accurate estimate of the response variable. The main idea behind this algorithm is to construct new base learners which can be optimally correlated with negative gradient of the loss function, relevant to the whole ensemble.

7. **Catboost:**

   CatBoost is a popular gradient boosting algorithm used in machine learning for classification, regression, and ranking tasks. It was developed by Yandex researchers and was released as an open-source library in 2017. CatBoost stands for "Category Boosting" and is particularly useful for handling categorical features in datasets. It uses an ordered boosting technique to handle categorical variables and can handle a wide range of data types, such as numerical, categorical, and text data.

3.7.2 Choosing Best Algorithm:

The CatBoost algorithm can be a good choice for a rainfall prediction project in machine learning due to its ability to handle complex data types, such as categorical variables, and its strong performance in a variety of applications.

Rainfall prediction involves analyzing a wide range of data, including atmospheric data, topographical data, and historical rainfall data. Many of these variables are categorical in nature, such as the type of cloud cover or the direction of the wind. CatBoost's "Ordered Boosting" technique is specifically designed to handle categorical variables, which can improve the accuracy of the predictions.

In addition, CatBoost is known for its strong performance in a variety of applications. It has been used successfully in a wide range of projects, including web search ranking, image classification, and recommender systems. Its ability to handle missing data effectively and prevent overfitting can also be advantageous in a complex project like rainfall prediction.

Overall, the selection of the CatBoost algorithm for a rainfall prediction project would depend on the specific requirements of the project, the available data, and the expertise of the team working on the project. However, given its strengths in handling categorical variables and complex data, CatBoost could be a promising choice for such a project.

Algorithm

1. Data Preparation: This step involves cleaning and processing the input data (which could be in the form of a table with columns representing features and labels). This step converts any categorical variables into a numerical form that the algorithm can work with.
2. Initialize Model: At this step, we define the parameters of the CatBoost model, such as the depth of the trees and how to handle categorical features. We create a "skeleton" model that will be refined during the training phase.
3. Training: Here, we use the training dataset to improve the model by iteratively building decision trees that correct the errors made by previous trees. We use gradient and hessian matrices to calculate how much the model needs to be updated at each

step. We also monitor metrics like accuracy to evaluate how well the model is performing.

4. Prediction: After training, we use the model to make predictions on new data. We use decision tree traversal algorithms to assign the new data points to leaf nodes, and then use statistical models to generate the predicted labels.

Overall, CatBoost classification involves preparing the data, defining the model, training the model, and then using the trained model to make predictions. The algorithm uses various techniques to handle categorical features and missing data, and iteratively improves the model until it reaches a satisfactory level of accuracy.

3.8 Implementation code

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn import preprocessing

import scipy.stats as stats

from sklearn.model_selection import train_test_split

from collections import Counter

from imblearn.over_sampling import SMOTE

from sklearn.metrics import accuracy_score,confusion_matrix,classification_report

from sklearn import metrics

from catboost import CatBoostClassifier

from sklearn.ensemble import AdaBoostClassifier;

from sklearn.ensemble import GradientBoostingClassifier

from sklearn.linear_model import LogisticRegression

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier

from sklearn.neighbors import KNeighborsClassifier


# Handle Missing Values

df.isnull().sum()*100/len(df)
```

```
#Correlation
corrmat = df.corr(method = "spearman")
plt.figure(figsize=(20,20))
#plot heat map
g=sns.heatmap(corrmat,annot=True)
#Graphs
for feature in continuous_feature:
    data=df.copy()
    sns.distplot(df[feature])
    plt.xlabel(feature)
    plt.ylabel("Count")
    plt.title(feature)
    plt.figure(figsize=(15,15))
    plt.show()
#A for loop is used to plot a boxplot for all the continuous features to see the outliers
for feature in continuous_feature:
    data=df.copy()
    sns.boxplot(data[feature])
    plt.title(feature)
    plt.figure(figsize=(15,15))
X_train, X_test, y_train, y_test = train_test_split(X,Y, test_size =0.2, stratify = Y,
random_state = 0)
sm=SMOTE(random_state=0)
X_train_res, y_train_res = sm.fit_resample(X_train, y_train)
print("The number of classes before fit {}".format(Counter(y_train)))
```

```
print("The number of classes after fit {}".format(Counter(y_train_res)))

#Catboost Algorithm

cat = CatBoostClassifier(iterations=2000, eval_metric = "AUC")

cat.fit(X_train_res, y_train_res)

y_pred = cat.predict(X_test)

print(confusion_matrix(y_test,y_pred))

print(accuracy_score(y_test,y_pred))

print(classification_report(y_test,y_pred))

#Logistic Regression

logreg = LogisticRegression()

logreg.fit(X_train_res, y_train_res)

y_pred2 = logreg.predict(X_test)

print(confusion_matrix(y_test,y_pred2))

print(accuracy_score(y_test,y_pred2))

print(classification_report(y_test,y_pred2))

#Random Forest

rf=RandomForestClassifier()

rf.fit(X_train_res,y_train_res)

y_pred1 = rf.predict(X_test)

print(confusion_matrix(y_test,y_pred1))

print(accuracy_score(y_test,y_pred1))

print(classification_report(y_test,y_pred1))
```

```
#KNN
knn=KNeighborsClassifier()
rf.fit(X_train_res,y_train_res)
y_pred3 = rf.predict(X_test)
print(confusion_matrix(y_test,y_pred3))
print(accuracy_score(y_test,y_pred3))
print(classification_report(y_test,y_pred3))
#Gradient Boosting
gb=GradientBoostingClassifier()
gb.fit(X_train_res,y_train_res)
y_pred4 = rf.predict(X_test)
print(confusion_matrix(y_test,y_pred4))
print(accuracy_score(y_test,y_pred4))
print(classification_report(y_test,y_pred4))
#Adaboost
gb=AdaBoostClassifier()
gb.fit(X_train_res,y_train_res)
y_pred5 = rf.predict(X_test)
print(confusion_matrix(y_test,y_pred5))
print(accuracy_score(y_test,y_pred5))
print(classification_report(y_test,y_pred5))
#DecisionTree
gb=DecisionTreeClassifier()
gb.fit(X_train_res,y_train_res)
```

```
y_pred6 = rf.predict(X_test)

print(confusion_matrix(y_test,y_pred6))

print(accuracy_score(y_test,y_pred6))

print(classification_report(y_test,y_pred6))
```

3.9 Confusion matrix

A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. A true positive (tp) is a result where the model predicts the positive class correctly. Similarly, a true negative (tn) is an outcome where the model correctly predicts the negative class. A false positive (fp) is an outcome where the model incorrectly predicts the positive class. And a false negative (fn) is an outcome where the model incorrectly predicts the negative class.

Sensitivity or Recall or hit rate or true positive rate (TPR)

It is the proportion of individuals who actually have belong to the cluster were identified as having in that cluster.

TPR = tp / (tp + fn)

Specificity, selectivity or true negative rate (TNR)

It is the proportion of individuals who actually do not belong to the cluster were identified as not belongs to that customer.

TNR = tn / (tn + fp) =1-FPR

Precision or positive predictive value (PPV)

If the test result is positive what is the probability that the customer has belong to that customer.

PPV = tp / (tp + fp)

Negative predictive value (NPV)

If the test result is negative what is the probability that the customer not belong to that cluster.

NPV = tn / (tn + fn)

Miss rate or false negative rate (FNR)

It is the proportion of the individuals with a known positive condition for which the test result is negative.

FNR = fn / (fp + tn)

Fall-out or false positive rate (FPR)

It is the proportion of all the customers having misplaced in their respective clusters.

FPR = fp/ (fp + tn)

False discovery rate (FDR)

It is the proportion of all the customers having misplaced in clusters

FDR = fp / fp + tp

Accuracy

The accuracy reflects the total proportion of individuals that are correctly classified.

ACC = (tp + tn) / (tp + tn + fp + fn)

F1 score

It is the harmonic mean of precision and sensitivity

F1 = 2tp / (2tp+ fp + fn)

4.OUTPUT SCREENS

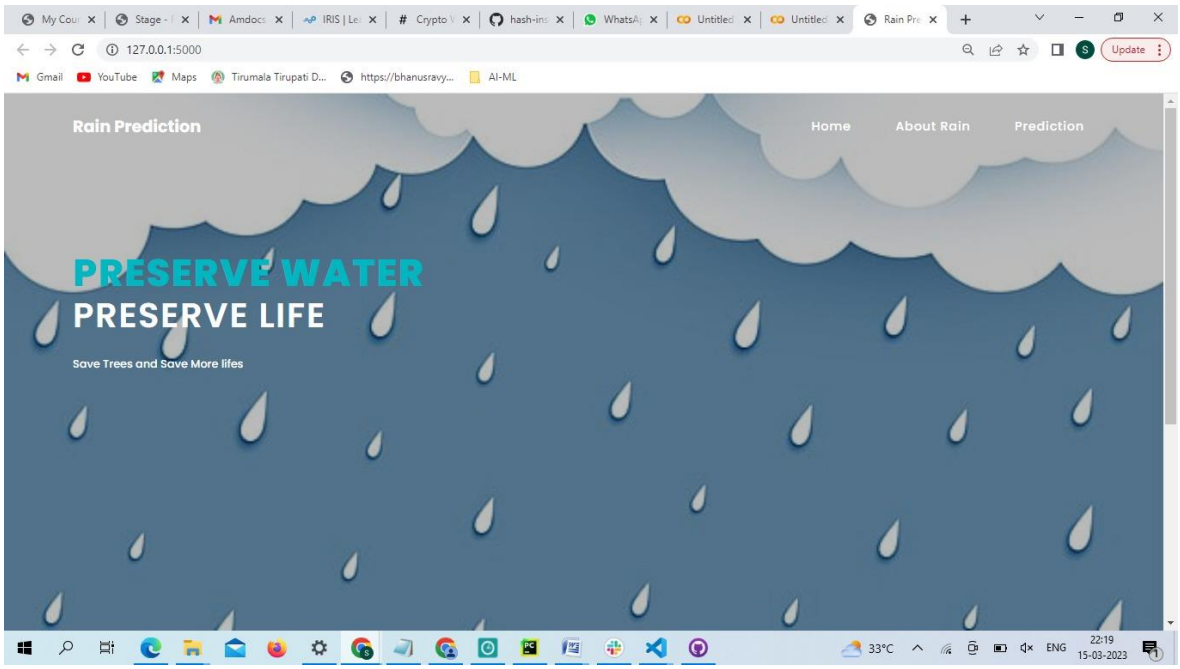


Fig: 4.1 Home Page

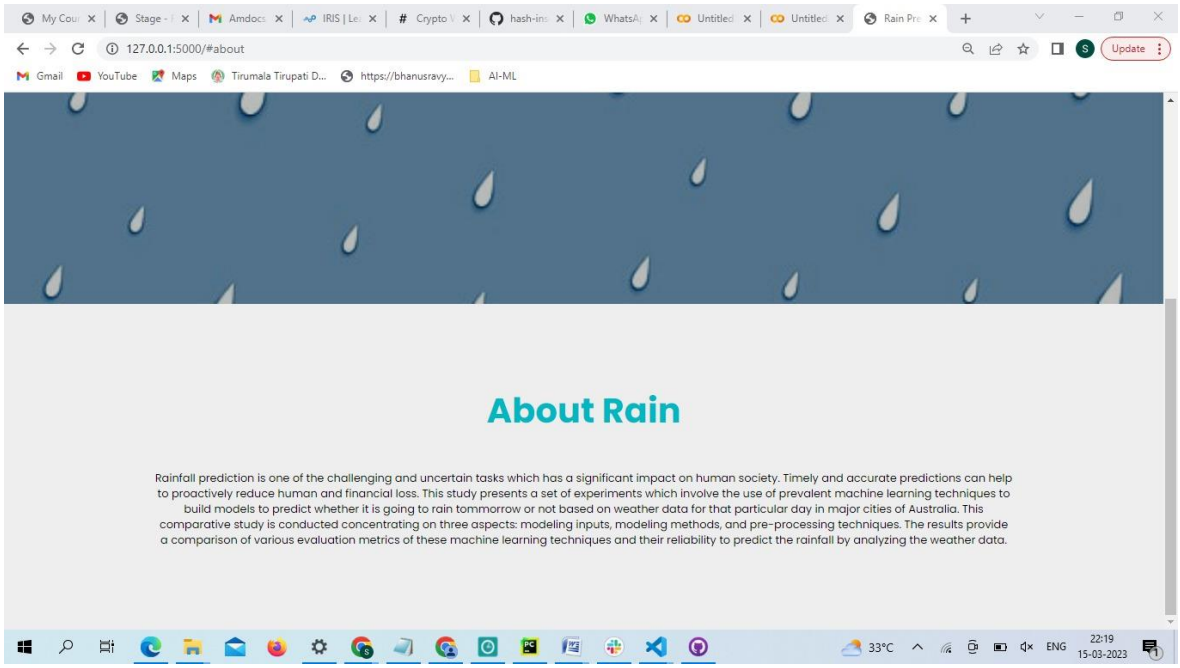


Fig:4.2 About Page

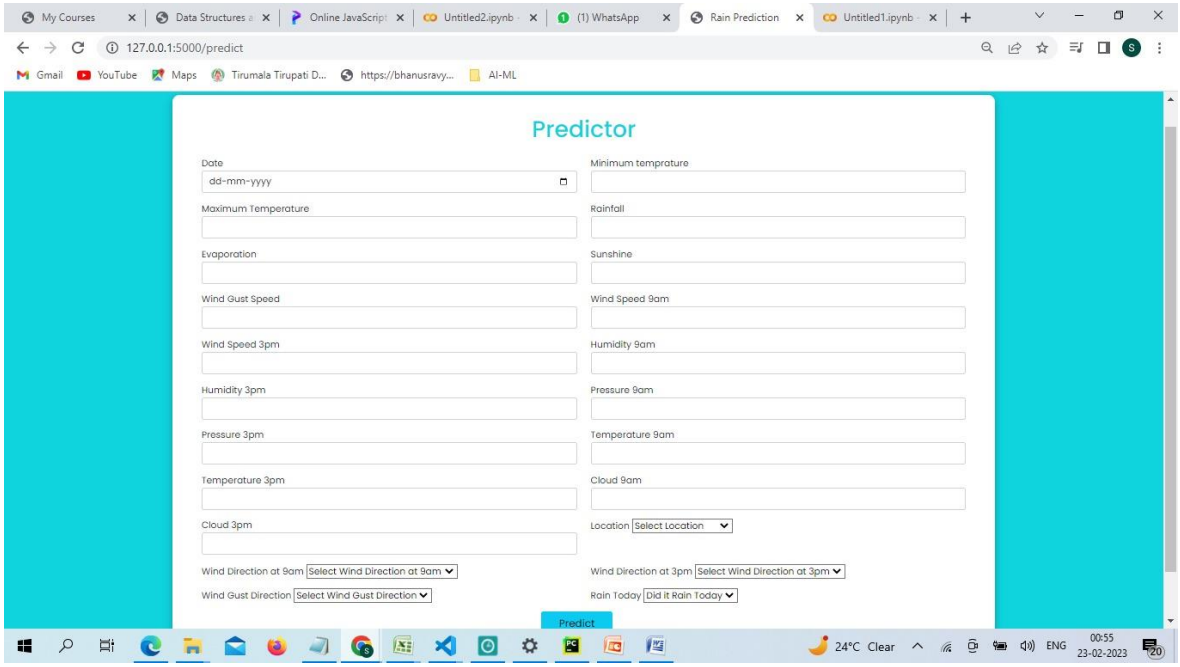


Fig:4.3 Prediction Page

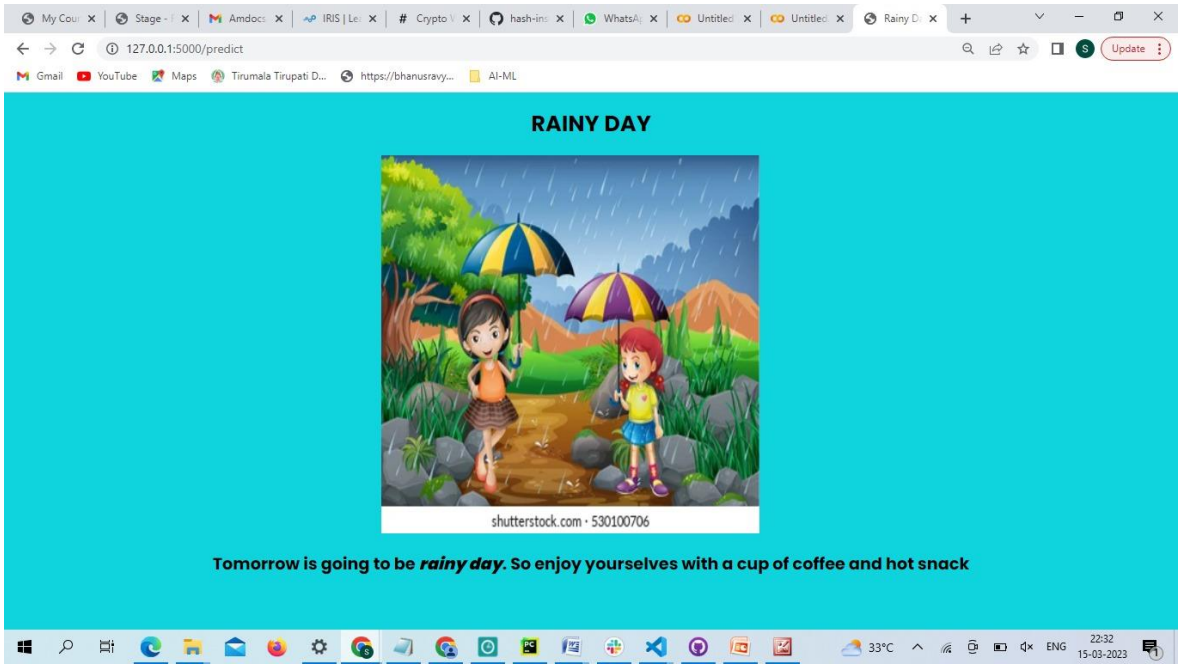


Fig:4.4 Rainy Day page

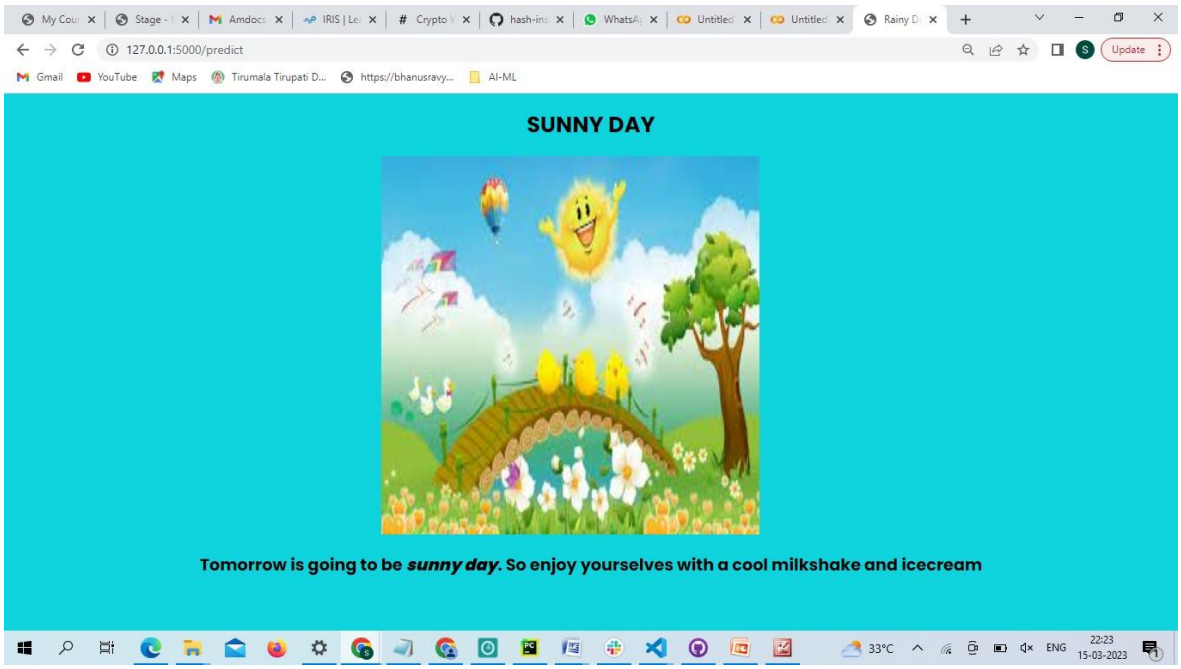


Fig:4.5 Sunny Day Page

5.CONCLUSION

In this project, we explored and applied several preprocessing steps and learned there impact on the overall performance of our classifiers. We also carried a comparative study of all the classifiers with different input data and observed how the input data can affect the model predictions. We can conclude that Australian weather is uncertain and there is no such correlation among rainfall and the respective region and time. We figured certain patterns and relationships among data which helped in determining important features. Refer to the appendix section.

6.FUTURE SCOPE

As we have a huge amount of data, we can apply Deep Learning models such as Multilayer Perceptron, Convolutional Neural Network, and others. It would be great to perform a comparative study between the Machine learning classifiers and Deep learning models.

7.BIBILOGRAPHY

1. World Health Organization: Climate Change and Human Health: Risks and Responses. World Health Organization, January 2003

2. Alcntara-Ayala, I.: Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. Geomorphology 47(24), 107124 (2002)

3. Nicholls, N.: Atmospheric and climatic hazards: Improved monitoring and prediction for disaster mitigation. Natural Hazards 23(23), 137155 (2001)

4. [Online] InDataLabs, Exploratory Data Analysis: the Best way to Start a Data Science Project. Available: https://medium.com/@InDataLabs/ why-start-a-data-science-project-with-exploratory-data-analysis-f90c0efcbe49

5. [Online] Pandas Documentation. Available: https://pandas.pydata.org/ pandas-docs/stable/reference/api/pandas.get_dummies.html

6. [Online] Sckit-Learn Documentation Available: https://scikit-learn.org/ stable/modules/generated/sklearn.feature_extraction.FeatureHasher. html

7. [Online] Sckit-Learn Documentation Available: https://scikit-learn.org/ stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

8. [Online] Sckit Learn Documentation Available: https://scikit-learn.org/ stable/modules/generated/sklearn.feature_selection.SelectKBest.html

Rainfall Prediction Using Machine Learning

N.Bhanu Sravya[1],P.Rafiya[2],K.Abhigna[3],A.Thanuja[4]

[1,2,3]Student, Department of CSE, Narasaraopeta Engineering College, Narasaraopeta, Guntur (D.T) A.P, India

[4]Professor, Department of CSE, Narasaraopeta Engineering College, Narasaraopeta, Guntur (D.T) A.P, India

nissankararaobhanusravya@gmail.com[1],rafiyapathan4085@gmail.com[2], karnatiabhignareddy@gmail.com[3],a.thanuja18@gmail.com[4]

ABSTRACT- "Rainfall Prediction Using Machine Learning" is the project's name. The dataset for this project is kept in Microsoft Excel, and the project is written in Python. Manymachine learning algorithms are used in this prediction to see which method makes the best accurate predictions. Forecasting rainfall is crucial in many areas of a nation and can aid in averting catastrophic natural catastrophes. Logistic Regression, Random Forest Classifier, Gradient Boosting, KNN, Decision Tree, Adaboost Classifier, and Catboost Classifier were allutilised to make this prediction. This project uses a total ofseven modules. The Australian rainfalldataset was utilised. The project's primarygoal is to evaluate different algorithms and identifythe top algorithmout ofthose algorithms. The farmers maygreatly benefit fromthis prediction by planting the appropriate crops based ontheir requirement for water.

KEYWORDS: Machine Learning, Logistic Regression,KNN, Random forest classifier, Gradient Boosting,Adaboost,Decision tree,Catboost.

1. INTRODUCTION

How to predict when it will rain is a topic that interests governments, corporations, risk management organisations, and the scientific community all at the same time. Rainfall is a climatic factor that affects a variety of human endeavours, such as tourism, forestry, construction, and agricultural production.

This project is used to predict the rainfall in the 49 cities of Australia.The prediction uses various algorithms. Forecasting rainfall is crucial in many areas of the nation and can aid in averting catastrophic natural catastrophes.

The goal of this study is to offer complete machine learning life cycle models. Here, we'll look at several model descriptions in more depth. The models in question are listed as follows:

1. DATA COLLECTION
2. DATA VISUALIZATION
3. DATA PREPROCESSING
4. MODEL SELECTION
5. PERFORMANCE EVALUATION.

The structure of the essay is as follows. In section 2, we first explain the dataset. Section 3 presents the strategies and approaches that were employed. In section 4, the outcomes are finally discussed.

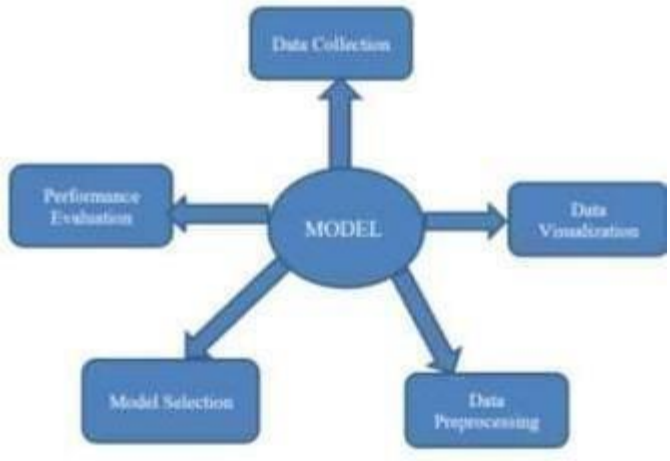


Fig.1(a) 5 steps involved in Model

2. DATASET

This section will cover every aspect of the dataset. We'll look at the location of the dataset, the properties it contains, and how each attribute is described in the dataset.

Several descriptions are provided for each characteristic.

The Attributes and datatype and no.of null values in each attribute of the dataset are described as shown below:

```
0   Date           142193 non-null  object
1   Location       142193 non-null  object
2   MinTemp        141556 non-null  float64
3   MaxTemp        141871 non-null  float64
4   Rainfall       140787 non-null  float64
5   Evaporation    81350 non-null   float64
6   Sunshine       74377 non-null   float64
7   WindGustDir    132863 non-null  object
8   WindGustSpeed  132923 non-null  float64
9   WindDir9am     132180 non-null  object
10  WindDir3pm     138415 non-null  object
11  WindSpeed9am   140845 non-null  float64
12  WindSpeed3pm   139563 non-null  float64
13  Humidity9am    140419 non-null  float64
14  Humidity3pm    138583 non-null  float64
15  Pressure9am    128179 non-null  float64
16  Pressure3pm    128212 non-null  float64
17  Cloud9am       88536 non-null   float64
18  Cloud3pm       85099 non-null   float64
19  Temp9am        141289 non-null  float64
20  Temp3pm        139467 non-null  float64
21  RainToday      140787 non-null  object
22  RISK_MM        142193 non-null  float64
23  RainTomorrow   142193 non-null  object
dtypes: float64(17), object(7)
memory usage: 26.0+ MB
```

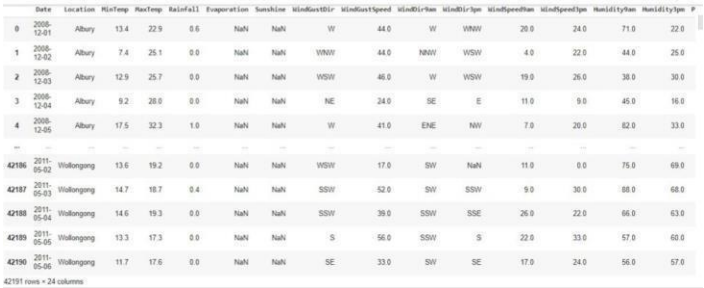
In the above dataset total we are having 23 attributes in the above metioned attributes our main aim is to predict wheather there will be rain tomorrow or not the main attribute is used for this prediction is **"RAINTOMORROW"** .

In our dataset we are having the null values in each and every attribute so we have to remove those null values .In order to remove those null values we have the concept of data preprocessing.In this data preprocessing we will be using the data cleaing technique.The description of the attributes are shown in the below picture.

Feature	Description
Date	The date of observation
Location	The common name of the location of the weather station
MinTemp	The minimum temperature in degrees celsius
MaxTemp	The maximum temperature in degrees celsius
Rainfall	The amount of rainfall recorded for the day in mm
Evaporation	The so-called Class A pan evaporation (mm) in the 24 hours to 9am
Sunshine	The number of hours of bright sunshine in the day.
WindGustDir	The direction of the strongest wind gust in the 24 hours to midnight
WindGustSpeed	The speed (km/h) of the strongest wind gust in the 24 hours to midnight
WindDir9am	Direction of the wind at 9am
WindDir3pm	Direction of the wind at 3pm
WindSpeed9am	Wind speed (km/hr) averaged over 10 minutes prior to 9am
WindSpeed3pm	Wind speed (km/hr) averaged over 10 minutes prior to 3pm
Humidity9am	Humidity (percent) at 9am
Humidity3pm	Humidity (percent) at 3pm
Pressure9am	Atmospheric pressure (hpa) reduced to mean sea level at 9am
Pressure3pm	Atmospheric pressure (hpa) reduced to mean sea level at 3pm
Cloud9am	Fraction of sky obscured by cloud at 9am.
Cloud3pm	Fraction of sky obscured by cloud at 3pm.
Temp9am	Temperature (degrees C) at 9am
Temp3pm	Temperature (degrees C) at 3pm
RainToday	1 if precipitation exceeds 1mm, otherwise 0
RISK_MM	The amount of next day rain in mm.
RainTomorrow	The target variable. Did it rain tomorrow?

Fig 2.1.weatherAUS.csv

The total no.of rows in the dataset is 42191 rows for 24 columns.Sample data in the dataset is shown in the below picture format.



	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	P
0	2008-12-01	Albury	13.4	22.9	0.6	NaN	NaN	W	44.0	W	WNW	20.0	24.0	71.0	22.0	
1	2008-12-02	Albury	7.4	25.1	0.0	NaN	NaN	WNW	44.0	NNW	WSW	4.0	22.0	44.0	25.0	
2	2008-12-03	Albury	12.9	25.7	0.0	NaN	NaN	WSW	46.0	W	WSW	19.0	26.0	38.0	30.0	
3	2008-12-04	Albury	9.2	28.0	0.0	NaN	NaN	NE	24.0	SE	E	11.0	9.0	45.0	16.0	
4	2008-12-05	Albury	17.5	32.3	1.0	NaN	NaN	W	41.0	ENE	NW	7.0	20.0	82.0	33.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
42186	2011-05-02	Wollongong	13.6	19.2	0.0	NaN	NaN	WSW	17.0	SW	NaN	11.0	0.0	75.0	69.0	
42187	2011-05-03	Wollongong	14.7	18.7	0.4	NaN	NaN	SSW	52.0	SW	SSW	9.0	30.0	88.0	68.0	
42188	2011-05-04	Wollongong	14.6	19.3	0.0	NaN	NaN	SSW	39.0	SSW	SSE	26.0	22.0	66.0	63.0	
42189	2011-05-05	Wollongong	13.3	17.3	0.0	NaN	NaN	S	56.0	SSW	S	22.0	33.0	57.0	60.0	
42190	2011-05-06	Wollongong	11.7	17.6	0.0	NaN	NaN	SE	33.0	SW	SE	17.0	24.0	56.0	57.0	

42191 rows × 24 columns

Fig 2.2.dataset

The irrelevant features in the above dataset is mentioned below:

1.Sunshine with 43% of null values.
2.Evaporation with 48% of null values.
3.cloud 3pm with 43% of null values.
4.cloud 9am with 38% of null values.

3.Methodology

Methodology is nothing but used methods and AI algorithms in our project here we are discussing the algorithms used in our project brefily.The algorithms used in our project are discussed below.Here we used the seven algorithms in order to predict the best one based on the accuracy percentage they got.the algorithms are:

1. KNN
2. Random Forest classifier
3. Logistic Regression
4. Gradient Boosting classifier
5. Adaboost
6. Decision Tree
7. Catboost

Now we will see about these algorithms one by one in detail brefily.

1. KNN:

The full form of KNN is K-Nearest Neighbour .This algorithm is one of the simplest Machine Learning algorithm.This comes under the Supervised Machine Learning Technique.This Algorithm can be used for both regression and classification.Among those two mostly this is used for classification problems.This algorithm can also be called as the **LAZY LEARNING** Algorithm.

2. Random Forest Classifier:

This is one of the popular Machine Learning Algorithm which comes under the Supervised Machine Learning Technique.In this Random Forest Classifier these will produce the more no.of tress among those trees we have to take the best tree that gives the more accuracy.

3. Logistic Regression:

This Logistic Regression is an example for the Supervised Machine Learning.This Algorithm mostly use to predict the probability for the occurring of binary event.There are three types of Logistic Regression.These are mentioned below:

1.Binary Logistic Regression
2.Multinomial Logistic Regression
3.Ordinal Logistic Regression

These are the three types of LOGISTIC REGRESSION.

4. Gradient Boosting Classifier:

This algorithm is a machine learning technique which is used in classification and regression.This Classifier is present in the ensemble model.This gives the outcome as the binary tree.Based on those we need to take the best part which we will get the less accuracy.In this Algorithm we will use the important parameter named **shrinkage.**

This Gradient Boosting Classifier is the Supervised Machine Learning Algorithm.

5. Adaboost Classifier:

This Adaboost Algorithm is a Boosting Technique this can be find in the Ensemble Method in Machine Learning.This Adaboost can be called as the Adaptive Boosting Algorithm.This Algorithm is First Successful boosting algorithm.This algorithm is developed for binary classification purpose.This is very important boosting technique.this converts the multiple “weak classifiers” into single “strong classifier”.

6. Decision Tree Classifier:

This Decision Tree Algorithm is a Supervised Machine Learning Algorithm.This can be used for both Classification and Regression.Mostly we use this for Classification problems.The format for this is tree-structured format.There will be two kinds of nodes.These are:

1.Decision Node
2.Leaf Node

In the Decision Node ther will be extention of tree,where as for the Leaf Node there will be no extention.This will consider as the final output.

7.Catboost Classifier:

This catboost Classifier is an open-source library.This Algorithm comes under the gradient Boosting classifer.where we can use the decision tree.this algorithm is developed by **YANDEX RESEARCHERS AND ENGINEERS**.This catboost classifier algorithm can be used easily.

DATA PREPROCESSING

The data preprocessing is nothing but which is used to convert the raw data into the clean dataset.For example rawdata is nothing but having the null values.The machine Lerning Algorithm can not understand those null values our aim is to remove those null values.For this process of removing null values we will use the data cleaning step in the data preprocessing steps.The data preprocessing can be applied to the dataset before we use this dataset in our algorithm.Like wise also the Ranforest Algorithm can not perform analysis if the dataset contains the null values.The data preprocessing can also be used in order to format our dataset in particular way.The steps involved in the data preprocessing are mentioned as shown below.

1. Having Dataset
2. Import Required Libraries
3. Loading Dataset
4. Identifying Missing Data
5. Encoding Categorical Data
6. Splitting Dataset into Train and Test Datasets.
7. Feature Scaling.

These are seven steps involved in the **data preprocessing** process.After completion of these seven steps we call this dataset as the clean dataset.Now this dataset can used inour required Machine Learning Algorithms.

4.Experiments And Results

In this final step we are going to evaluate the accuracy for the Australian Dataset by using the different machine learning algorithms.The Algorithms we used are KNN,Random forest,Decision Tree,Catboost,Adaboost,Gradient Boosting,Logistic Regression.

Before this we need to do the Data Preprocessing step.we need to train and test our dataset set to get the accurate results.In this step we will find which algorithm is best to use in our project based upon the accuracy score we get for different machine learning algorithms.

Now,we will observe code for the Catboost algorithm and the same code is used for all the algorithms but,we need to change the importing statements.The sample code is provided below:

```
From sklearn.ensemble import CatBoostClassifier

model =CatBoostClassifier(iterations=2000, eval_metric = "AUC")
model.fit(x_train, y_train)

y_pred= model.predict(x_test)

from sklearn.metrics import accuracy_score
ac = accuracy_score(y_pred, y_test)
#output - 0.86
```

Now,we will observe Accuracy for all the algorithms.

ALGORITHM	ACCURACY
Logistic Regression	79%
Decision Tree	73%
Random Forest	81%
KNN	80%
Gradient Boosting	81%
AdaBoost	80%
CatBoost	86%

By observing above table comparing algorithms we observe that catBoost classifier has highest accuracy and Decision tree has least accuracy.So for our project we took CatBoost classifier Algorithm.

CONCLUSION

In this work, we explored and applied many preprocessing techniques to find out how they impacted the overall performance of our classifiers. We also compared every classifier using different inputs, making note of how the entering data can affect the predictions made by the model.

We can infer that Australian weather is erratic and that there is no connection between rainfall and a certain location or time. We found a number of links and trends in the data, allowing us to pinpoint important traits.

Because of the large quantity of data we have, we may employ Deep Learning models like Multilayer Perceptrons, Convolutional Neural Networks (CNN), and others. It would be great to compare Deep Learning models and Machine Learning classifiers.

REFERENCES

1. World Health Organization: Climate Change and Human Health: Risks and Responses. World Health Organization, January 2003
2. Alcntara-Ayala, I.: Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. Geomorphology 47(24), 107124 (2002)
3. Nicholls, N.: Atmospheric and climatic hazards: Improved monitoring and prediction for disaster mitigation. Natural Hazards 23(23), 137155 (2001)
4. [Online] InDataLabs, Exploratory Data Analysis: the Best way to Start a Data Science Project. Available: https://medium.com/@InDataLabs/ why-start-a-data-science-project-with-exploratory-data-analysis-f90c0efcbe49
5. [Online] Pandas Documentation. Available: https://pandas.pydata.org/ pandas-docs/stable/reference/api/pandas.get_dummies.html
6. [Online] Sckit-Learn Documentation Available: https://scikitlearn.org/stable/modules/generated/sklearn.feature_extraction.FeatureHasher. html
7. [Online] Sckit-Learn Documentation Available: https://scikitlearn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

Rainfall Prediction

by sravya bhanu

Submission date: 17-Mar-2023 01:03PM (UTC-0400)
Submission ID: 2039472694
File name: rainfall_prediction_1.pdf (601.62K)
Word count: 1788
Character count: 9862

Rainfall Prediction Using Machine Learning

N.Bhanu Sravya[1],P.Rafiya[2],K.Abhigna[3],A.Thanuja[4]

[1,2,3]Student, Department of CSE, Narasaraopeta Engineering College, Narasaraopeta, Guntur (D.T) A.P, India

[4]Professor, Department of CSE, Narasaraopeta Engineering College, Narasaraopeta, Guntur (D.T) A.P, India

nissankararaobhanusravya@gmail.com[1], rafiyapathan4085@gmail.com[2], karnatiabhignareddy@gmail.com[3], a.thanuja18@gmail.com[4]

ABSTRACT- "Rainfall Prediction Using Machine Learning" is the project's name. The dataset for this project is kept in Microsoft Excel, and the project is written in Python. Many machine learning algorithms are used in this prediction to see which method makes the best accurate predictions. Forecasting rainfall is crucial in many areas of a nation and can aid in averting catastrophic natural catastrophes. Logistic Regression, Random Forest Classifier, Gradient Boosting, KNN, Decision Tree, Adaboost Classifier, and Catboost Classifier were all utilised to make this prediction. This project uses a total of seven modules. The Australian rainfall dataset was utilised. The project's primary goal is to evaluate different algorithms and identify the top algorithm out of those algorithms. The farmers may greatly benefit from this prediction by planting the appropriate crops based on their requirement for water.

KEYWORDS: Machine Learning, Logistic Regression,KNN, Random forest classifier, Gradient Boosting,Adaboost,Decision tree,Catboost.

1. INTRODUCTION

How to predict when it will rain is a topic that interests governments, corporations, risk management organisations, and the scientific community all at the same time. Rainfall is a climatic factor that affects a variety of human endeavours, such as tourism, forestry, construction, and agricultural production.

This project is used to predict the rainfall in the 49 cities of Australia.The prediction uses various algorithms. Forecasting rainfall is crucial in many areas of the nation and can aid in averting catastrophic natural catastrophes.

The goal of this study is to offer complete machine learning life cycle models. Here, we'll look at several model descriptions in more depth. The models in question are listed as follows:

1. DATA COLLECTION

2. DATA VISUALIZATION

3. DATA PREPROCESSING

4. MODEL SELECTION

5. PERFORMANCE EVALUATION.

The structure of the essay is as follows. In section 2, we first explain the dataset. Section 3 presents the strategies and approaches that were employed. In section 4, the outcomes are finally discussed.

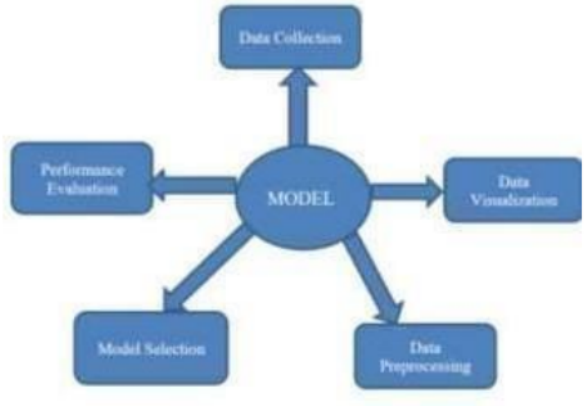


Fig.1(a) 5 steps involved in Model

2. DATASET

This section will cover every aspect of the dataset. We'll look at the location of the dataset, the properties it contains, and how each attribute is described in the dataset.

Several descriptions are provided for each characteristic.

The Attributes and datatype and no.of null values in each attribute of the dataset are described as shown below:

```
 0   Date           142193 non-null  object
 1   Location       142193 non-null  object
 2   MinTemp        141556 non-null  float64
 3   MaxTemp        141871 non-null  float64
 4   Rainfall       140787 non-null  float64
 5   Evaporation    81350 non-null   float64
 6   Sunshine       74377 non-null   float64
 7   WindGustDir    132863 non-null  object
 8   WindGustSpeed  132923 non-null  float64
 9   WindDir9am     132180 non-null  object
 10  WindDir3pm     138415 non-null  object
 11  WindSpeed9am   140845 non-null  float64
 12  WindSpeed3pm   139563 non-null  float64
 13  Humidity9am    140419 non-null  float64
 14  Humidity3pm    138583 non-null  float64
 15  Pressure9am    128179 non-null  float64
 16  Pressure3pm    128212 non-null  float64
 17  Cloud9am       88536 non-null   float64
 18  Cloud3pm       85099 non-null   float64
 19  Temp9am        141289 non-null  float64
 20  Temp3pm        139467 non-null  float64
 21  RainToday      140787 non-null  object
 22  RISK_MM        142193 non-null  float64
 23  RainTomorrow   142193 non-null  object
dtypes: float64(17), object(7)
memory usage: 26.0+ MB
```

In the above dataset total we are having 23 attributes in the above metioned attributes our main aim is to predict wheather there will be rain tomorrow or not the main attribute is used for this prediction is **"RAINTOMORROW"** .

In our dataset we are having the null values in each and every attribute so we have to remove those null values .In order to remove those null values we have the concept of data preprocessing.In this data preprocessing we will be using the data cleaing technique.The description of the attributes are shown in the below picture.

Feature	Description
Date	The date of observation
Location	The common name of the location of the weather station
MinTemp	The minimum temperature in degrees celsius
MaxTemp	The maximum temperature in degrees celsius
Rainfall	The amount of rainfall recorded for the day in mm
Evaporation	The so-called Class A pan evaporation (mm) in the 24 hours to 9am
Sunshine	The number of hours of bright sunshine in the day.
WindGustDir	The direction of the strongest wind gust in the 24 hours to midnight
WindGustSpeed	The speed (km/h) of the strongest wind gust in the 24 hours to midnight
WindDir9am	Direction of the wind at 9am
WindDir3pm	Direction of the wind at 3pm
WindSpeed9am	Wind speed (km/hr) averaged over 10 minutes prior to 9am
WindSpeed3pm	Wind speed (km/hr) averaged over 10 minutes prior to 3pm
Humidity9am	Humidity (percent) at 9am
Humidity3pm	Humidity (percent) at 3pm
Pressure9am	Atmospheric pressure (hpa) reduced to mean sea level at 9am
Pressure3pm	Atmospheric pressure (hpa) reduced to mean sea level at 3pm
Cloud9am	Fraction of sky obscured by cloud at 9am.
Cloud3pm	Fraction of sky obscured by cloud at 3pm.
Temp9am	Temperature (degrees C) at 9am
Temp3pm	Temperature (degrees C) at 3pm
RainToday	1 if precipitation exceeds 1mm, otherwise 0
RISK_MM	The amount of next day rain in mm.
RainTomorrow	The target variable. Did it rain tomorrow?

Fig 2.1.weatherAUS.csv

The total no.of rows in the dataset is 42191 rows for 24 columns.Sample data in the dataset is shown in the below picture format.

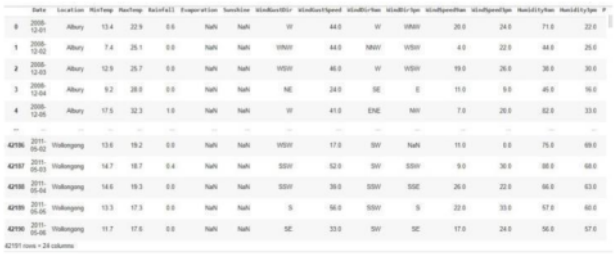


Fig 2.2.dataset

The irrelevant features in the above dataset is mentioned below:

1.Sunshine with 43% of null values.
2.Evaporation with 48% of null values.
3.cloud 3pm with 43% of null values.
4.cloud 9am with 38% of null values.

3.Methodology

Methodology is nothing but used methods and AI algorithms in our project here we are discussing the algorithms used in our project brefily.The algorithms used in our project are discussed below.Here we used the seven algorithms in order to predict the best one based on the accuracy percentage they got.the algorithms are:

1.KNN
2.Random Forest classifier
3.Logistic Regression
4.Gradient Boosting classifier
5.Adaboost
6.Decision Tree
7.Catboost

Now we will see about these algorithms one by one in detail brefily.

1.KNN:

The full form of KNN is K-Nearest Neighbour .This algorithm is one of the simplest Machine Learning algorithm.This comes under the Supervised Machine Learning Technique.This Algorithm can be used for both regression and classification.Among those two mostly this is used for classification problems.This algorithm can also be called as the **LAZY LEARNING** Algorithm.

2.Random Forest Classifier:

This is one of the popular Machine Learning Algorithm which comes under the Supervised Machine Learning Technique.In this Random Forest Classifier these will produce the more no.of tress among those trees we have to take the best tree that gives the more accuracy.

3.Logistic Regression:

This Logistic Regression is an example for the Supervised Machine Learning.This Algorithm mostly use to predict the probability for the occurring of binary event.There are three types of Logistic Regression.These are mentioned below:

1.Binary Logistic Regression
2.Multinomial Logistic Regression
3.Ordinal Logistic Regression

These are the three types of LOGISTIC REGRESSION.

4.Gradient Boosting Classifier:

This algorithm is a machine learning technique which is used in classification and regression.This Classifier is present in the ensemble model.This gives the outcome as the binary tree.Based on those we need to take the best part which we will get the less accuracy.In this Algorithm we will use the important parameter named **shrinkage.**

This Gradient Boosting Classifier is the Supervised Machine Learning Algorithm.

5.Adaboost Classifier:

This Adaboost Algorithm is a Boosting Technique this can be find in the Ensemble Method in Machine Learning.This Adaboost can be called as the Adaptive Boosting Algorithm.This Algorithm is First Successful boosting algorithm.This algorithm is developed for binary classification purpose.This is very important boosting technique.this converts the multiple "weak classifiers" into single "strong classifier".

6.Decision Tree Classifier:

This Decision Tree Algorithm is a Supervised Machine Learning Algorithm.This can be used for both Classification and Regression.Mostly we use this for Classification problems.The format for this is tree-structured format.There will be two kinds of nodes.These are:

1.Decision Node
2.Leaf Node

In the Decision Node ther will be extention of tree,where as for the Leaf Node there will be no extention.This will consider as the final output.

7.Catboost Classifier:

This catboost Classifier is an open-source library.This Algorithm comes under the gradient Boosting classifer.where we can use the decision tree.this algorithm is developed by **YANDEX RESEARCHERS AND ENGINEERS**.This catboost classifier algorithm can be used easily.

DATA PREPROCESSING

The data preprocessing is nothing but which is used to convert the raw data into the clean dataset.For example rawdata is nothing but having the null values.The machine Lerning Algorithm can not understand those null values our aim is to remove those null values.For this process of removing null values we will use the data cleaning step in the data preprocessing steps.The data preprocessing can be applied to the dataset before we use this dataset in our algorithm.Like wise also the Ranforest Algorithm can not perform analysis if the dataset contains the null values.The data preprocessing can also be used in order to format our dataset in particular way.The steps involved in the data preprocessing are mentioned as shown below.

1. Having Dataset
2. Import Required Libraries
3. Loading Dataset
4. Identifying Missing Data
5. Encoding Categorical Data
6. Splitting Dataset into Train and Test Datasets.
7. Feature Scaling.

These are seven steps involved in the **data preprocessing** process.After completion of these seven steps we call this dataset as the clean dataset.Now this dataset can used inour required Machine Learning Algorithms.

4.Experiments And Results

In this final step we are going to evaluate the accuracy for the Australian Dataset by using the different machine learning algorithms.The Algorithms we used are KNN,Random forest,Decision Tree,Catboost,Adaboost,Gradient Boosting,Logistic Regression.

Before this we need to do the Data Preprocessing step.we need to train and test our dataset set to get the accurate results.In this step we will find which algorithm is best to use in our project based upon the accuracy score we get for different machine learning algorithms.

Now,we will observe code for the Catboost algorithm and the same code is used for all the algorithms but,we need to change the importing statements.The sample code is provided below:

```
From sklearn.ensemble import CatBoostClassifier
model =CatBoostClassifier(iterations=2000, eval_metric = "AUC")
model.fit(x_train, y_train)

y_pred= model.predict(x_test)

from sklearn.metrics import accuracy_score
ac = accuracy_score(y_pred, y_test)
#output - 0.86
```

Now,we will observe Accuracy for all the algorithms.

ALGORITHM	ACCURACY
Logistic Regression	79%
Decision Tree	73%
Random Forest	81%
KNN	80%
Gradient Boosting	81%
AdaBoost	80%
CatBoost	86%

By observing above table comparing algorithms we observe that catBoost classifier has highest accuracy and Decision tree has least accuracy.So for our project we took CatBoost classifier Algorithm.

CONCLUSION

In this work, we explored and applied many preprocessing techniques to find out how they impacted the overall performance of our classifiers. We also compared every classifier using different inputs, making note of how the entering data can affect the predictions made by the model.

We can infer that Australian weather is erratic and that there is no connection between rainfall and a certain location or time. We found a number of links and trends in the data, allowing us to pinpoint important traits.

Because of the large quantity of data we have, we may employ Deep Learning models like Multilayer Perceptrons, Convolutional Neural Networks (CNN), and others. It would be great to compare Deep Learning models and Machine Learning classifiers.

REFERENCES

1. World Health Organization: Climate Change and Human Health: Risks and Responses. World Health Organization, January 2003
2. Alcntara-Ayala, I.: Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. Geomorphology 47(24), 107124 (2002)
3. Nicholls, N.: Atmospheric and climatic hazards: Improved monitoring and prediction for disaster mitigation. Natural Hazards 23(23), 137155 (2001)
4. [Online] InDataLabs, Exploratory Data Analysis: the Best way to Start a Data Science Project. Available: https://medium.com/@InDataLabs/ why-start-a-data-science-project-with-exploratory-data-analysis-f90c0efcbe49
5. [Online] Pandas Documentation. Available: https://pandas.pydata.org/ pandas-docs/stable/reference/api/pandas.get_dummies.html
6. [Online] Sckit-Learn Documentation Available: https://scikitlearn.org/stable/modules/generated/sklearn.feature_extraction.FeatureHasher. html
7. [Online] Sckit-Learn Documentation Available: https://scikitlearn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html

Rainfall Prediction

ORIGINALITY REPORT

11%	9%	6%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	arxiv.org Internet Source	3%
2	medium.com Internet Source	1%
3	S. Siva Sunayna, S. N. Thirumala Rao, M. Sireesha. "Chapter 25 Performance Evaluation of Machine Learning Algorithms to Predict Breast Cancer", Springer Science and Business Media LLC, 2022 Publication	1%
4	Submitted to Monash University Student Paper	1%
5	Submitted to British University in Egypt Student Paper	1%
6	Damir Filipović, Ludger Overbeck, Thorsten Schmidt. "DYNAMIC CDO TERM STRUCTURE MODELING", Mathematical Finance, 2011 Publication	1%
7	www.researchgate.net Internet Source	1%

8	morioh.com Internet Source	1%
9	www.conftool.pro Internet Source	1%
10	virgoady7.medium.com Internet Source	1%
11	github.com Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

International Conference on

PAPER ID
NECICAIEA2K23042

Artificial Intelligence and Its Emerging Areas

NEC-ICAIEA-2K23

17th & 18th March, 2023

Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that N.Bhanu Sravya , Narasaraopeta Engineering College has presented the paper title Rainfall Prediction in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

International Conference on

PAPER ID
NECICAIEA2K23042

Artificial Intelligence and Its Emerging Areas

NEC-ICAIEA-2K23

17th & 18th March, 2023

Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that P.Rafiya , Narasaraopeta Engineering College has presented the paper title Rainfall Prediction in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

International Conference on

PAPER ID
NECICAIEA2K23042

Artificial Intelligence and Its Emerging Areas

NEC-ICAIEA-2K23

17th & 18th March, 2023

Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that K.Abhigna , Narasaraopeta Engineering College has presented the paper title Rainfall Prediction in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar

NARASARAOPETA ENGINEERING COLLEGE
(AUTONOMOUS)

Approved by AICTE, Permanently Affiliated to JNTUK, Kakinada, NIRF Ranking (251-300 Band), Accredited by NBA (Tier-I) & NAAC with 'A+' Grade
Kotappakonda Road, Yellamanda (Post), Narasaraopet - 522601, Palnadu Dist., Andhra Pradesh, INDIA. Website:www.nrtec.in

International Conference on

PAPER ID
NECICAIEA2K23042

Artificial Intelligence and Its Emerging Areas

NEC-ICAIEA-2K23

17th & 18th March, 2023

Organized by Department of Computer Science and Engineering in Association with CSI

Certificate of Presentation

This is to Certify that A.Thanuja , Narasaraopeta Engineering College has presented the paper title Rainfall Prediction in the International Conference on Artificial Intelligence and Its Emerging Areas-2K23 [NEC-ICAIEA-2K23], Organized by Department of Computer Science and Engineeringin Association with CSI on 17th and 18th March 2023 at Narasaraopeta Engineering College, Narasaraopet, A.P., India.

Convenor
Dr.S.V.N.Srinivasu

Chief-Convenor
Dr.S.N.Tirumala Rao

Principal, Patron
Dr. M. Sreenivasa Kumar