**VISVESVARAYA TECHNOLOGICAL UNIVERSITY**

Jnana Sangama, Machhe, Belagavi, Karnataka 590018



A

Project report

on

# IMPACT OF TEMPERATURE VARIATIONS OVER THE BAY OF BENGAL ON THE CLIMATE OF EASTERN COAST OF INDIA.

*Submitted in partial fulfillment of the requirement*

*for the award of the degree of*

Bachelor of Engineering

in

# Information Science & Engineering

by

**Bhavana (1BG14IS008)**

**H Sudhanva (1BG14IS017)**

**Sai Navaneeth V (1BG14IS041)**

**Satish Kumar M S (1BG14IS045)**

Under the Guidance of

**Dr. Saritha Chakrasali**

Professor

Dept. of ISE

B.N.M.I.T



*Vidyaya Amrutham Ashnuthe*

**B.N.M. Institute of Technology**

12th Main, 27th Cross, Banashankari II Stage, Bangalore 560 070.

Department of Information Science and Engineering

2017-2018

**B.N.M. Institute of Technology**

12th Main, 27th Cross Banashankari II Stage, Bangalore - 560070

DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING



*Vidyaya Amrutham Ashnuthe*

##### CERTIFICATE

# Certified that the of project work entitled, “Impact of temperature variations over the Bay of Bengal on the eastern coast of India” is carried out by Ms. Bhavana (1BG14IS008) , Mr. H Sudhanva (1BG14IS017) , Mr. Sai Navaneeth V (1BG14IS041) , Mr. Satish Kumar M S (1BG14IS045) , the bonafide students of B.N.M Institute of Technology in partial fulfillment for the award of Bachelor of Engineering in Information Science & Engineering of the Visvesvaraya Technological University, Belagavi during the year 2017-2018. It is certified that all corrections / suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The project has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said Degree.

**Dr. Saritha Chakrasali Dr. Surabhi Narayan**

**Professor , Dept. of ISE Prof & Head, Dept. of ISE**

**BNMIT** **BNMIT**

**Name of Examiners Signature with date**

**1.**

**2.**

**ABSTRACT OF THE PROJECT**

The increasing research in the fields of Artificial Intelligence and Machine Learning has given rise to numerous weather prediction models. Yet, the problem of accurately predicting or forecasting the weather still persists. Numerical weather prediction, is taking the existing numerical data on weather conditions and applying machine learning algorithms on it to forecast the weather. Weather forecasting has traditionally been done by physical models of the atmosphere, which is unstable to perturbations, and thus is inaccurate for large periods of time. Weather is a continuous, data-intensive, multidimensional, dynamic and chaotic process and these properties make weather prediction a big challenge. Machine Learning techniques are more robust to perturbations, it would be more ideal to explore their applications in the field of weather forecasting and to potentially generate more accurate forecasts of weather for large periods of time. This project attempts to predict the changes in the temperature of cities in the coastal region using Machine Learning algorithms, by analyzing the statistical climate data of the city Chennai, along with the corresponding sea surface temperature of Bay of Bengal. This work aims at studying the relationship and establishing a pattern between the climatic changes in the land and its associated sea surface temperature using Data Mining techniques and Machine Learning algorithms.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| SL NO. | TITLE | PAGE NO. |
| **1** | **Introduction**   * 1. Overview   2. Objective   3. Motivation |  |
| **2** | **Literature Survey**   * 1. Project Background History   2. Existing System      1. Limitations of the Existing System   3. Proposed System      1. Overview of Machine Learning      2. Outline of Machine Learning   4. Advantages of Proposed System |  |
| **3** | **System Requirements Specification**  3.1. Hardware Requirements  3.2. Software Requirements  3.3. Functional Requirements  3.4. Non-Functional Requirements |  |
| **4** | **Cost Estimation of the Project** |  |
| **5** | **System Design and Development**  5.1. System Design  5.2. DFD/UML Diagrams  5.1.1. Dataflow Diagram |  |
| **6** | **Schedule of the project** |  |
| **7** | **Implementation**  7.1. Introduction  7.2. Technologies and Tools Used  7.2.1. Python  7.2.2. Spyder  7.2.3. Anaconda  7.3. Dataset Description  7.4. List of modules  7.4.1. Data Collection  7.4.2. Data Cleaning and Filtering  7.4.3. Feature selection  7.4.4. Implementing Regression Model  7.4.5. Testing  7.5. Description of Modules  7.6. Main Algorithm |  |
| **8** | **Testing and Validation**  8.1. Testing Process  8.2. Test cases and Validation |  |
| **9** | **Results and Snapshots** |  |
| **10** | **Conclusion and Future Enhancements**  10.1. Conclusion  10.2. Future Directions |  |
|  | References |  |

CHAPTER – 1

**INTRODUCTION**

**CHAPTER 1**

**INTRODUCTION**

* 1. **Overview**

Artificial intelligence (AI) traditionally refers to an artificial creation of human-like intelligence that can learn, reason, plan, perceive, or process natural language. Artificial intelligence is further defined as “narrow AI” or “general AI”. Narrow AI, which we interact with today, is designed to perform specific tasks within a domain (e.g. language translation). General AI is hypothetical and not domain specific, but can learn and perform tasks anywhere. This project focuses on advances in narrow AI, particularly on the development of new algorithms and models referred to as *machine learning*. Computer algorithms organize enormous amounts of data into information and services, based on certain instructions and rules. It’s an important concept to understand, because in machine learning, learning algorithms – not computer programmers – create the rules. Instead of programming the computer every step of the way, this approach gives the computer instructions that allow it to learn from data without new step-by-step instructions by the programmer.

Data analytics refers to qualitative and quantitative techniques and processes used to enhance productivity and business gain. Data is extracted and categorized to identify and analyze behavioral data and patterns, and techniques vary according to organizational requirements

Weather simply refers to the condition of air on earth at a given place and time. The application of science and technology are to predict the state of the atmosphere in future time for a given location which is important due to its effectiveness in human life. Today, weather forecasts are made by collecting quantitative data about the current state of the atmosphere and using scientific understanding of atmospheric processes to project how the atmosphere will evolve. The chaotic nature of the atmosphere implies the need of massive computational power required to solve the equations that describe the atmospheric conditions. Various papers related to weather forecasting using Machine Learning algorithms and other data mining concepts were studied. The relevant papers have been discussed below.

* 1. **Objective**

The objective of our project is to analyze related weather data using data mining techniques and study the correlation between the sea surface temperature and the corresponding land surface temperature over the Bay of Bengal.

* 1. **Motivation**

The motivation behind this project is to be able to recognize and establish a pattern between sea and land climate conditions upon considering a large data set for a specific location. Thus, being able to forecast weather more efficiently while improving the process of doing so.

CHAPTER – 2

**LITERATURE SURVEY**

**CHAPTER 2**

**LITERATURE SURVEY**

**2.1. Project Background History**

Atmospheric sciences is the study of the [Earth's atmosphere](https://en.wikipedia.org/wiki/Atmosphere_of_Earth), its processes, the effects other systems have on the atmosphere, and the effects of the atmosphere on these other systems. [Meteorology](https://en.wikipedia.org/wiki/Meteorology) includes [atmospheric chemistry](https://en.wikipedia.org/wiki/Atmospheric_chemistry) and [atmospheric physics](https://en.wikipedia.org/wiki/Atmospheric_physics) with a major focus on [weather forecasting](https://en.wikipedia.org/wiki/Weather_forecasting). [Climatology](https://en.wikipedia.org/wiki/Climatology) is the study of atmospheric changes (both long and short-term) that define average climates and their change over time, due to both natural and anthropogenic [climate variability](https://en.wikipedia.org/wiki/Climate_change).

Experimental instruments used in atmospheric sciences include satellites, rocketsondes, radiosondes, weather balloons and lasers. With the help of these instruments, relevant and accurate weather data collection is possible.

As a subset of artificial intelligence (AI), machine learning algorithms enable computers to learn from data, and even improve themselves, without being explicitly programmed. Thanks to statistics, machine learning became very famous in 1990s. The intersection of computer science and statistics gave birth to probabilistic approaches in AI. This shifted the field further toward data-driven approaches. Having large-scale data available, scientists started to build intelligent systems that were able to analyze and learn from large amounts of data. Machines are getting more and more intelligent and AI is expanding to more businesses and industries.

Previous weather analysis models used the complicated blend of mathematical instruments which was insufficient in order to get higher classification rate. In contrast, simple analytical models are well-suited the same tasks. As a consequence, researchers focus on the automated analysis of weather and climate data. [14]

**2.2. Existing System**

The related papers for this project are discussed below.

* Mark Holmstrom, Dylan Liu, Christopher Vo, in their paper, “Machine Learning Applied to Weather Forecasting” [1] explore the applications of machine learning techniques to weather forecasting and potentially generate more accurate weather forecasts for large period of time. This paper predicts the maximum temperature and the minimum temperature for seven days, given weather data for the past two days of a particular region. A linear regression model and a variation of functional regression are used in this paper. The two models are compared with one another and the results from both, with a professional weather forecasting service. The linear regression model implemented in this paper is the base algorithm used in this project.

**2.1.1. Limitations of Existing System**

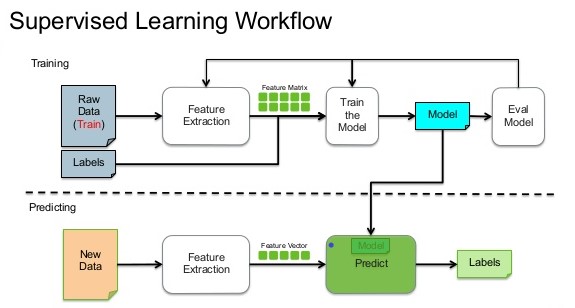
The following limitations were observed after the literature survey:

* No correlation between land temperature and sea surface temperature is being established.
* The linear regression model is inherently a high variance model and thus, would be unstable to outliers. Thus, a larger data set would be required to improve the model.
* The functional regression model is high bias and traditionally requires a large dataset, the weather data of only past two days is insufficient to capture any trends in the weather.

**2.3. Proposed System**

**2.3.1 Overview of Machine Learning**

Data collection allows us to gather records of measurements that are already conducted by data mining techniques. The data mining stage is divided into three phases. At each phase the algorithms are used to analyze the available meteorological datasets. The testing method adopted for this research is percentage split, that train on a percentage of the dataset, cross validate on it and test on the remaining percentage. Thereafter interesting patterns representing knowledge will be identified. With the identified patterns, we apply the supervised machine learning techniques such as linear regression and functional regression to build a Machine Learning model as shown in Fig 2.1. This Machine Learning model will be able to predict climatic changes which affect the temperature of Chennai city in coastal region due to corresponding change in Bay of Bengal temperature.



**Fig2.1.:Supervised Learning Workflow [11]**

**2.3.2. Outline of Machine Learning**

/\* Stuff \*/

**2.4. Advantages of Proposed System**

/\*Stuff \*/

CHAPTER -3

SYSTEM REQUIREMENT

SPECIFICATION

CHAPTER 3

SYSTEM REQUIREMENTS

**3.1. Hardware Requirements**

* **System:** Intel core 7th Gen i5 Processor.
* **Hard Disk:** 500Gb HDD
* **RAM:** 8Gb recommended
* **Monitor:** 15 VGA Color
* **GPU:** Nvidia GT 630M 1Gb VRAM

**3.2. Software Requirements**

* **O/S:** Windows 7 and above
* **Language:** Python
* **Additional Packages:** Python SciKit, Anaconda, TensorFlow, Spyder.

**3.3. Functional Requirements**

**3.4. Non-Functional Requirements**

CHAPTER – 4

**COST ESTIMATION**

**OF THE PROJECT**

**CHAPTER - 5**

**SYSTEM DESIGN AND**

**DEVELOPMENT**

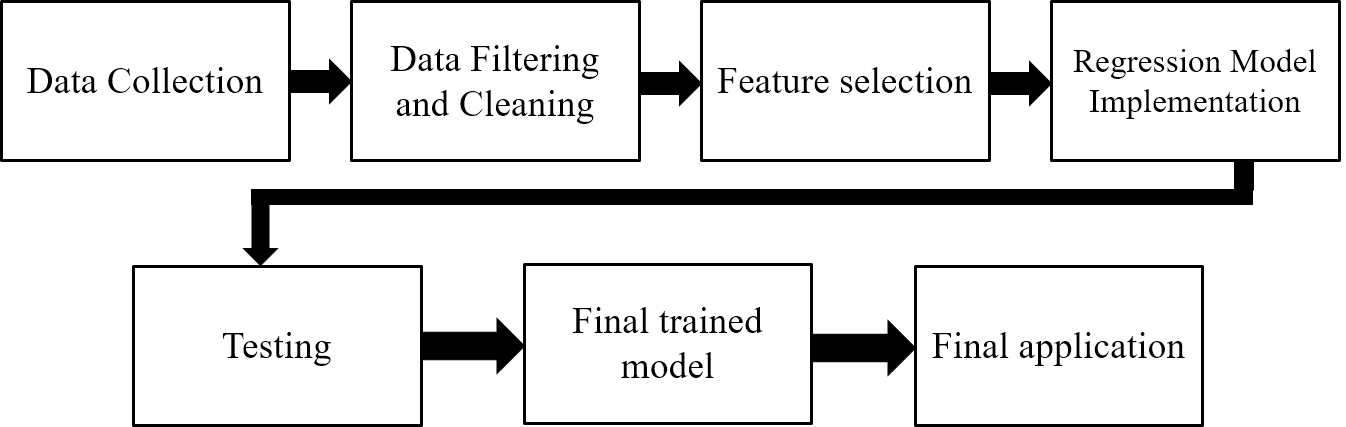
**CHAPTER 5**

**SYSTEM DESIGN AND DEVELOPMENT**

**5.1. System Design**

The proposed system design consists of seven connected components they are data collection, data filtering and cleaning, feature selection, Regression model implementation, testing, final trained model, final application. Each component is dependent on each other to perform the task.

In the data collection component, the dataset is formed and the filtering of the dataset is done to this collected data. The irrelevant features are removed based on the availability of the data for any particular attribute. The process of selecting the features is a bit complicated. Building of a regression model will be based on the relevant features. The obtained dataset will be divided into test dataset as well as training dataset, the latter will be used in the final trained model to check the prediction. Comparative study between the sea surface temperature and the land surface temperature will be carried on using the obtained model.

****

**Fig 4.1 System design**

**5.2. DFD/UML Diagrams**

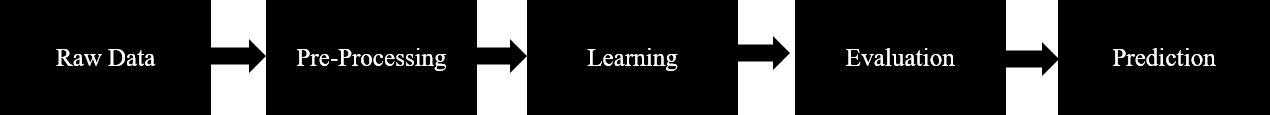
**5.2.1. Data flow Diagram**

A Data Flow Diagram (DFD) is traditional graphical representation of the “flow” data through an information system. A DFD is often used as a preliminary step to create an overview of the system. DFD’s can also be used for visualization of data processing. It shows how information enters and leaves the system, what changes the information and where information is stored.

Data flow diagrams can be used to provide the end used with a physical idea of where the data they input ultimately has an effect upon the structure of the whole system from order to dispatch report. The development of the system can be easily determined through the data flow diagram model.

A data flow shows the flow of information from its source to its destination. A data flow is represented by a line, with arrowheads showing the direction of flow. Information always flows to or from a process. It is usually beginning with a context diagram as the level 0 of DFD diagram, a simple representation of the whole system.

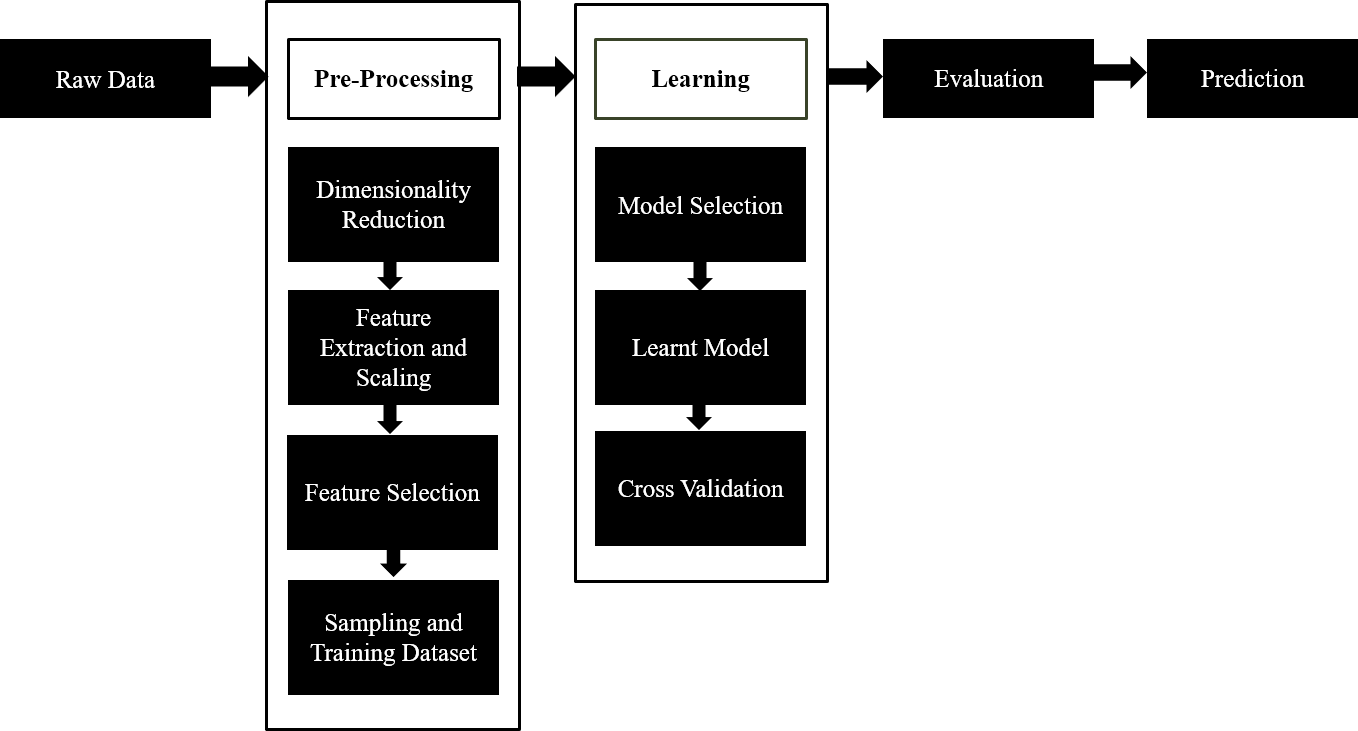
**Level 0**



**Fig 5.2.1 Level 0 DFD Diagram**

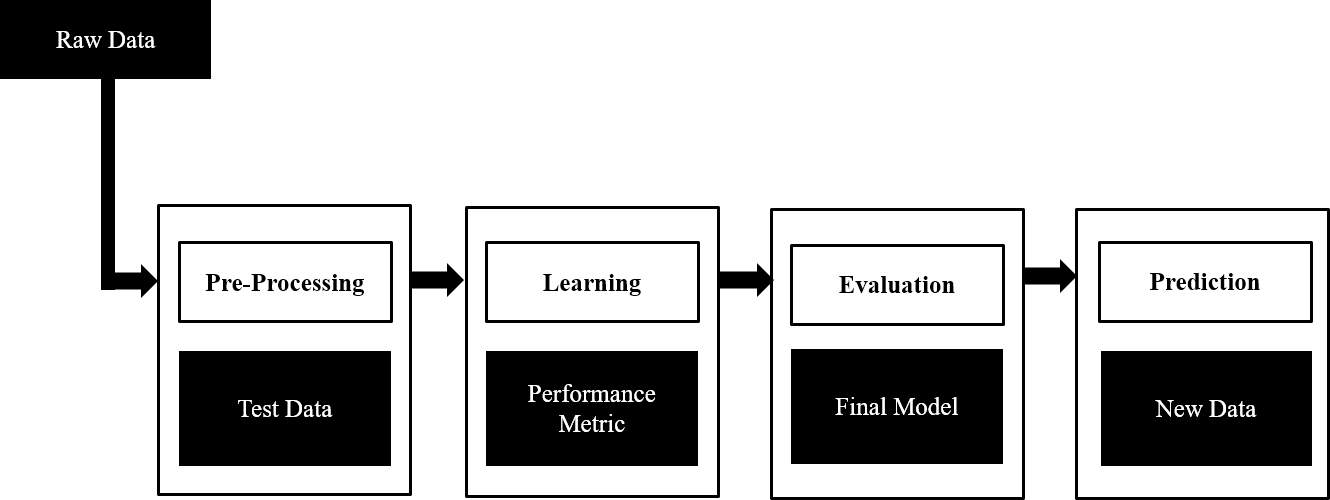
To elaborate further from that level, we drill down to a level 1 diagram with lower level functions decomposed from the major functions of the system.

**Level 1**



This could continue to evolve to become a level 2 diagram when further analysis is required. Progression to level 3, 4 and so on is possible but anything beyond level 3 is not very common.

**Level 2**

**Fig 5.2 Data Flow Diagram**

**CHAPTER - 6**

**SCHEDULE OF**

**THE PROJECT**

**CHAPTER 6**

**SCHEDULE OF THE PROJECT**

Based on the system design, the project implementation is divided into the following modules:

Module-1 : Data Collection

Module-2 : Data Cleaning and Filtering

Module-3 : Feature Selection

Module-4 : Implementing Regression model

Module-5 : Testing

The timeline followed for this project is given by Fig 6.1.



**Fig 6.1: Timeline of the project**

CHAPTER – 7

**IMPLEMENTATION**

**CHAPTER 7**

**IMPLEMENTAION**

\

**7.1. Introduction**

Implementation is the phase of the system in which the application is written in a programming language based on the design developed in the earlier phase. The implementation should be straight forward as all the decisions about the design will be done before the actual coding begins. We have used python, Spyder and anaconda.

**7.2. Technologies and tools used:**

**7.2.1. Python**

Python is an object-oriented, high-level programming language with integrated dynamic semantics. It is extremely attractive in the field of data science because it offers many built in libraries and functions that help in the data science operations.  It is easy to read and translate Python code much easier than other languages. In turn, this reduces the cost of program maintenance and development because it allows teams to work collaboratively without significant language and experience barriers. Additionally, Python supports the use of modules and packages, which means that programs can be designed in a modular style and code can be reused across a variety of projects. Once you've developed a module or package you need, it can be scaled for use in other projects, and it's easy to import or export these modules. One of the most promising benefits of Python is that both the standard library and the interpreter are available free of charge, in both binary and source form.

**7.2.2. Spyder**

Spyder is a powerful interactive development environment for the Python language with advanced editing, interactive testing, debugging and introspection features. Additionally, Spyder is a numerical computing environment thanks to the support of IPython and popular Python libraries such as NumPy, SciPy, or matplotlib. To install this with conda we type the command on cmd or terminal: conda install -c anaconda spyder

**7.2.3. Anaconda**

Anaconda is a Python distribution that makes it easy to install Python plus a number of its most often used 3rd party libraries in a flexible way on a Windows or Linux machine. It is a free and open source distribution of the Python and R programming languages for data science and machine learning related applications (large-scale data processing, predictive analytics, scientific computing), that aims to simplify package management and deployment. Package versions are managed by the package management system conda, which makes it quite simple to install, run, and update complex data science and machine learning software libraries like Scikit-learn, TensorFlow, and SciPy.

**7.3. Dataset description**

[Weather Underground](https://www.wunderground.com/) is a company that collects and distributes data on various weather measurements around the globe. The company provides a swath of API's that are available for both commercial and non-commercial uses. Weather Underground provides many different web service API's to access data from but, the one we will be concerned with is their history API. The history API provides a summary of various weather measurements for a city and state on a specific day.

Land weather conditions dataset required for this project is collected from Weather Underground. Each record contains mean temperature, mean dew point, mean pressure, maximum humidity, minimum humidity, maximum temperature, minimum temperature, maximum dew point, minimum dew point, maximum pressure, minimum pressure, precipitation details. Land weather conditions dataset containing all these features is collected for a period of 2015 to 2017 for Chennai city.

**7.4. List of Modules**

**7.4.1. Data Collection**

Data collection is the process of gathering and measuring information on targeted variables in an established systematic fashion, which then enables one to answer relevant questions and evaluate outcomes. Data collection is a component of research in all fields of study including [physical](https://en.wikipedia.org/wiki/Physical_science) and [social sciences,](https://en.wikipedia.org/wiki/Social_science) [humanities,](https://en.wikipedia.org/wiki/Humanities) and [business.](https://en.wikipedia.org/wiki/Business) While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same. The goal for all data collection is to capture quality evidence that allows analysis to lead to the formulation of convincing and credible answers to the questions that have been posed.

The dataset that is used for this work is obtained from the following sources:

* Land temperature [12]
* Sea surface temperature [13]

The data used for this project is collected from Weather Underground's free tier API web service. Weather Underground provides many different web service API's to access data from but, the one we will be concerned with is their history API. The history API provides a summary of various weather measurements for a city and state on a specific day.

Data collection process takes the following steps:

Step 1: Initialize base url, api key, target date, feature name list.

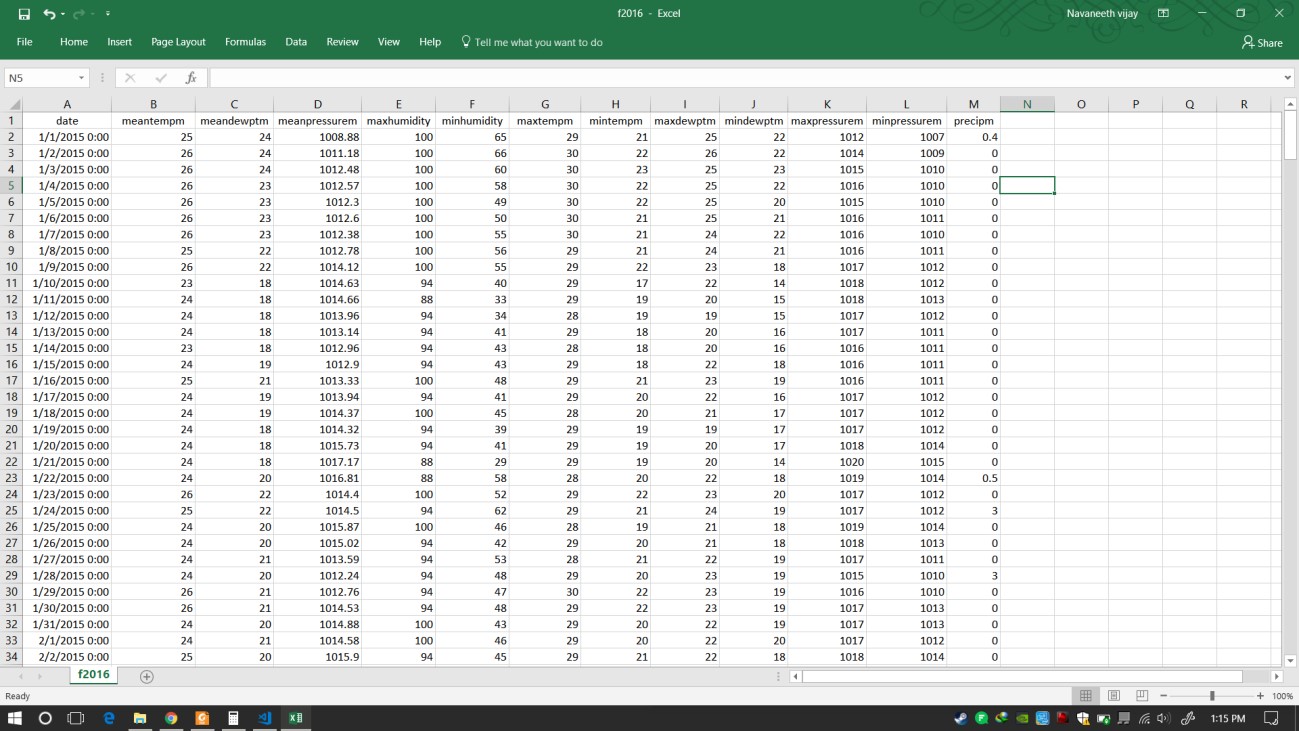
Step 2: open a csv file in write mode set the fieldnames to feature list values

Step 3:for each in range(days) send a request to the base url defined convert the api response to json format write the json data to the opened csv file increment the target date to next day

end for

Step 4: end

Thus, weather details of Chennai such as, mean temperature, mean dew point, mean pressure, maximum humidity, minimum humidity, maximum temperature, minimum temperature, maximum dew point, minimum dew point, maximum pressure, minimum pressure and precipitation is collected. A sample output of the above code is shown in Fig 4.3.



**Fig 7.1: Snapshot of Data collected for the project**

**7.4.2. Data Cleaning and Filtering**

Data filtering in IT can refer to a wide range of strategies or solutions for refining data sets. This means the data sets are refined into simply what a user (or set of users) needs, without including other data that can be repetitive, irrelevant or even sensitive. Different types of data filters can be used to amend reports, query results, or other kinds of information results.

Data cleaning, (or data cleansing, data scrubbing) is an aspect of [data processing](http://psychology.wikia.com/wiki/Data_processing) and is the process of detecting and correcting (or removing) corrupt or inaccurate [records](http://psychology.wikia.com/wiki/Storage_record?redlink=1&action=edit&flow=create-page-article-redlink) from a record set, [table,](http://psychology.wikia.com/wiki/Table_(database)) or [database.](http://psychology.wikia.com/wiki/Database) Used mainly in databases, the term refers to identifying incomplete, incorrect, inaccurate, irrelevant, etc. parts of the data and then replacing, modifying, or deleting this [dirty data](http://psychology.wikia.com/wiki/Dirty_data?redlink=1&action=edit&flow=create-page-article-redlink)

The steps used for the process of data filtering and cleaning are as follows:

Step 1: Make a list containing features to remove

Step 2: Make a list containing features to keep

Step 3: Convert the selected feature objects into numerical values

Set not available values to 'coerce’.

Step 4: Transpose the DataFrame Set.

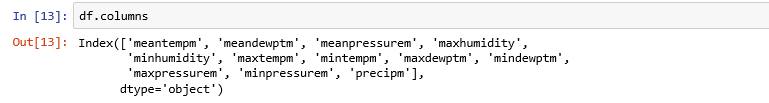
Step 5: Calculate IQR and remove outliers

Step 6: Data visualization.

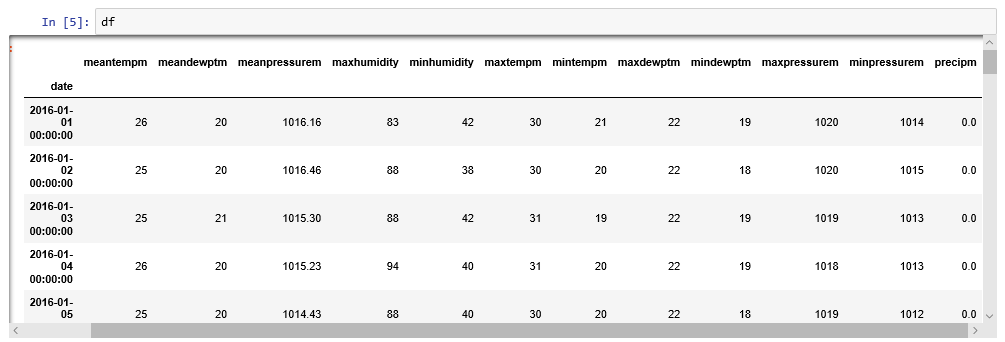
Step 7: End

Some snapshots to follow along with the algorithm are given below:

The Pandas DataFrame is a very useful data structure for many programming tasks which are most popularly known for cleaning and processing data to be used in machine learning projects (or experiments).Fig 7.2 and 7.3 show the initializing of the data frame:

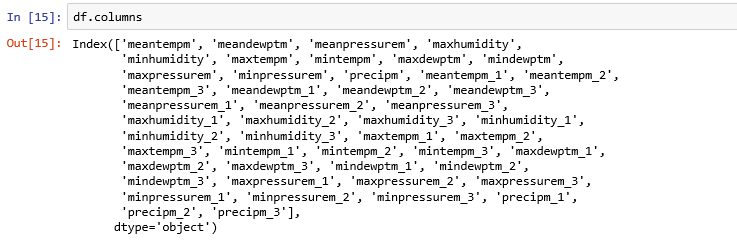


**Fig 7.2 : Initializing the data frames**

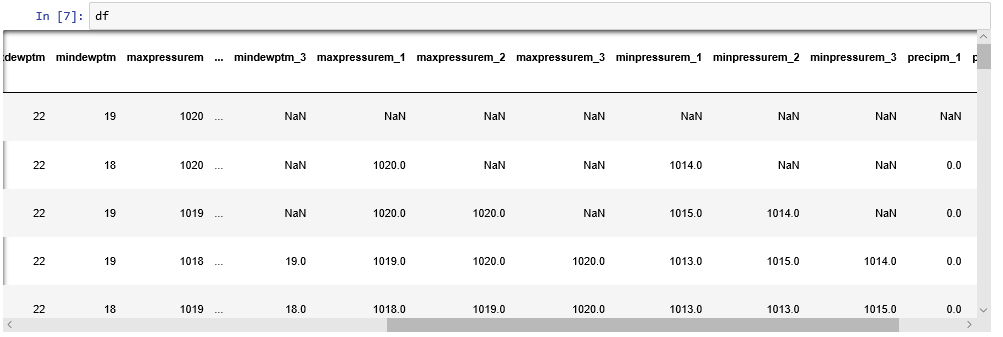


**Fig 7.3 : Initializing the data frames**

Selecting meaningful features to investigate, paired with a thoughtful assumption of likely patterns in data. Fig 7.4 and 7.5 show deriving features:

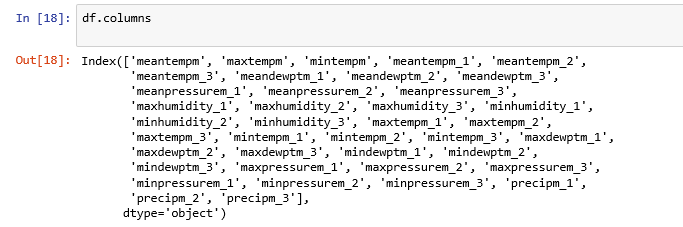


**Fig 7.4 : Deriving features**



**Fig 7.5: Deriving features**

For each day (row) and for a given feature (column) find the value for that feature N days prior. For each value of N (1-3 in our case) make a new column for that feature representing the Nth prior day's measurement. Fig 7.6 shows retaining features:



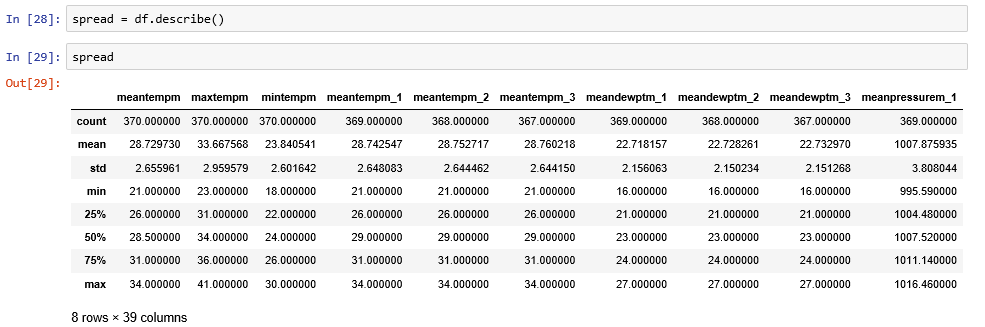
**Fig 7.6 : Retaining features**

The data type of every column is of type "object". Thus, there is a need to convert all of these feature columns to floats for the type of numerical analysis must performed later. Object to numeric conversion is shown in Fig 7.7.



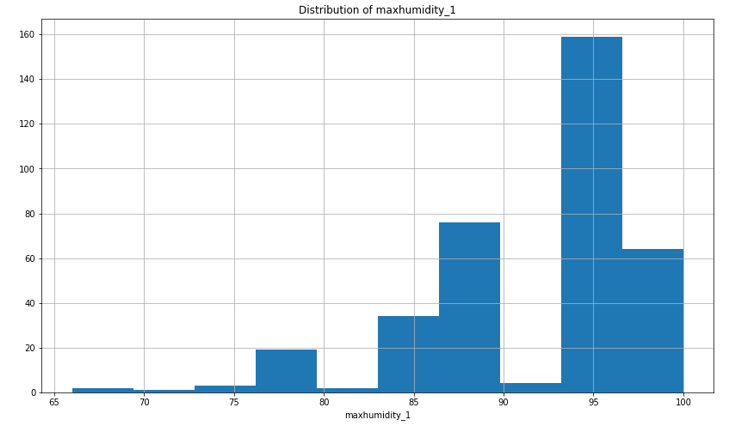
**Fig 7.7: Object to numeric**

The DataFrame method describe() will produce a DataFrame containing the count, mean, standard deviation, min, 25th percentile, 50th percentile (or median), the 75th percentile and, the max value. This can be very useful information to evaluating the distribution of the feature data. Describe function is show in Fig 7.8.



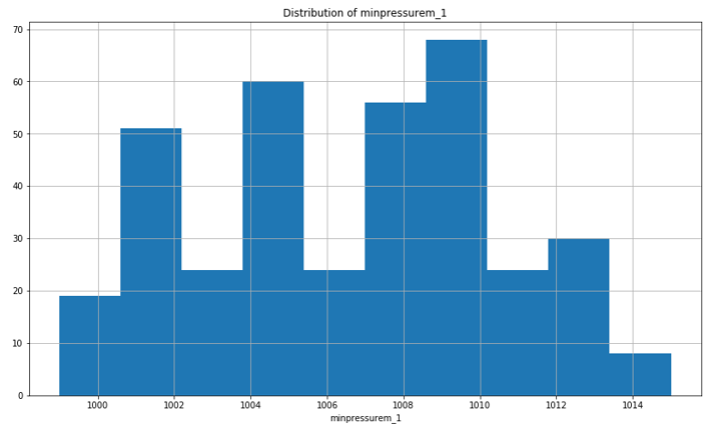
**Fig 7.8: Describe function**

The first set of features all appear to be related to max humidity. Looking at the data one can tell that the outlier for this feature category is due to the apparently very low min value. To take a closer look at it, use a histogram. Looking at the histogram of the values for maxhumidity, Fig 7.9,  the data exhibits quite a bit of negative skew. This must be kept in mind when selecting prediction models and evaluating the strength of impact of max humidities.

****

**Fig 7.9 : histogram of values of maxhumidity**

Now consider the minimum pressure feature distribution shown in Fig 7.10. This plot exhibits another interesting feature. From this plot, the data is [multimodal](https://en.wikipedia.org/wiki/Multimodal_distribution), from which it can be considered that there are two very different sets of environmental circumstances apparent in this data.

****

**Fig 7.10 : Mnimum pressure feature distribution**

**7.4.3. Feature Selection**

In [machine learning](https://en.wikipedia.org/wiki/Machine_learning) and [statistics](https://en.wikipedia.org/wiki/Statistics), feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant [features](https://en.wikipedia.org/wiki/Feature_(machine_learning)) (variables, predictors) for use in model construction. Feature selection techniques are used for four reasons:

* simplification of models to make them easier to interpret by researchers/users,
* shorter training times,
* to avoid the [curse of dimensionality](https://en.wikipedia.org/wiki/Curse_of_dimensionality),
* enhanced generalization by reducing [overfitting](https://en.wikipedia.org/wiki/Overfitting) (formally, reduction of [variance](https://en.wikipedia.org/wiki/Bias-variance_tradeoff))

**7.4.4. Implementing Regression Model**

Linear regression is a [linear](https://en.wikipedia.org/wiki/Linear) approach for modelling the relationship between a scalar [dependent variable](https://en.wikipedia.org/wiki/Dependent_variable) *y* and one or more [explanatory variables](https://en.wikipedia.org/wiki/Explanatory_variable) (or independent variables) denoted *X*.

Linear regression has many practical uses. Most applications fall into one of the following two broad categories:

* If the goal is prediction, or forecasting, or error reduction, linear regression can be used to fit a predictive model to an observed data set of *y* and *X* values. After developing such a model, if an additional value of *X* is then given without its accompanying value of *y*, the fitted model can be used to make a prediction of the value of *y*.

Given a variable *y* and a number of variables *X*1, ..., *Xp* that may be related to *y*, linear regression analysis can be applied to quantify the strength of the relationship between *y* and the *Xj*, to assess which *Xj* may have no relationship with *y* at all, and to identify which subsets of the *Xj* contain redundant information about *y*.

**7.4.5. Testing**

Software testing is a method of assessing the functionality of a [software](http://searchsoa.techtarget.com/definition/software) [program](http://searchsoftwarequality.techtarget.com/definition/program). There are many different types of software testing but the two main categories are [dynamic testing](http://whatis.techtarget.com/definition/dynamic-testing) and [static testing](http://whatis.techtarget.com/definition/static-testing).

Dynamic testing is an assessment that is conducted while the program is [executed](http://searchcio-midmarket.techtarget.com/definition/executable); static testing, on the other hand, is an examination of the program's [code](http://whatis.techtarget.com/definition/code) and associated [documentation](http://searchsoftwarequality.techtarget.com/definition/documentation). Dynamic and static methods are often used together.

**7.5. Description of Modules**

**7.6. Main Algorithm**

**CHAPTER – 8**

**TESTING AND**

**VALIDATION**

**CHAPTER 8**

**TESTING AND VALIDATION**

**8.1. Testing Process**

**8.2. Test cases and Validation**

**CHAPTER – 9**

**RESULTS AND**

**SNAPSHOTS**

**CHAPTER 9**

**RESULTS AND SNAPSHOTS**

**CHAPTER -10**

**CONCLUSION AND**

**FUTURE ENHANCEMENTS**

**CHAPTER 10**

**CONCLUSION AND FUTURE ENCHANCEMENTS**

**10.1. Conclusion**

**10.2. Future Directions**

**REFERENCES**

[1] Mark Holmstrom, Dylan Liu, Christopher Vo. Machine Learning Applied to Weather Forecasting. Stanford University. 2016.

[2] Siddharth S. Bhatkhande, Roopa G. Hubballi. Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques. *IJARCCE* (2016).

[3] Aditya Grover, Ashish Kapoor and Eric Horvitz. A Deep Hybrid Model for Weather Forecasting.

[4] John K. Williams and D. A. Ahijevych, C. J. Kessinger, T. R. Saxen, M. Steiner and S. Dettling. A machine learning approach to finding weather regimes and skillful predictor combinations for short-term storm forecasting. National Center for Atmospheric Research, Boulder, Colorado.

[5] Aastha Sharma, Setu Chaturvedi and Bhupesh Gour. A Semi- Supervised Technique for Weather Condition Prediction using DBSCAN and KNN.International Journal of Computer Applications (0975 – 8887) Volume 95– No. 10, June 2014.

[6] Mrs. C. Beulah Christalin Latha, Dr. (Mrs.) Sujni Paul, Dr.E.Kirubakaran and Mr. Sathianarayanan. A Service Oriented Architecture for Weather Forecasting Using Data Mining.Int. J. of Advanced Networking and Applications. Volume: 02, Issue:02, Pages:608-613. 2010.

[7] C. Johansson, M. Bergkvist, O. De Somer, D. Geysen , N. Lavesson and D. Vanhoudt. Operational demand forecasting in district heating systems using ensembles of online machine learning algorithms.The 15th International Symposium on District Heating and Cooling September 4-7, 2016, Seoul, Republic of Korea (South Korea).

[8] Daniel Bejarano and Adriano Quiroga. Wind Prediction: Physical model improvement through support vector regression.Stanford University. December 2013.

[9] Emilcy Hern ́andez Victor Sanchez-Anguix Vicente Julian Javier Palanca and N ́estor Duque.Rainfall prediction: A Deep Learning approach.

[10] Kiran Kumar. R 1, Usha Rani. R.Weather Prediction through Machine Learning.Dept of Computer science, Krishna University, Machilipatnam. AP, India.

[11] [Morgun Ivan](http://en.proft.me/) “Types of machine learning algorithms”.

[12] “Chennai, IN” in Weather Underground, The Weather Company, 2016. [Online]. Available: <https://www.wunderground.com/weather/in/chennai>

[13] NOAA's National Centers for Environmental Information (NCEI), 2015. [Online].

Available:<https://www.ncdc.noaa.gov/cdo-web/datatools/marine>

[14]  Nasimul Hasan ,  Md. Taufeeq Uddin  ,[Nihad Karim Chowdhury](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Nihad%20Karim%20Chowdhury.QT.&newsearch=true) “Automated weather event analysis with machine learning”