**VISVESVARAYA TECHNOLOGICAL UNIVERSITY**

Jnana Sangama, Machhe, Belagavi, Karnataka 590018



A

Project Synopsis

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 synopsis of project work entitled correlation of coastal climate change associated with patterns detected in the sea surface temperature is carried out by Mr./Ms. Bhavana , H Sudhanva , Sai Navaneeth V , Satish Kumar M S USN 1BG14IS008, 1BG14IS017 , 1BG14IS041 , 1BG14IS045 respectively, 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 synopsis has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said Degree.

**Dr. Saritha Chakrasali**

**Professor** , Dept. of ISE

BNMIT

**ABSTRACT OF THE PROJECT**

The increasing research in the fields of Artificial Intelligence and Machine Learning has given rise to numerous weather prediction models. But 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 cities in eastern coast of India such as Chennai, Visakhapatnam etc., along with the corresponding sea water 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 water 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 | 103 |
| **2** | **LITERATURE SURVEY**   * 1. Existing system      1. Limitations of existing system   2. Proposed system      1. Advantages of proposed system | 214 |
| **3** | **SYSTEM REQUIREMENTS SPECIFICATION**   * 1. Hardware requirements   2. Software requirements | 120 |
| **4** | **SYSTEM DESIGN AND DEVELOPMENT**   * 1. System design   2. UML diagrams      1. Dataflow diagram | 121 |
| 5 | IMPLEMENTATION   * 1. Introduction   2. Technologies and tools used      1. Python      2. Spyder      3. Anaconda   3. Dataset description   4. List of modules      1. Data collection      2. Data cleaning and filtering      3. Feature extraction      4. Implementing regression model      5. Final model   5. Description of modules   6. Main algorithm | 210 |

CHAPTER – 1

**INTRODUCTION**

**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**

**LITERATURE SURVEY**

**2.1. 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 using linear regression model. The linear regression model implemented in this paper is the base algorithm used in this project.
* Siddharth S. Bhatkhande, Roopa G. Hubballi in their paper “Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques” [2] investigate the use of data mining techniques in forecasting attributes like maximum temperature, minimum temperature, which is carried out using Decision Tree algorithms and meteorological data collected between 2012 and 2015 from the different cities. On available datasets, they apply the Decision Tree Algorithm for deleting the inappropriate data. The concepts of data mining implemented here will be used as reference in this project.
* Kiran Kumar. R and Usha Rani. R in their paper, “Weather Prediction through Machine Learning” [10], design an effective rainfall prediction agent model using support vector machine and multiple linear regressions.

**2.1.1 Limitations of existing system**

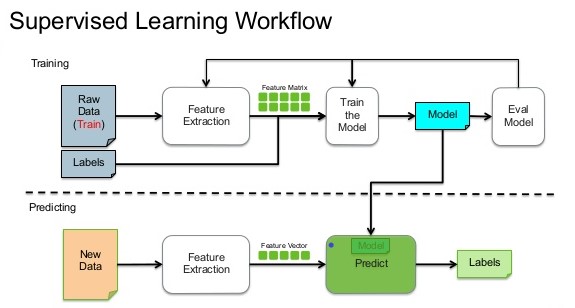
The following limitations were observed after the literature survey:

* No correlation between climatic condition of the land and sea surface temperature is being established.
* The linear regression model is inherently a high variance model and thus, would be unstable to outliers.
* 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.

**2.2 Proposed system**

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 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.



**/\*------\*/**

**2.2.1 Advantages of existing system**

CHAPTER – 3

**SYSTEM REQUIREMENTS**

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.

CHAPTER – 4

**SYSTEM DESIGN AND**

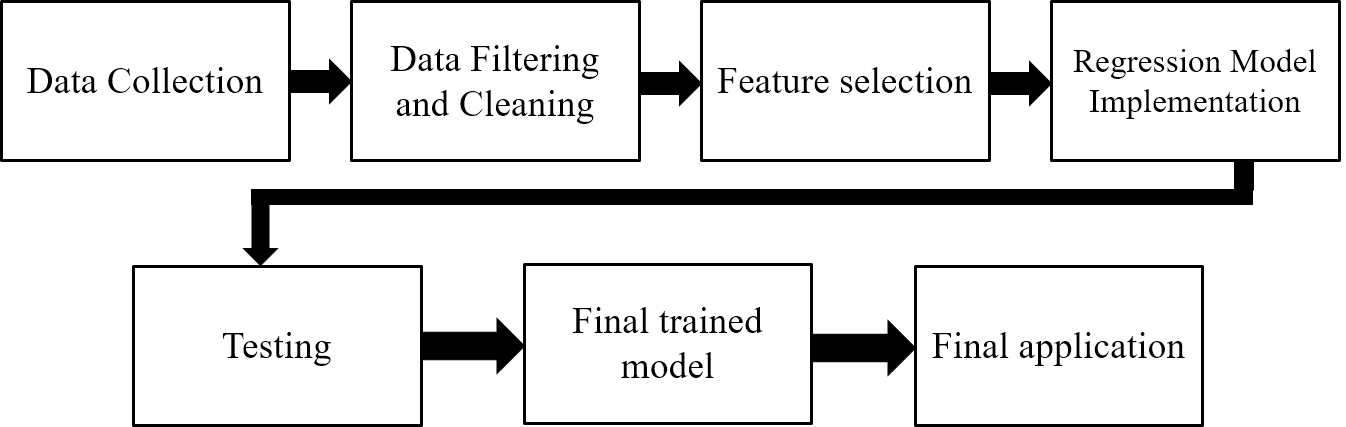
**DEVELOPMENT**

**SYSTEM DESIGN AND DEVELOPMENT**

**4.1 System design**

Our 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**

**4.2 UML diagrams**

**4.2.1 Data flow diagram**

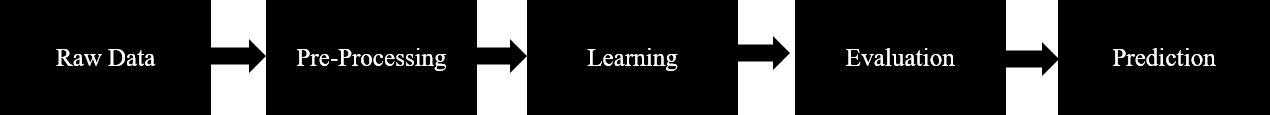
A Data Flow Diagram (DFD) is traditional graphical representation of the “flow” data through an information system. A DFD is often used as a preliminary step to create an overview of the system. DFD’s can also be used for visualization of data processing. It shows how information enters and leaves the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

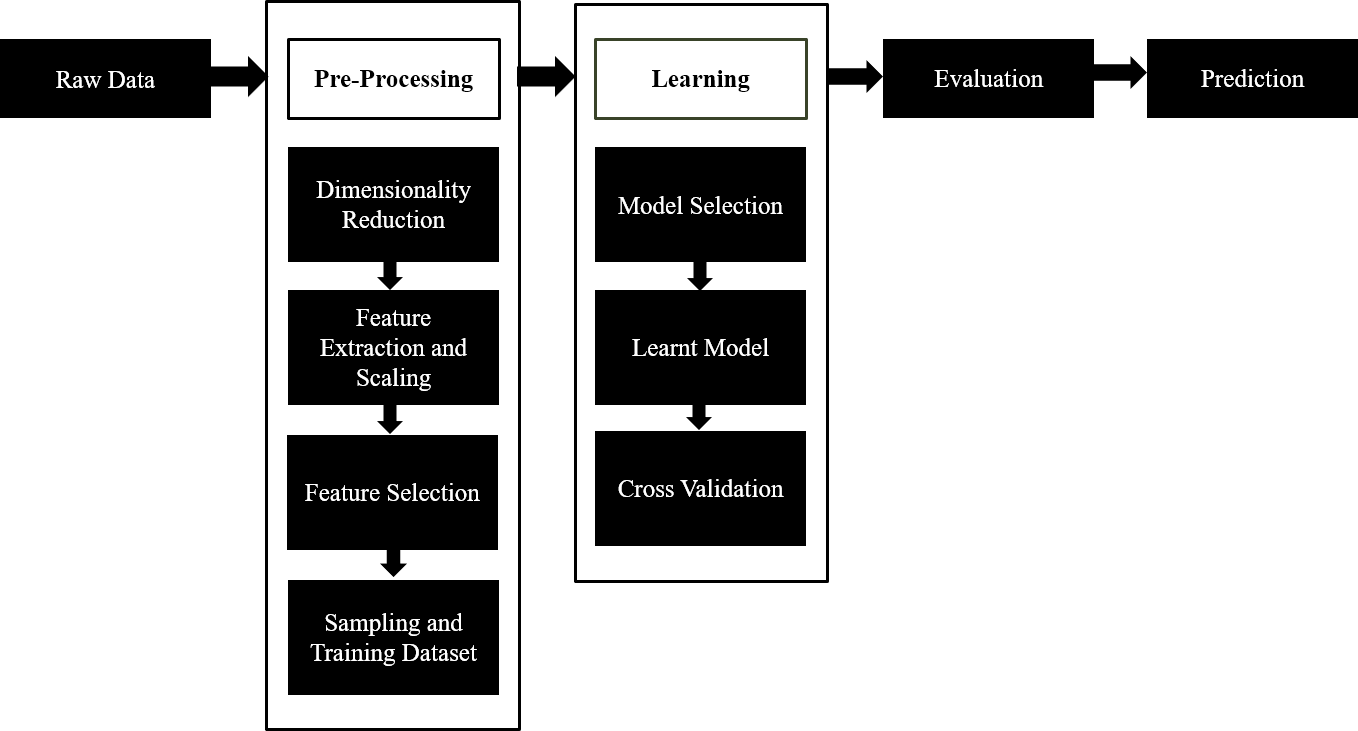
A data flow shows the flow of information from its source to its destination. A data flow is represented by a line, with arrowheads showing the direction of flow. Information always flows to or from a process. It is usually beginning with a context diagram as the level 0 of DFD diagram, a simple representation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

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. 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.

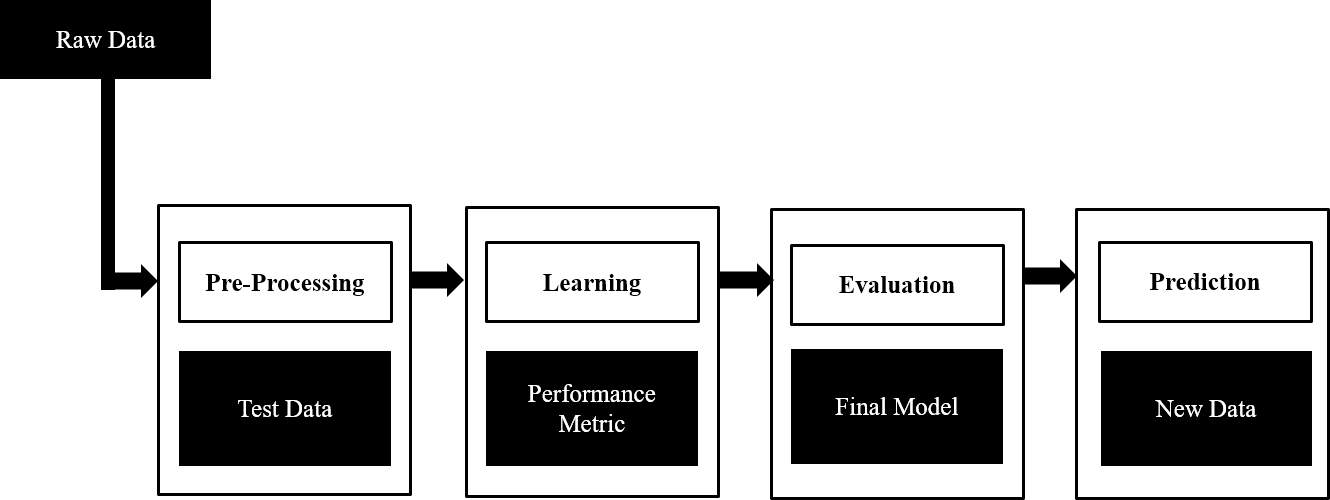
**Level 0**



**Level 1**



**Level 2**

****

**Fig 4.2 Data Flow Diagram**

CHAPTER – 5

**IMPLEMENTATION**

**IMPLEMENTAION**

\

**5.1 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.

**5.2 Technologies and tools used:**

**5.2.1 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.

**5.2.2 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

**5.2.3 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.

**5.3 Dataset description**

**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] <http://en.proft.me/2015/12/24/types-machine-learning-algorithms/>

[12] <https://www.wunderground.com/>

[13] <https://seatemperature.info/>