**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**

(Approved by AICTE, Affiliated to VTU, ISO 9001:2008 Certified and Accredited as grade A Institution by NAAC)

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

Ph: 91-80-26711780/81/82 Email: principal@bnmit.in www.bnmit.org

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 project work entitled correlation of coastal climate change associated with patterns detected in the sea surface temperature is carried out by Mr./Ms. Bhavana (IBG14IS008) , H Sudhanva (1BG14IS017), Sai Navaneeth V (1BG14IS041), 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 synopsis 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.**

**CONTENTS**

1. Abstract of the project 1
2. Introduction 2
   1. Literature Survey 3
   2. Motivation 5
   3. Problem statement 5
   4. Goals & Objectives 5
   5. Proposed System 6
   6. Timeline 7
   7. Overall flow of the Project 7

2.7.1 Module 1- Data Collection 7

2.7.2 Module 2- Data Filtering and Cleaning 9

2.7.3 Module 3- Feature Selection 10

2.7.4 Module 4- Implementing Linear Regression Model 11

2.7.5 Module 5- Testing 12

* 1. System requirements 13
  2. References 14

Chapter 1

**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 aim 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.

Chapter 2

**INTRODUCTION**

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.

**2.1. LITERATURE SURVEY**

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.

Further, the other related papers are discussed below:

* C. Johansson, M. Bergkvist, O. De Somer, D. Geysen, N. Lavesson and D. Vanhoudt in their paper, “Operational demand forecasting in district heating systems using ensembles of online machine learning algorithms” [7], state that Heat demand forecasting is in one form or another an integrated part of most optimization solutions for district heating and cooling (DHC). This paper presents the current status and results from extensive work in the development, implementation and operational service of online machine learning algorithms for demand forecasting.
* Aastha Sharma, Setu Chaturvedi and Bhupesh Gour in their paper, “A Semi- Supervised Technique for Weather Condition Prediction using DBSCAN and KNN” [5], propose a semi-supervised weather prediction technique to validate the predictions done for certain atmospheric parameters taken for four years on a day wise basis in a certain city.
* Aditya Grover, Ashish Kapoor and Eric Horvitz in their paper, “A Deep Hybrid Model for Weather Forecasting” [3], state that weather forecasting is a canonical predictive challenge that has depended primarily on model-based methods. They study specifically the power of making predictions via a hybrid approach that combines discriminatively trained predictive models with a deep neural network that models the joint statistics of a set of weather-related variables.
* John K. Williams and D. A. Ahijevych, C. J. Kessinger, T. R. Saxen, M. Steiner and S. Dettling in their paper, “A machine learning approach to finding weather regimes and skillful predictor combinations for short-term storm forecasting” [4], it is shown that the random forest machine learning approach provides a tool for identifying a set of skillful predictors for thunderstorm initiation as well as providing a performance benchmark.
* Mrs. C. Beulah Christalin Latha, Dr. (Mrs.) Sujni Paul, Dr.E.Kirubakaran and Mr. Sathyanarayanan in their paper, “A Service Oriented Architecture for Weather Forecasting Using Data Mining” [6], treat weather as a continuous, data-intensive, multidimensional,dynamic process, that makes weather forecasting a formidable challenge.Their paper proposes a novel method to develop a service oriented architecture for a weather information system and forecast weather using data mining techniques.
* Daniel Bejarano and Adriano Quiroga in their project, “Wind Prediction: Physical model improvement through support vector regression” [8], concentrates on wind speed prediction through the combination of support vector regression and the weather research and forecast model was explored.
* Emilcy Hern´andez, Victor Sanchez-Anguix, Vicente Julian, Javier Palanca, and N´estor Duque in their paper, “Rainfall prediction: A Deep Learning approach” [9], they introduce an architecture based on Deep Learning for the prediction of the accumulated daily precipitation for the next day. More specifically, it includes an autoencoder for reducing and capturing non-linear relationships between attributes, and a multilayer perceptron for the prediction task.
* 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.

**2.2 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.

**2.3 PROBLEM STATEMENT**

The project aims to achieve the following objectives:

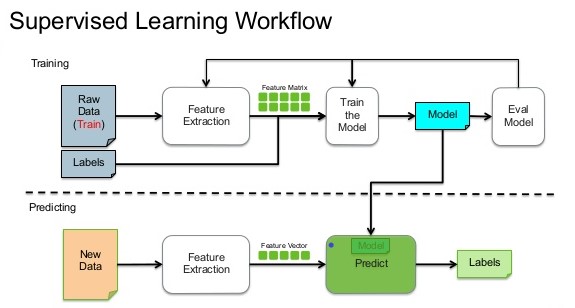
* To analyze related weather data using data mining techniques.
* To obtain correlation between sea surface temperature to corresponding coastal temperature.
* Implementing supervised machine learning algorithm on the available dataset.
* To forecast the weather conditions of the respective city.

**2.4 GOALS & OBJECTIVES**

* To be able to recognize pattern between sea and land temperature.
* To consider a larger dataset.
* Improving weather forecasting process.
* Being able to provide warnings in time to save people as well as property.
* Damage control made easy.

**2.5 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.

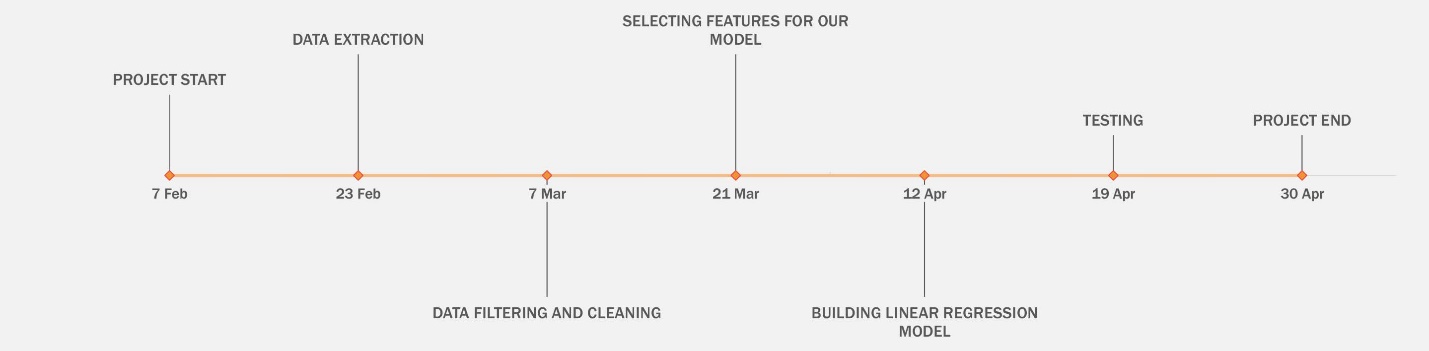


**Fig 1: Supervised Learning Workflow [11]**

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

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

**2.6 TIMELINE**

****

**2.7 OVER ALL FLOW OF THE PROJECT**

Module-1 : Data Collection  
Module-2 : Data Cleaning and Filtering   
Module-3 : Feature Extraction  
Module-4 : Implementing Regression model  
Module-5 : Final model

**2.7.1 MODULE 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. [14]

For the purpose of this project, the following code has been implemented for the data collection process:

Wundeground.py:

import requests

from pyquery import PyQuery as pq

import csv

i = 0

idx = 0

fname = [

"2010.csv",

"2011.csv",

"2012.csv",

"2013.csv",

"2014.csv",

"2015.csv",

"2016.csv",

]

BaseUrl = "https://www.wunderground.com/history/airport/VOMM"

StartYear = [2010,2011,2012,2013,2014,2015,2016]

EndYear = [2010,2011,2012,2013,2014,2015,2016]

StartDayMonth = "/1/1/"

EndDate = "CustomHistory.html?dayend=31&monthend=12"

YearEnd = "&yearend="

UrlEnd = "&req\_city=&req\_state=&req\_statename=&reqdb.zip=&reqdb.magic=&reqdb.wmo="

for value in StartYear:

url = BaseUrl+str(StartYear[idx])+StartDayMonth+EndDate+YearEnd+str(EndYear[idx])+UrlEnd

response = requests.get(url)

fh = open(fname[idx],"w")

doc = pq(response.content)

tablehead = [th.text() for th in doc('#observations\_details td:not(:nth-child(21))').items()]

for item in tablehead:

i = i+1

fh.write("%s," % item)

if(i%20 == 0):

fh.write('\n')

idx = idx + 1

print("Done bro")

fh.close()

**2.7.2 MODULE 2- DATA FILTERING AND CLEANING**

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. [15]

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). [16]

**2.7.3 MODULE 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)) [17]

**2.7.4 MODULE 4- IMPLEMENTING LINER 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*. [18]

**2.7.5 MODULE 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. [19]

**2.8 SYSTEM REQUIREMENTS**

**Hardware requirements:**

* **System:** Intel core 7th Gen i7 Processor.
* **Hard Disk:** 150Gb Solid State Drive
* **RAM:** 8Gb recommended
* **Monitor:** 15 VGA Colour
* **GPU:** Nvidia GT 630M 1Gb VRAM

**Software requirements:**

* **O/S:** Windows 7 and above
* **Language:** python, R and MATLAB
* **Additional Packages:** python scikit, python anaconda, tensorflow

**2.9 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/>

[14] <https://en.wikipedia.org/wiki/Data_collection>

[15] <https://www.techopedia.com/definition/26202/data-filtering>

[16] <http://psychology.wikia.com/wiki/Data_cleaning>

[17] <https://en.wikipedia.org/wiki/Feature_selection>

[18] <https://en.wikipedia.org/wiki/Linear_regression>

[19] <http://whatis.techtarget.com/definition/software-testing>