**CHAPTER 1**

**INTRODUCTION**

# 1.1 Introduction

Agriculture can be seen as the oldest occupation in the world and it has been evolving ever since along with the human race. Nepal is also an agricultural country and 66% of people are directly involved in farming [1]. Till this date, there’re so many research and developments in this field. There are almost all kinds of technologies which makes work easier in this field. Despite these developments and advancements, agriculture is facing so many challenges. As of now greatest challenges are climate change, degrading soil condition, and production should be increased making lesser investment and time.

Weather Analysis for Crop Prediction is a system combining of machine learning and web field to provide reliable and accurate measure for predicting crops which is suitable under the soil condition and weather scenario. The users will be suggested with the best crop suited by the weather of the place and soil attributes. The system uses open street map API in order to get weather attributes of different places within Nepal. These obtained weather attributes along with the soil attributes are fed to machine learning algorithm in order to compute the most suitable crop for the selected region.

This project classifies the weather conditions and maps the crops suitable under those classified weather condition. Then, soil attributes are taken as input and prediction model which is trained to predict suitable crops for that soil type and weather predicts the crops for that data. This enables the farmers to minimize the loss of crops due to adverse weather conditions and geographical factors too. This application focuses on analyzing the past trends of weather of a particular place to predict the future weather and assist farmers in choosing the right crop with the maximum productivity.

# 1.2 Problem Definition

Farmers now a days are having difficult times in choosing the right crop with maximum productivity. They rely on traditional calendar method to choose the crops for different months throughout the year. Relying on calendar was decent method for selecting the crop till some decade back. This however, is proving to lose the productivity year after year. The weather condition now can be felt different from the weather some years back. Moreover, there are several other factors affecting the crop yield which are found to be ignored by most of the farmers.

It is being difficult to harvest suitable crops in right time due to the recent emerging problems like climate change and global warming. And as soil condition also keeps changing, farmers usually don’t know about that and keeps planting same crops in same field for years which also reduce the soil fertility and crops production. So, it’ll be beneficial in every aspect to predict crops for coming weather based on soil types.

The climate of world has changed a lot due to several factors. Therefore, the yield of crops isn’t found to be as expected. Weather has adverse effect on the yield of crops. Given the weather condition and type of soil, there will be at least one most profitable crop. The farmers can be assisted by suggesting the best crop for a particular area along with the forecast of next x days.

# 1.3 Objectives

The crucial part of this project is to solve underlying problems in agriculture sector and aid farmers. To overcome the problems faced by farmers in present days, certain objectives were drawn before the start of the project.

Major objectives of this project are:

* To predict upcoming weather for crop prediction.
* To classify weather and find suitable crops based on weather and soil analysis.

# 1.4 Scope and Limitation

## 1.4.1 Scope

This project will provide a reliable and accurate model for predicting suitable crops based on weather and soil type. So, this’ll partially solve the problem of crop losses due to climate change.

## 1.4.2 Limitation

Despite of its significant benefits and scopes, it has some limitations which are:

* There exist some exceptional cases where crops predicted by this system may not be suitable for that condition as other factors may also affects the crops.
* As this system only analyze the past weather data, this can’t accurately predict upcoming weather.

# 1.5 Report organization

Report organization is an important part of a report formation. It gives the overall pattern of the report, which contains summary information of the overall document. This document is categorized into several chapters and further is divided into sub chapters including all details of project.

* First chapter is about the introduction of the whole report. It includes short introduction of the system, scope and limitations, background study and objectives of the system.
* The second chapter includes the research methodologies in the project. It also includes feasibility study, functional and non-functional requirements. Data and Process Modeling are also included in this chapter.
* Third chapter is about the system design. It contains database design, interface design etc.
* The fourth chapter is about the implementation and testing procedures. It contains the detail about the tools that are required to design the system. In the testing section, different testing processes are included.
* The fifth chapter includes the maintenance and support of the system.
* The last chapter includes conclusion of the whole project. It also provides information about what further can be achieved from this project.

**CHAPTER 2**

**REQUIREMENT ANALYSIS AND FEASIBILITY STUDY**

# 2.1 Literature Review

Several articles and research paper where studies before the project actually took off. Many online as well as offline papers and journals were studies and the similar applications were checked out. Some major articles we consulted are listed below:

This review article written by Dr. Bharath Mishra focuses mostly on various attributes to be taken in consideration while applying data mining techniques in the field of agriculture and also concludes that the multidisciplinary approach of integrating computer science with agriculture will help in forecasting/ managing agricultural crops effectively [2].

Sally Jo Cunningham emphasizes on the usage of data mining techniques and its process model to derive innovative applications in the field of agriculture. He also visualizes the applications of data mining, the goal might be to use a model predictively, to provide automated classification of new instances. He finally concludes that the usage of Weka tool and its comprehensive suite of facilities for applying data mining techniques to large data sets [3].

Jayanta *Basak*, the author of this paper discusses about the central problem in weather and climate modelling, which is to predict the future states of the atmospheric system. Also, the atmospheric correlation is being discussed in this paper. The author has provided techniques for determining the strongest independent components in the multidimensional data set [4].

Thomas Van Klompenburg, Ayalew Kassahun and Cagatey Catal in their research paper did the Systematic Literature Review (SLR) to extract and synthesize the algorithms that have been used in crop yield prediction studies. They analyzed the most used features and algorithms while performing the crop yield study. They found out soil, temperature and rainfall to be the 3 most used features and Artificial Neural Network (ANN) to be the most used algorithm. They further studied the ANN algorithms and sorted 3 most used ANN algorithms which are Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM) and Deep Neural Network (DNN) [5].

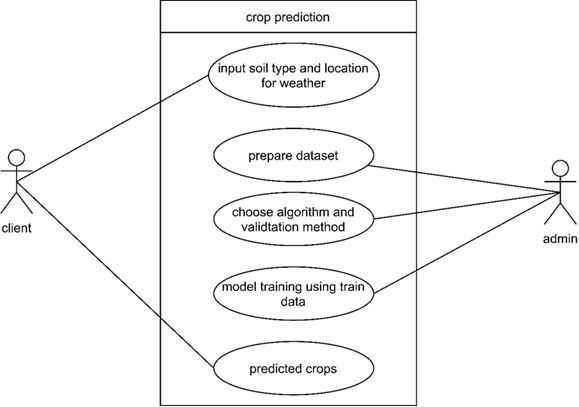
# 2.2 Requirement Analysis

While developing a system and before implementing it, it is necessary to analyze the whole system requirements. It is categorized into mainly two parts, namely: functional and non-functional requirements.

## 2.2.1 Functional Requirements

A functional requirement describes what a software system should do when it is given input. Some of the functional requirements are given below:

* User should enter soil types and location to get predicted crops.
* System should classify weather and also classify crops suitable on those weather.
* System should predict upcoming weather and predict suitable crops for that weather and provided soil types.



*Fig 2.1: - Use case diagram of proposed system*

# 2.2.2 **Non-functional Requirements**

A non-functional requirement describes how the system performs a certain function. Non-functional requirements generally specify the system’s quality attributes or characteristics.

Typical non-functional requirements include:

# 2.3 Feasibility Analysis

To evaluate whether the end product is feasible on the completion or not, the feasibility study needs to be done before actually starting the project. Feasibility study helps us draw a concrete decision whether the product is going to be sustain in the market or not and whether to start the project or not.

We performed some study and analyzed the system and get to know that it is feasible to make the system. Mainly four types of feasibility studies were done with this system analysis, namely: Economic feasibility, Operational feasibility, Technical feasibility and Schedule Feasibility.

## 2.3.1 Economic Feasibility

This system can be developed on low end computers and can be deployed on server with lower specs as only classification and prediction algorithms are used and won’t work on neural networks or deep learning so no any high-end computers with higher GPU’s, TPU’s and huge computation power is required. Thus, this system is economically feasible. Also, client doesn’t need to make any sort of payments in order to find the suitable crops using this system. Thus, the system is economically feasible.

## 2.3.2 Operational Feasibility

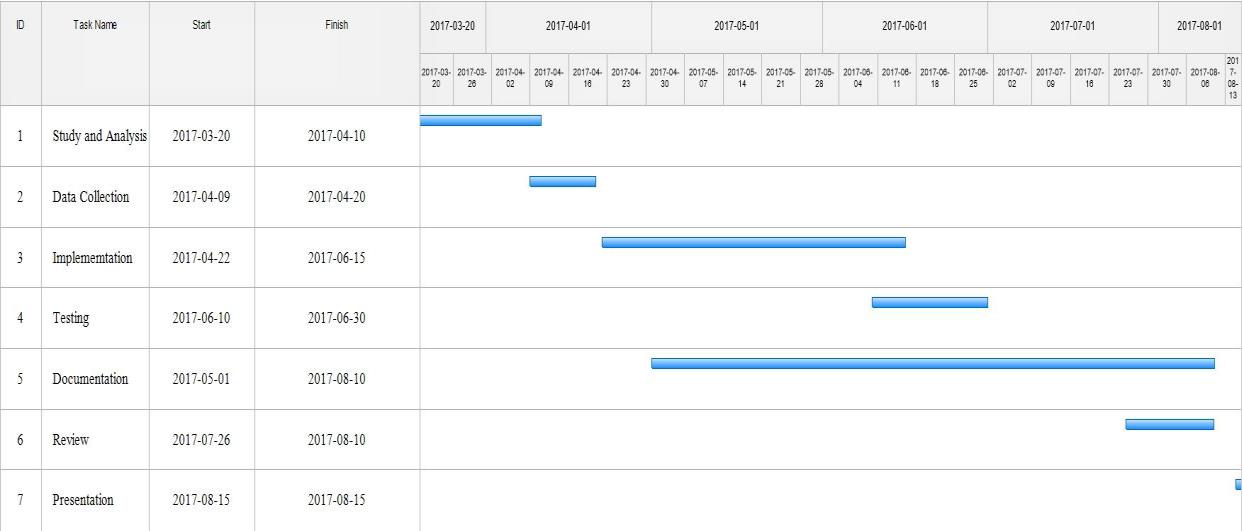
This system provides a simple web interface, which can be easily used by any type of users having basic idea of interacting with websites. This system will tend to provide accurate results according to the way the system needs to do. The system is operable in any browser in any operating system. Hence, this system is operationally feasible.

## 2.3.3 Technical Feasibility

Though this system will have predication and classification algorithms, this system will be built using simple algorithms so, there won’t be any technical complexities. Also, from the user’s perspective, user with basic computing knowledge can use this system. User can simply fill a small form with required information and find the most suitable crops under the required conditions.

## 2.3.4 Schedule Feasibility

A system is said to be scheduled feasible if it is implemented within the planned scheduled. We carried out the study on how much it will take to complete the task after studying the requirements and proposed plan.



*Fig 2.2: - Gantt chart for the proposed system*

We proposed the rough timeline so that we it would help us to perform our different project activities. Following Gantt chart shows the proposed schedule to perform the project:

Thus, creating the schedule and working on it makes it easier for developers to finish the project on time. Hence, this project is feasible if we completed this task within the designed Gantt chart.

# 2.4 Structuring System Requirement

In order to have a detailed understanding of how the system works, structuring the system requirement plays a pivotal role. Flow of the program execution and the data entities used can be clearly understood by structuring the system. The system involves data pre-processing and manipulation and calculation of data throughout the execution till it provides the desired output. These data items and flow of the data in the system are shown using different structuring models such as entities relationship diagram (ER diagram), Dataflow Diagram (DFD), Context diagram etc.

## 2.4.1 Process Modeling

Process modeling involves graphically representing the processes, or actions, that capture, manipulate, store, and distribute data between a system and its environment and among components within a system. Here, the different processes in a system are connected with sources and sinks so that the actual flow of data in the system is displayed. A common form of a process model is a data-flow diagram (DFD). A data-flow diagram is a graphic that illustrates the movement of data between external entities and the processes and data stores within a system. Although several different tools have been developed for process modeling, we focus solely on data-flow diagrams because they are useful tools for process modeling. Here we use level 0 DFD in order to illustrate the process of the system in detail.

## 2.4.1 Data collection

The application works on machine learning approach. Therefore, data is an integral part of the system. The system requires bulk information related to weather and crops. Data collection was a challenge in order to build the system.

Among the different methods of data collection, we used documents and records because the crops and weather data need to be accurate (or closely accurate) as far as possible in order to find the best suited crop.

Besides, the weather data collection was done via an API of Apple inc. known as darksky. Darksky provides the weather data of all major cities in the world. Many cities of Nepal were found listed in the darksky. So, for weather data, the system completely relies in the darksky API.

**CHAPTER 3**

**SYSTEM DESIGN**

# 3.1 System Design

System design is the process of designing the elements of a system such as the architecture, modules and components, the different interfaces of those components and the data that goes through that system. The major elements of a system include data, architecture, modules, components and interfaces.

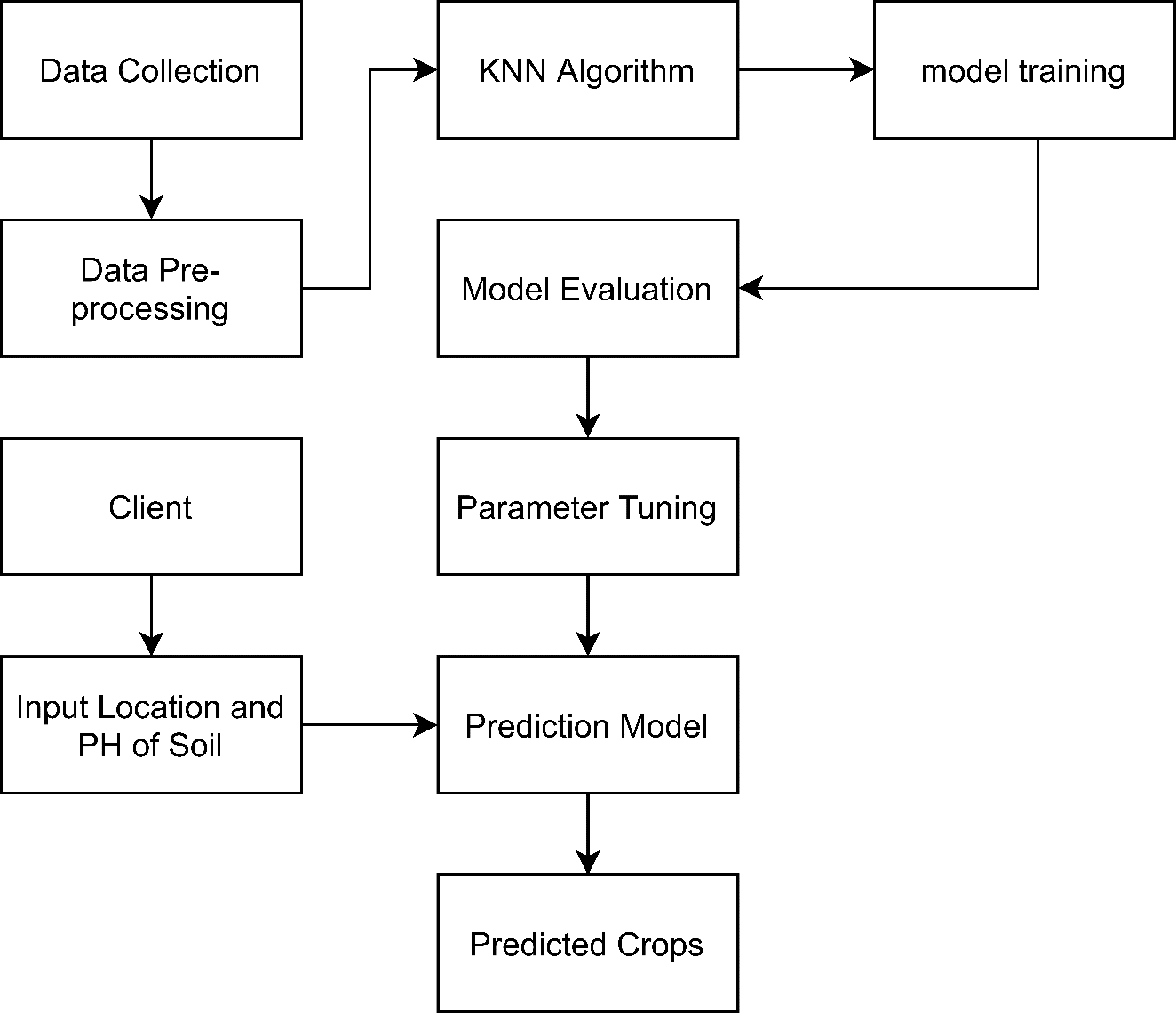
## 3.1.1 Interface Design

Interface is the shared boundary across which the components of the system exchange information and relate. Interface design is the specification of the interaction between a system and its environment. During interface design, the internal functions and process of the systems are completely ignored and the system is treated as a black box. Attention is focused on the dialogue between the target system and the users, devices, and other systems with which it interacts. The design problem statement produced during the problem analysis step should identify the people, other systems, and devices which are collectively called agents.

## 3.1.2 Process Design

Process design involves the representation of the way of process execution in the system. It is similar to the data flow diagram where the entire process of the system is decomposed into many small processes to show clear flow of execution in step-by-step manner.

The decomposition diagram below shows how the data flows in the system right from the start of the application to the point where user gets the end result



**Figure 3.1.2.1:** Process diagram for proposed work

**CHAPTER 4**

**IMPLEMENTATION AND TESTING**

# 4.1 Background Study

The system is developed by integrating many programming languages and scripting languages. The interface of the web application is developed in c# (dot net core) framework and some JavaScript and CSS. The main algorithm is however implemented in python.

Therefore, in order to develop the system, the basic knowledge of C#, Python and JavaScript was necessary for the developing team members. For the end users, no specific knowledge of programming language is required. Basic computer operating skills and a brief training is adequate to use the system.

# 4.2 Implementation Tools

After all the studies and analysis, the appropriate tools and materials needed to develop the system were sorted out. This included the selection of programming language and other tools which are essential in order to develop a proper system as expected. From the meeting among the team members, we found out the appropriate tools and platforms required to develop the Crop Prediction web application.

Different tools and technologies that we have used are given below:

4.2.1 Python:Python was used as the core programming language while developing the system. It is called as core programming language for the project because the main computing algorithm (KNN) was implemented using python. Python was used for creating the machine learning model which classifies the weather provided on the basis of the train data pre-fed to the system.

4.2.2 Flask**:** Flask is the web framework for python. Flask in this project was used in order to create an API which was fed to the web application which directly interacts with the user

## 4.2.3 MS SQL

We used MS SQL for storing and communicating information with the database. All the datasets and the location related data are stored in MS SQL server. We used MS SQL version 15.

## 4.2.4 Draw.io

In the system structuring part of the report organization, we have drawn ER Diagram, DFDs and Use case diagrams with the help of the case tool draw.io. It is an easy tool for constructing such diagrams providing necessary images, shapes etc. It also provides some standards that are available in designing.

# 4.3 Algorithm Implementation

Different algorithms and papers were studied and KNN algorithm was found to be most suitable for crops prediction using soil and weather analysis.



## K-Nearest Neighbors

KNN is a type of instance-based learning or the lazy learning where the function is only approximated locally and all computation is deferred until classification. In this particular project, the KNN algorithm is specifically used in order to classify the weather data into different classifications. This is also responsible for recommending the right crops as per the provided weather conditions and soil types.

The case is classified by a majority of votes of its neighbor and the case is assigned to the class common amongst its K nearest neighbors measured by a distance function. If k=1, then the case is simply assigned to the class of its nearest neighbor. In this case we use Euclidian distance as a distance function.

**Euclidean Distance:** Euclidian distance between two points measures the length of segments connecting the two points. It can be given as:

Where,

Xi and Yi are the end points of the line segments,

K is the limit of iterative summation

Once the Euclidian distance is evaluated, the training set can be used to classify the weather. The major steps involved in KNN algorithm are:

1. Load the data
2. Choose and initialize appropriate value of K
3. For getting the classification, iterate from 1 to total number of training data
4. Calculate distance between test data and each row of training data using Euclidian distance.
5. Sort the calculated distances in ascending order based on distance values
6. Get top K rows from the sorted array
7. Get the most frequent class of these rows
8. Return the predicted class

To select appropriate value of K for the available data, the KNN algorithm is run several times with different value of K. The value of K is chosen in such a way that it reduces the number of errors encountered while maintaining the algorithm’s ability to accurately make predictions when it is given data it hasn’t seen before.

# 4.4. Testing