AgriSense

A Project Report

Submitted by

Vidhi Agrawal
Darpan Bendre
Sadhika Bhatia
Darshan Daliya

Under the Guidance of

Prof. Kamal Mistry

In partial fulfilment for the award of the degree of

BACHELORS OF TECHNOLOGY

COMPUTER ENGINEERING

At

MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT AND ENGINEERING

APRIL 2021

DECLARATION

We, Vidhi Agrawal, Darpan Bendre, Sadhika Bhatia, Darshan Daliya, Roll No. B003, B008, B010, B018 B.Tech (Computer Engineering), VIII semester understand that plagiarism is defined as anyone or combination of the following:

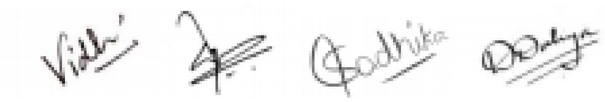
1. Un-credited verbatim copying of individual sentences, paragraphs or illustration (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Un-credited improper paraphrasing of pages paragraphs (changing a few words phrases, or rearranging the original sentence order)

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did wrote what. (Source:IEEE, The institute, Dec. 2004)

4. I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

5. I affirm that no portion of my work can be considered as plagiarism and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the seminar/ project report may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Signature of the Student: 

Name: Vidhi Agrawal, Darpan Bendre, Sadhika Bhatia, Darshan Daliya

Roll No. B003, B008, B010, B018

Place: Mumbai

Date: 13-04-2021

CERTIFICATE

This is to certify that the project entitled “AgriSense” is the bonafide work carried out by Vidhi Agrawal (B003), Darpan Bendre (B008), Sadhika Bhatia (B010), Darshan Daliya (B018) of B.Tech (Computer Engineering), MPSTME (NMIMS), Mumbai, during the VIII semester of the academic year 2020-2021, in partial fulfillment of the requirements for the award of the Degree of Bachelors of Engineering as per the norms prescribed by NMIMS. The project work has been assessed and found to be satisfactory.

Prof. Kamal Mistry

Internal Mentor

Examiner 1

Examiner 2

Dean

ABSTRACT

Agriculture contributes to over 17% of the Indian economy. It is one field in which the right use of technology can create a massive impact in the lives of farmers who have been growing the same crops in the same regions for decades. In regions where the conditions are traditionally suited to grow high profit cash crops farmers have been flourishing but this can only be said for a very small number of farmers. In most regions farming is not profitable and is full of uncertainties. While farmers know that historically a certain crop is suited for the region but the information they lack is that over the years due to the changes in agricultural techniques and research on seed production, there is a chance that a certain highly profitable crop which has never been grown in a particular region can be grown on their land. We are developing a Region Compatibility based Crop Recommendation system and Market Price Predictor to give the farmers all the information they require to find out if a particular variety cash crop can be grown on their land. This can also be used by companies who want to do contract farming in a region or by insurance companies. We will be studying data about a particular region, namely its important factors like Precipitation, Land Surface Temperature and elevation and matching that area with a similar area where a highly profitable crop is grown. We will be extracting this data using remote sensing. We will finally give a comparison in percentage to show how similar the two regions are, this will help the farmers decide if they should be planting a certain crop. We will also give a market analysis to show how much the farmer stand to earning he chooses to grow a certain crop. We are striving to bridge the gap between the farmers and technology.

Table of Contents

List of Tables

List of Figures

Abbreviations

Abbreviation	Description
GEE	Google Earth Engine
LST	Land Surface Temperature
DSM	Digital Soil Mapping
EVI	Enhanced Vegetation Index
VOD	Vegetation Optical Depth
MODIS	Moderate Resolution Imaging Spectro Diameter
SMAP	Soil Moil Moisture Active Passive
SAR	Synthetic Aperture Radar
PA	Precision Agriculture
NPK	Nitrogen Phosphorus Potassium
CHIRPS Pentad	Climate Hazard Group Infrared Precipitation with Station Data
SRTM	Shutter Radar Topography Mission
ML	Machine Learning
ROI	Region of Interest
PCA	Principle Component Analysis
KNN	K- Nearest Neighbours
SVM	Support Vector Machine
SVC	Support Vector Classifier
CNN	Convolutional Neural Network
CART	Classification and Regression Tree
BPNN	Back Propagation Neural Network
RBFN	Radial Basis Function Network
RMSE	Root Mean Squared Error
MSE	Mean Squared Error
MAE	Mean Absolute Error
BI	Business Intelligence

Chapter 1
Introduction

1.1 Motivation

Agriculture remains to be a great player in the generation of revenue and a source of food for many people all over the world, especially India. Over 70 per cent of the rural households depend on agriculture. Agriculture contributes about 17% to the total GDP and provides employment to over 60% of the population. Over the past years, this sector has seen a lot of changes and advancement in the different farming approaches and techniques. For example, nowadays, there is the use of organic fertilizer, the consumption of reduced amounts of pesticides, the use of different tractors and machinery. The availability of such inputs has seen the need for the use of natural resources and process with aim of improving agricultural output and reducing costs. Even with the economy depending so much on Agriculture it is one of the fields in which the use of technology is negligible. The use of modern technology in agriculture comes with a lot of benefits. They don't have knowledge about other cash crops that can be grown as they don't have enough information related to it.

The lack of information accessibility and knowledge that farmers have pertaining to the kind of crops that lead to a higher yield and more profitability. Over generations farmers belonging to a particular region continue to grow the same crops over and over again. There are two reasons for this.

1. They don't have knowledge about other cash crops that can be grown as they don't have enough information related to it.
2. They aren't sure of the price they will be getting and have no idea if it will be more profitable to grow the alternate crops and hence aren't willing to change as they think it is a big risk.

Due to these lakhs of farmers are struggling to make ends meet. Agriculture being the backbone of the Indian Economy, there is a need to empower our farmers with information obtained through technology.

Thus, to address this problem faced by farmers will be developing "Region Compatibility based Crop Recommendation system and Market Price Prediction" through which where farmers will be able to deduce the possibility of growing alternate crops based on satellite imaging and also know the increase in profit that they can generate based on market price prediction. We will be using Google Earth Engine to extract data, Power BI to visualize our results, MS Excel to compile data and Python for

coding.
We have done a review of literature through which we researched and studied the different cash crops that are grown in India and the conditions required to grow them. We then selected the crops which will give the highest profitability based on historic data collected through various reputable sources
We have then implemented the system by running this data against the specific regions chosen by the user i.e. a farmer in this case to give a percentage match.

1.2 Definition of Problem

'To design Region Compatibility based Crop Recommendation system and Market Price Predictor"

This is a platform which where farmers will be able to deduce the possibility of growing alternate crops based on satellite imaging and also know the increase in profit that they can generate based on market price prediction. Region compatibility as the name suggest will help farmers decide if a particular land is suitable to grow an alternate crop. It will predict this based on the data we extract from google earth engine. We will be studying historical data (12 years) of crops and the optimum conditions that they have been growing in for years. We will then use remote sensing to map these particular regions and then compare how similar they are to a region on which a farmer is wanting to grow that crop. We will also be basing the recommendation based on the higher profit that he stands to gain. We will be doing a thorough price analysis to make sure that the farmers make a well-informed decision. We will be showing a percentage comparison where that particular crop is growing and how similar the land being compared too is in terms of elevation, temperature, precipitation and soil quality

We will be using Google Earth Engine to extract data, Power BI to visualise our results, MS Excel to compile data and Python for coding.

1.3 Project Scope

With the uncertain climatic conditions causing massive losses for farmers we feel there is a need to empower them with technology through technology. Hence we decided to work on three major fronts (Farmer Awareness, Comparison and Market Prediction

1.3.1 Farmer Awareness

Farmers don't have enough knowledge about ways in which they can maximize output from their fields. They also don't have enough information as to which alternate crops can grow in their region. There is always a possibility that a particular crop could be grown in a region which would give a higher yield and would also be more profitable. Using satellite imaging our platform will help them study their geographical region, compare it with alternate regions and find out how feasible it is to grow another crop.

1.3.2 Comparison

Even if a farmer is told that a particular crop can be grown in a particular region there is a chance that the crop fails. It is a big risk which a farmer can't afford to take. Using our platform a farmer will be able to able to compare another location/region where a particular crop is growing and the factors such as soil and weather conditions of both regions will be compared to give an exact percentage showing how similar the two regions are. This process will be done for all cash crops. So a farmer can compare crops grow in different regions and he will know which one has the least chance of failure.

1.3.3 Market Prediction

The main concern for a farmer remains if he will receive a fair price for the produce grown by him. Through our platform he will be able to know the predicted prices of the market for the alternate crop being grown depending on the month of harvest. We will be analyzing the data over the last 5 years to predict trends for the same.

We have done a review of literature through which we researched and studied the different cash crops that are grown in India and the conditions required to grow them. We then selected the crops which will give the highest profitability based on historic data collected through various reputable sources. We have then implemented the system by running this data against the specific regions chosen by the user i.e a farmer in this case to give a percentage match.

Chapter 2

Review of Literature

More and more researchers have begun to identify the problem in Indian agriculture and are increasingly dedicating their time and efforts to help alleviate the issue. Our main target was to understand the working of the google earth engine and find out how well it can be used to extract relevant data.

While researching we came across [1] in which comparison of pixel-based approaches to crop mapping in Ukraine was done and explored efficiency of the Google Earth Engine (GEE)cloud platform for solving “Big Data “problem and providing high resolution crop classification map for large territory. We found that Google Earth Engine (GEE) provided very good performance in enabling access to remote sensing products through the cloud platform they went hand in hand with [2] which the various applications of Google earth engine in the Agriculture sector were explored. Various aspects such as vegetation mapping and monitoring, landcover mapping and other remote sensing agricultural applications along with a little insight into the different satellites that we can refer to calculate the stated aspects. When we looked more into the various ways in which remote sensing can be used to measure vegetation depth we came across [3]The proposed combination of multi-sensor (optical and microwave) remote sensing data for crop yield estimation and forecasting used two novel approaches. We the lag between Enhanced Vegetation Index (EVI) derived from MODIS and Vegetation Optical Depth (VOD) derived from SMAP was used as a new joint metric combining the information from the two satellite sensors in a unique feature or descriptor. The second approach used avoided summarizing statistics and uses machine learning to combine full time series of EVI and VOD. Results confirmed the value of using both EVI and VOD at the same time, and the advantage of using automatic machine learning models for crop yield/production estimation. In line with this in [4]we studied that nine spectral and temperature bands from relatively low-resolution satellite images were used as their feature for predicting county-level corn and soybean yields in US and they demonstrated that it is possible to achieve better crop yield prediction accuracy using MODIS satellite imagery by employing more complex models. We had gained information regarding how satellite imaging can be used for various purposes but to study how to actually map the crops based on multiple sources or satellites we looked into[5].

Finally, machine learning methods were used to map the spatial distribution of soil nutrients. Based on this we understood that when we can to potentially map the crops to certain region soil plays a very important role in that aspect. To explore Mapping Cropping Practices on a National we assessed the use of annual, quarterly, and eight-day temporal features for mapping five cropping practices on annual croplands across Turkey in [6]. Five cropping practices: Spring and winter cropping, summer cropping, semi-aquatic cropping, double cropping, and greenhouse cultivation were studied. Our study presents an open and readily available framework for detailed cropland mapping over large areas, which bears the potential to inform assessments of land use intensity, as well as land and water resource demands. This helped us in the mapping of regions and extraction of data

Table 2.1: Crop-based Research[7]

Crop/Factor	**Temp (ºC)**	**Soil Type**	**Rainfall/Humidity**	**Elevation**	**Time**	**Region**
Mushroom	25-30	Clayey and Loamy	55-70%	900m	14 weeks	Orissa, Karnataka, Maharashtra, North East states
Cashew	25-49	Laterite, Red and Coastal Sandy	50cm 250cm	700m	18 weeks	Gujarat, Maharashtra, Karnataka, Tamil Nadu, Karnataka
Rapeseed	10-30	Clayey and Loamy	350-550mm	650m	24 weeks	Rajasthan, Gujarat, West Bengal
Jute	24-37	Alluvial	1000mm	500-1800m	24 weeks	Bihar, Orissa, Andhra Pradesh, Meghalaya

Avocado	12-30	Range of Soils	2000mm	700-900m	5-6 years	Karnataka, Kerala, Tamil Nadu
Rubber	20-35	Range of soil	2000mm	600m	6-7 years	Kerala, North East states
Tea	20-30	Well drained, deep, friable loams	1500mm	2000m	8 months	North East, Kerala, Uttarakhand
Coffee	17-25	Rich moist, well drained, loamy	1000-2000mm	200-300m Robusta 1000-2000-Arabica	75 days to sprout, 2 years for final use	Karnataka, Kerala, Tamil Nadu, Orissa, Andhra Pradesh, Telangana, North East

[8]considered a facet of precision agriculture that concentrates on plant-driven crop management. By monitoring soil, crop and climate in a field and providing a decision support system that is able to learn, it is possible to deliver treatments, such as irrigation, fertilizer and pesticide application, for specific parts of a field in real time and proactively. In this context, we have applied machine learning techniques to automatically extract new knowledge in the form of generalized decision rules towards the best administration of natural resources like water. The machine learning application model suggested in this paper is based on an inductive and iterative process of discovering knowledge on the basis of which, patterns and associations having arisen initially are re-examined to expand the pre-existing knowledge. The result of this study was the creation of an effective set of decision rules used to predict the plants' state and the prevention of unpleasant impacts from the water stress in plants.

In the diverse procedures presented in this paper[9] include KNN, Similarity- based Models, Ensemble-based Models, Neural Networks, etc. These algorithms take into account various different factors that are external in nature like meteorological data, temperature and others like soil profile and texture to give best recommendations which not only lead to better yields but also minimum use of resources and capital. The most important attributes using techniques like principal component analysis (PCA), linear discriminant analysis (LDA). These extracted features are used to train models like Naive Bayes, Random Forest, KNN, etc. using training data and performance is evaluated on test data using techniques such as cross-validation, accuracy, RMSE, precision, recall. [10] discusses a hybrid classifier model which is used in optimizing the feature and help in the predicting the two classes of yield that is good and bad yield. The model uses SVM_GWO i.e. grey wolf optimizer along with SVM to design a classification model with better accuracy than SVM. The accuracy obtained by this model is 77%. It describes a comparative study related to how the new model has worked in respect to the basic SVM regression model. [11]presents the impact of machine learning in precision agriculture. Dataset considered for the study is related to boneyards. The study presents a comparison of an innovative machine learning methodology compared to a baseline used classically on vineyard and agricultural objects. The final overall accuracy of the algorithm is of 84:275%. The vines are very well classified. Few of them are wrongly classified. The DTE is very stable throughout all iterations and provides a stable accuracy. Machine Learning used to enrich and improve the detection of precise agricultural objects is also discussed in this study and opens new perspectives for the future of high precision agriculture.

Table 2.1: Algorithm-based Research

Paper	**Input**	**Output**	**Algorithm**	**Accuracy/Error Rate**
AgroConsultant[12]	Soil Type, Aquifer, Soil pH, Top soil Characteristics, Season	Recommends a cash crop and a cereal	• Decision Tree • KNN • Random Forest • Neural networks	90.20% 89.78% 90.43% 91%

Synergistic integration of optical and microwave satellite data for crop yield estimation[3]	Growing and senescence stages of crops: corn, soybean and wheat.	Estimate the county-level surveyed total production, as well as individual yields of the major crops grown in the region	• Regularized Linear Regression(RLR) • Kernel Ridge Regression(KRR)	61% 71%
Understanding Satellite-Imagery-Based Crop Yield Predictions[4]	Multispectral Images	Corn and Soybean yield	• CNN	5.24 RMSE
Crop Suitability and Fertilizers Recommendation Using Data Mining Techniques[13]	• District, State, Season • Soil Characteristics	• Crop Suitability • Fertilizer Recommendation	• Random Forest • K means	68%
Large Scale Crop Mapping from Multi-Source Remote Sensing Images in Google Earth Engine[5]	Multispectral Images	Growth Trends of 3 crops in a particular region	CART- Classification and Regression Tree	84.25%

Using Google's cloud-based platform for digital soil mapping[14]	29,784 soil profiles from the NASIS database (USDA–NRCS, 2014) which contains laboratory measurements of topsoil organic carbon content	prediction of a categorical property (soil class) based on a classification algorithm, and prediction of a continuous property (soil organic carbon content) via a regression technique.	Serial Classifier algorithm (similar to random forests)	84.25%

[15]proposes a random forest based adaptive model for real-time electricity price forecasting. The main contribution lies in the following two aspects. Firstly, the model provides a confidence interval associated with the prediction. Secondly, the model can adjust to the latest forecasting scenarios by updating itself with new observations. A case study has been presented to prove the validity of the proposed model.

A random forest-based stock forecasting algorithm is presented in [16], and their operational characteristics are analysed using two KOSPI stock datasets. The algorithm is designed to automatically adjust the number of decision trees and the maximum depth of the trees through a data training process.

In [17] the vegetable price of Beijing wholesale Market, three models which are the BPNN, neural network based on GA and RBFN are established separately. Based on that, an integrated prediction model is constructed.

Through this we gained through knowledge about the use of Google Earth Engine, how mapping could be done efficiently and effectively using satellite imaging. We further gained an understanding of the Machine learning algorithms used to develop a robust recommendation system. Furthermore we looked into existing research on market price analysis. This gave us a fair idea of the usage of various ML models based on applications and how it can be applied to build our target product.

Chapter 3

Analysis & Design

3.1 Architecture

The architecture of the system is divided into two section the client side and the server-side. On the client-side the user interacts with the web interface whereas on the server-side actual computations are done.

Client-side: Input parameters are taken from the user and corresponding output for compatibility percentage and price analysis is displayed on the web interface of the system.

Server-side: From the client-side a database is prepared through Remote sensing(Google Earth Engine) based on the current yield region of the respective crops((Phenological features). This data is used to train the SVM ML model and a compatibility score of crops is calculated for the desired user inputs.

The second part of the system consists of the Random Forest ML model which uses historic Price Data for the price prediction of coffee across different markets. The calculated result is then passed to the web interface where it is displayed to the user as output.

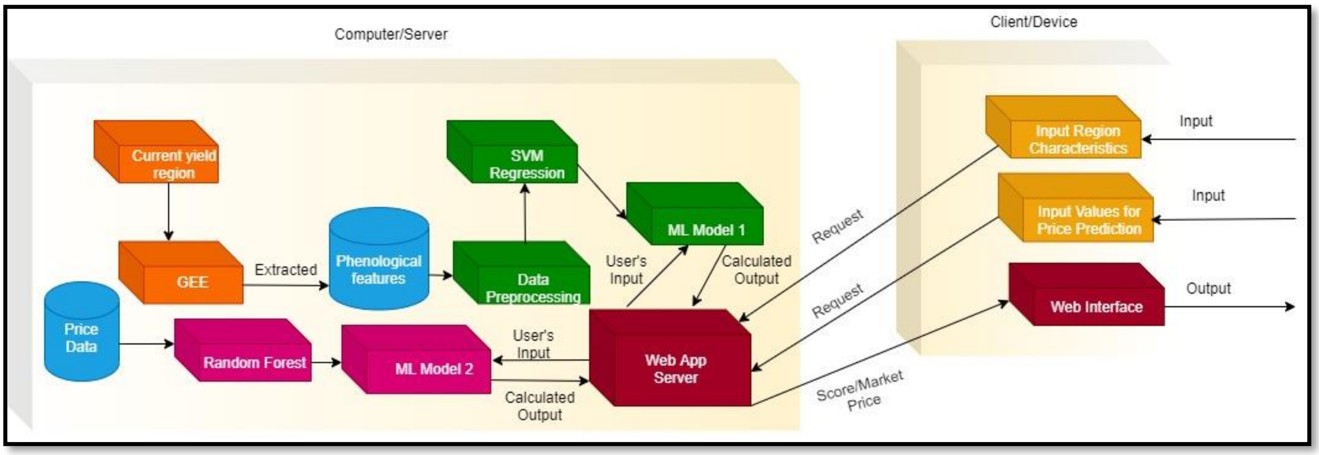


Fig 3.1: Architecture of the Proposed System

Software analysis and design includes all activities, which help the transformation of requirement specification into implementation.

3.2 Use Case Diagram

The use case is essentially a primary example of how the proposed software application or system is meant to be used, from the user's point of view. A use case diagram will typically show system 'actors' (humans or other entities external to the system) and how they interact with the system. Technically, each action such a system actor can perform with the application or system is considered to be a separate use case. In our model we have three actors:

i) User-Client

ii) Administrator (has unrestricted access)

iii) Database Repository

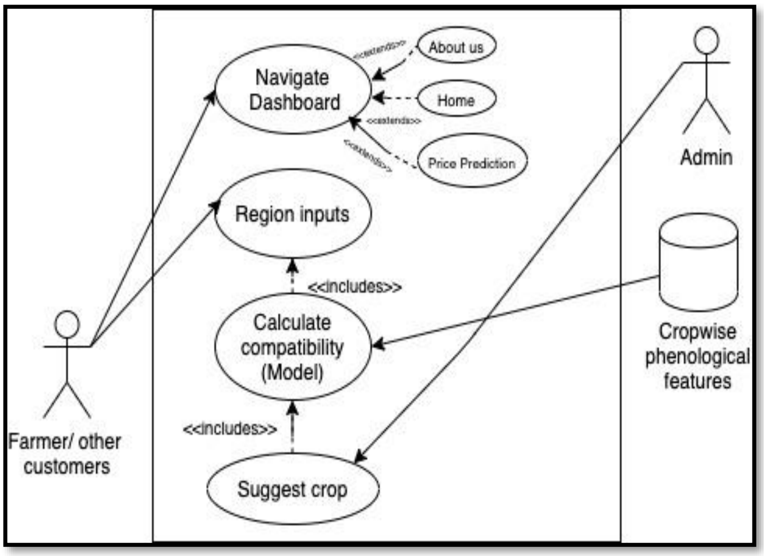


Fig 3.2.1: User Story 1- Check crop compatibility

User Scenario 1 we are considering the scenario in which the user is checking the compatibility of a particular crop for a given region.

Actor 1 accessing the navigation dashboard and the region inputs is the farmer or any other customer like a financial company, Agronomist etc. Actor 2 is the admin who has access to the machine learning model based Compatibility module which is linked to the final output of suggesting the crop. The machine learning model is able to make the decision using data like phenological features stored in the data repository.

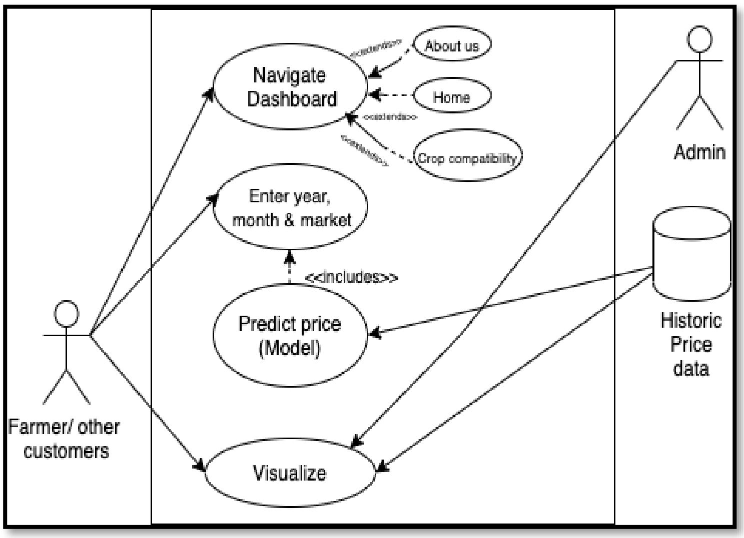


Fig 3.2.2: User Story 2- Predict and Review Market Prices

User Scenario 2 we are considering the scenario in which the user is checking market prediction for a particular crop.

Actor 1 accessing the navigation dashboard and entering input details like year, month and market is the farmer or any other customer like a financial company, Agronomist etc. Actor 2 is the admin who has access to the machine learning model based price predicting module which is linked to the final output of predicting the market price. The machine learning model is able to make the decision using historical data stored in the data repository.

3.3 Sequence Diagram

Sequence Diagrams are interaction diagrams that detail how operations are carried out. They capture the interaction between objects in the context of a collaboration. Sequence Diagrams are time focus and they show the order of the interaction visually by using the vertical axis of the diagram to represent time what messages are sent and when. Below is the Sequence Diagram for our system.

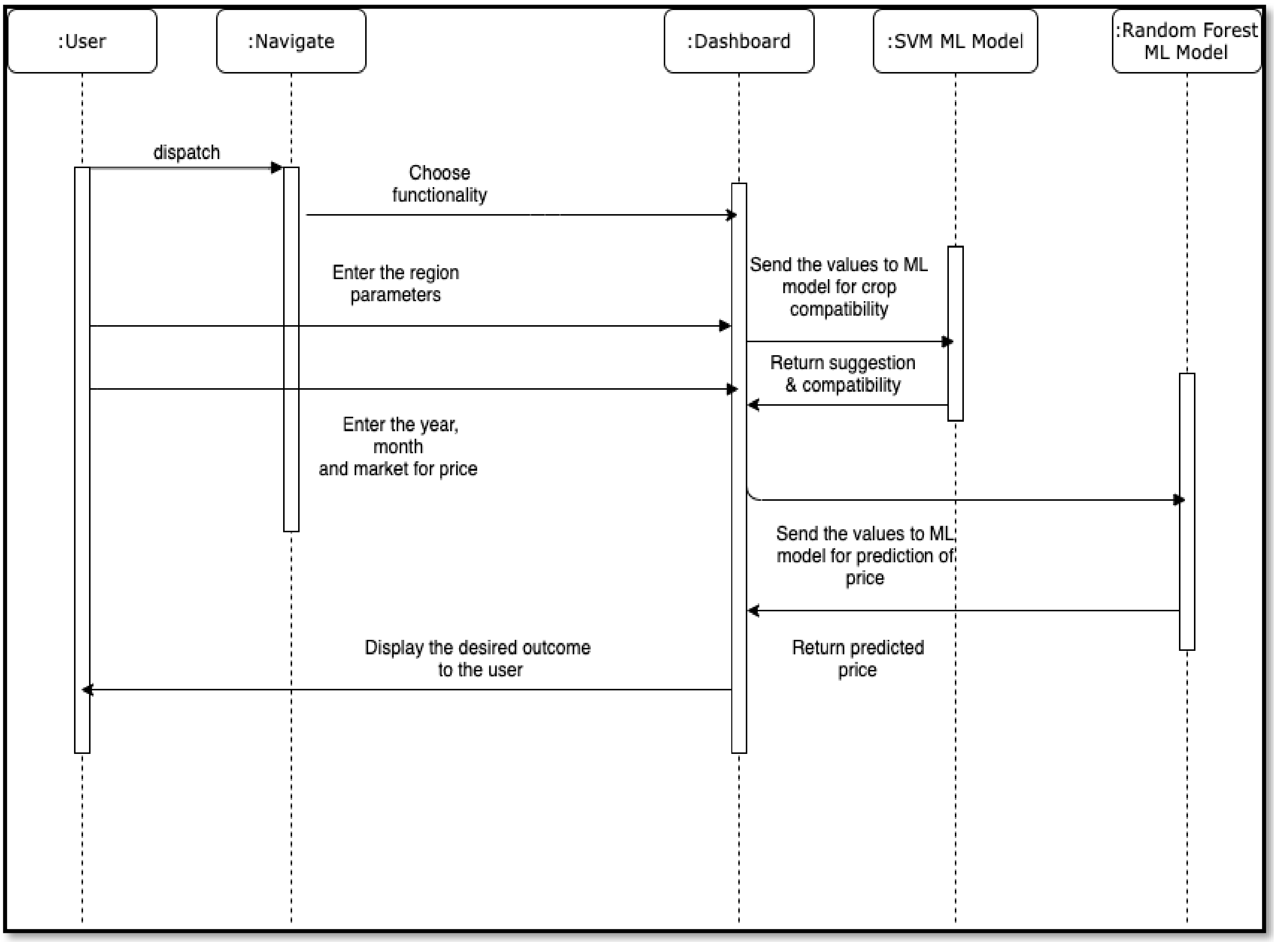


Fig 3.3: Sequence Diagram

The above diagram shows sequential interaction at every stage within the system. The arrows represent the flow of data amongst the different modules within the system, it shows what input is given to the module and what output is generated accordingly

Chapter 4
Implementation

Implementation method is a systematically structured approach to effectively integrate a software based service or component into the workflow of an organizational structure or an individual end-user.

4.1 Environment

4.1.1 Software Specifications

- **Google Earth Engine:**

Google Earth Engine is a planetary-scale platform for Earth science data & analysis. It combines a multi petabyte catalogue of satellite imagery and geospatial datasets with planetary-scale analysis capabilities and makes it available for scientists, researchers, and developers to detect changes, map trends, and quantify differences on the Earth's surface.

- **Jupyter Notebook:**

The *Jupyter Notebook* is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modelling, data visualization, machine learning, and much more.

- **Visual Studio Code:**

Visual Studio Code is a free source-code editor made by Microsoft for Windows, Linux and macOS. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.

- **Power BI:**

Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights

- **Draw.io:**

Draw.io is an online diagram editor built that enables the user to create flowcharts, UML, entity relation, network diagrams, mock-ups and more.

4.1.2 Hardware Specifications:

Hardware requirements are of minimum consideration in this project. We need a good internet connection and a system with the following features:

- Windows based platform:
  - Version:
  - Windows 10 or Windows 8.1
  - Windows 8
  - Windows 7 (32-bit and 64 bit version required)
  - Windows Vista Windows XP SP3 (32-bit version only)
  - Processor: At least 1 GHz
  - RAM: At least 2 GB
- Mac based platform:
  - Version: Mac OS X v10.6.6 or later
  - Processor: An Intel Core 2 Duo, Core i3, Core i5, Core i7, or Xeon Processor
  - RAM: 2 GB

4.2 System 1: Region Compatibility

GEE enabled us to get a global-scale insight which helped us to prepare our database by remote sensing. We used interactive time-lapse viewer to travel back in time and analyse the change in environmental patterns over the past years. After a rigorous literature survey, we finalised on the following:

Crops

1. Coffee
2. Avocado

Attributes

1. Land Surface Temperature
2. Precipitation
3. Elevation

Districts

1. Chikmagaluru (Karnataka)

2. Madikeri (Karnataka)

3. Nilgiris (Tamil Nadu)

4. Tirap (Arunachal Pradesh)

5. Changlang (Arunachal Pradesh)

6. Wayanad (Kerala)

7. Munnar (Sikkim)

8. Hassan (Karnataka)

9. Yercaud (Karnataka)

10. Kodaikanal (Tamil Nadu)

Duration

12 Years (2007-2019)

4.2.1 Importing Rasters

After selecting the factors, we want to analyse (precipitation, land-surface temperature and elevation) our first step was to select the target satellites for the remote sensing of this data.

The three rasters that we are working with:

1. **CHIRPS Pentad: Climate Hazards Group InfraRed Precipitation with Station Data (version 2.0 final)-** Analyzing long term trends in precipitation data for a specific region of interest (ROI)

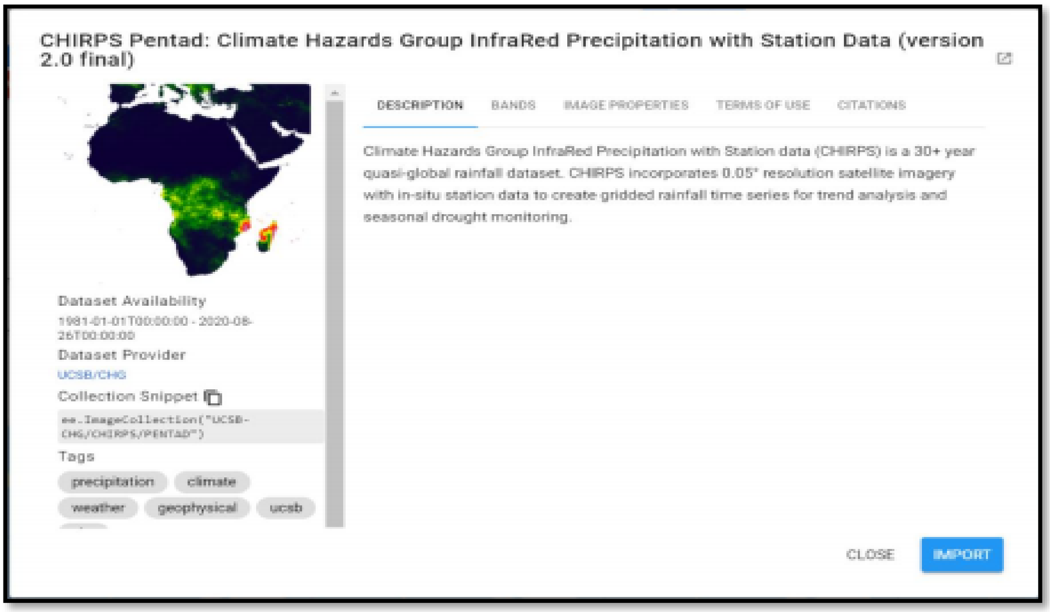


Fig 4.2.1.1: Precipitation Raster

2. **MOD11A2.006 Terra Land Surface Temperature and Emissivity 8-Day Global 1km-** Analyzing long term trends in land surface temperature data for a specific region of interest (ROI)

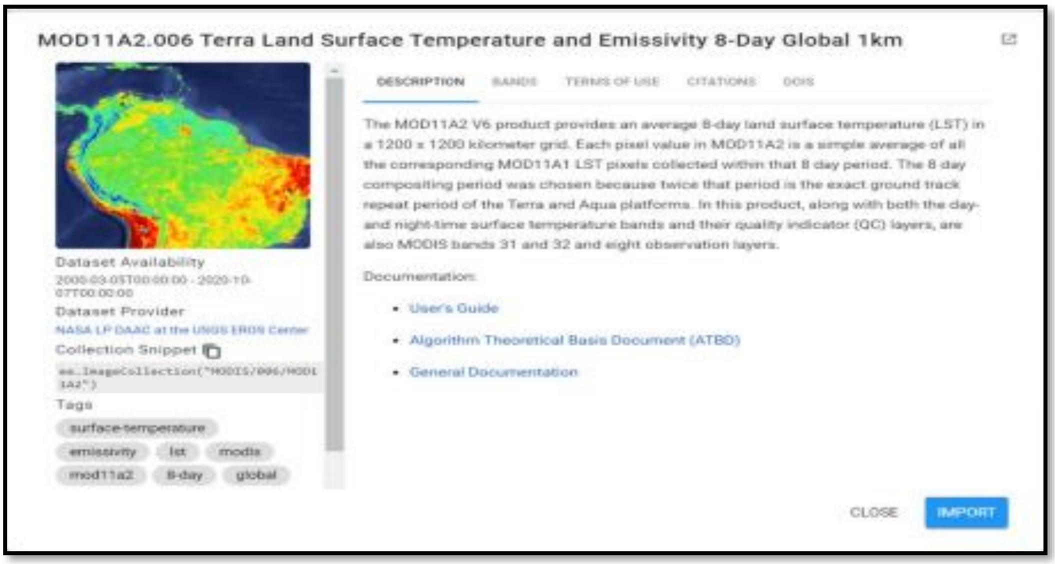


Fig 4.2.1.2: LST Raster

3. **NASA SRTM Digital Elevation 30m-** Analyzing elevation data for a specific region of interest (ROI)

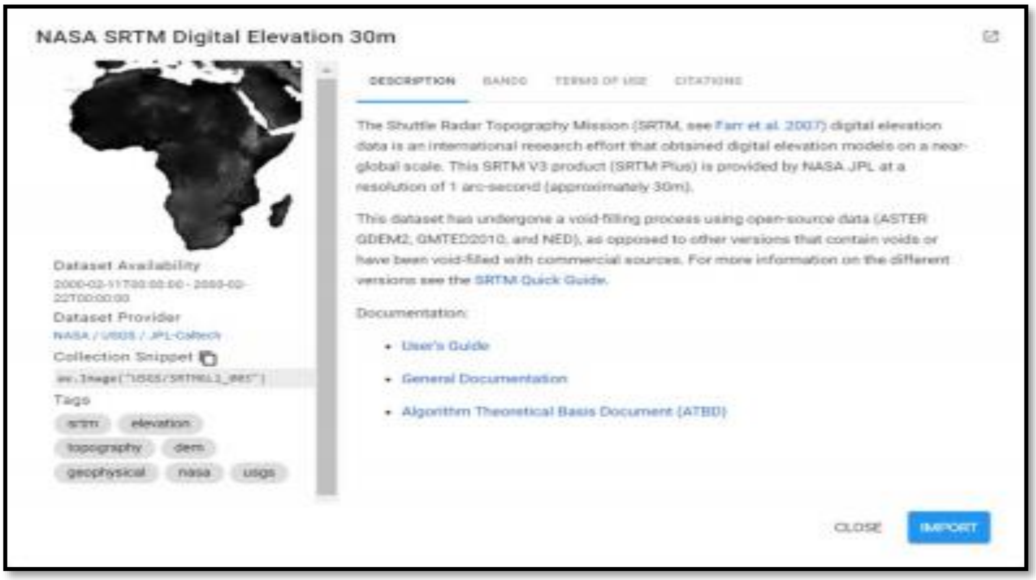


Fig 4.2.1.3: Elevation Raster

4.2.2 Defining ROI

Our next step is to select a particular region of interest for which we want to gather the data. We have analysed over 7 districts for each crop over a span of 12 years. ROI is defined by making a polygon from geometry imports option as shown in the image below.

Fig 4.2.2: ROI Selection

4.2.3 Code to extract Data

Once we have selected the desired ROI, we need to extract data (precipitation, land-surface temperature and elevation) for that particular region. This is achieved by writing specific code for the particular parameter in the script box of GEE using either JavaScript or Python.

Code for different parameters is as follows:

1. **Land-Surface Temperature:** Through this code snippet we are extracting the land-surface temperature in degree Celsius for the specified ROI for a span of 12 years. We have plotted a time-series chart for the same to understand it better.
2. **Precipitation-** Through this code snippet we are extracting the annual mean precipitation in mm for the specified ROI for a span of 12 years. We have plotted a time-series chart for the same to understand it better.
3. **Elevation-** Through this code snippet we are extracting the mean elevation in meters for the specified ROI for a span of 12 years.

4.2.4 Output

After the successful execution of the codes we generated the following outputs:

1. Land-surface temperature (LST) time series chart- The chart shows graphical representation of

LST in degree Celsius for the particular ROI over a span of 12 years.

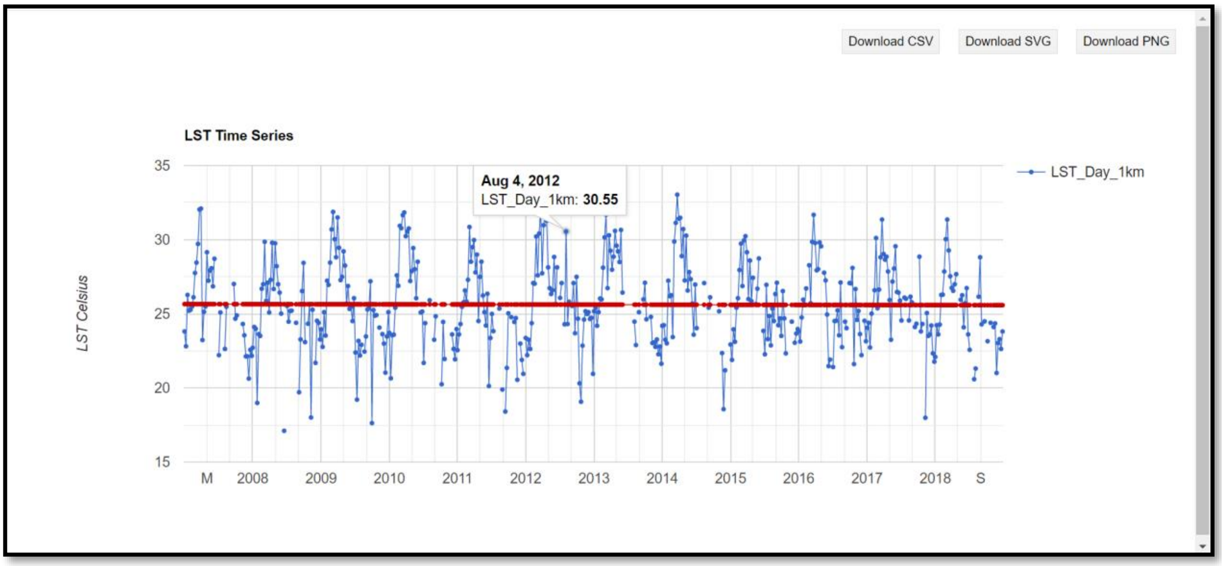


Fig 4.2.4.1: Graphical Representation LST

2. Precipitation full time series chart- The chart shows graphical representation of precipitation in mm for the particular ROI over a span of 12 years.

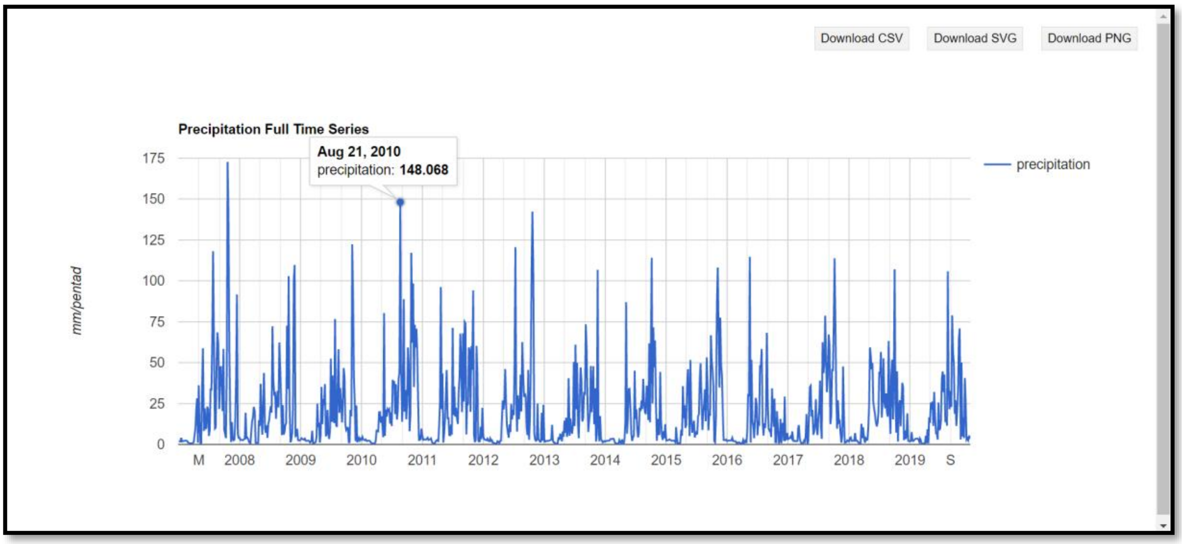


Fig 4.2.4.2: Graphical Representation Precipitation

3. **Elevation-** Mean elevation is obtained in meters for the particular ROI.

Fig 4.2.4.3: Elevation Analysed

4.2.5 Data Collected

The following data is collected and similarly we have collected data for other districts

Temperature

	A	B	C
14	Apr 7, 2014	39.07	
15	Apr 15, 2014	36.61	
16	Apr 23, 2014	33.17	
17	May 1, 2014	34.69	
18	May 9, 2014	34.69	
19	May 17, 2014	33.65	
20	May 25, 2014	34.21	
21	Jun 2, 2014	31.95	
22	Jun 10, 2014	28.45	
23	Jun 18, 2014	27.13	
24	Jun 26, 2014	35.63	
25	Jul 4, 2014		
26	Jul 12, 2014		
27	Jul 20, 2014		
28	Jul 28, 2014	26.05	
29	Aug 5, 2014		
30	Aug 13, 2014		
31	Aug 21, 2014	28.21	
32	Aug 29, 2014		
33	Sep 6, 2014	27.97	
34	Sep 14, 2014	29.81	
35	Sep 22, 2014	30.77	

Hassan_Karnataka_t1

Precipitation

	A	B
21	Apr 6, 2014	8.949
22	Apr 11, 2014	23.702
23	Apr 16, 2014	15.032
24	Apr 21, 2014	2.55
25	Apr 26, 2014	22.051
26	May 1, 2014	27.183
27	May 6, 2014	38.282
28	May 11, 2014	1.179
29	May 16, 2014	20.347
30	May 21, 2014	19.464
31	May 26, 2014	61.447
32	Jun 1, 2014	28.355
33	Jun 6, 2014	12.774
34	Jun 11, 2014	11.601
35	Jun 16, 2014	9.37
36	Jun 21, 2014	6.519
37	Jun 26, 2014	36.495
38	Jul 1, 2014	14.122
39	Jul 6, 2014	16.756
40	Jul 11, 2014	17.043
41	Jul 16, 2014	11.164
42	Jul 21, 2014	17.088

Hassan_Karnataka_p1

Fig. 4.2.5: Dataset Generated

4.2.6 Data Pre-processing/Preparation

Data pre-processing is a data mining technique which is used to transform the raw data in a useful and efficient format. It is mainly done to make the data consistent and ready for computation. Main steps in data pre-processing are:

1. Dataset Compilation: In this step all the individual datasets for precipitation, LST and elevation for all the different districts is compiled together into a single dataset for each crop

2. Data Cleaning: The data can have many irrelevant and missing parts. To handle this part, data cleaning is done. It involves handling of missing data, noisy data etc. In our case we had missing values as shown in the dataset obtained previously. To overcome this, we used the interpolate function of pandas which is a data-centric python package. Interpolate() was used to fill missing values within the data frame.

3. Data Consistency: The data being still inconsistent even after cleaning as the revisit period for MOD11A2.006 Terra Land Surface Temperature and Emissivity 8-Day Global 1km is 8 days and

CHIRPS Pentad: Climate Hazards Group InfraRed Precipitation with Station Data (version 2.0 final) is 6 days there was a difference in number of values. We used the moving averages technique to smoothen this inconsistency and obtain a consistent dataset.

4. **Creation of Target Variable:** Crop_class target variable has been created which holds values 0/1/2 depending on the crop

Our Machine Learning Model is trained to recognize various patterns over a set of training data, providing it an algorithm that it can use to reason over and learn from the phenological dataset. Once we have trained the model, we can use it to reason over data that it hasn't seen before and make sound predictions generating a compatibility score so that we can achieve our goal of mapping crops that are highly profitable but sparsely grown.

4.2.7 Machine Learning Model

Algorithm used- SVM Regressor

We have implemented SVM Regression for crop recommendation and calculating the probability of the suggested crop growing on a piece of land. In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyse data for classification and regression analysis. SVR is a powerful algorithm that allows us to choose how tolerant we are of errors, both through an acceptable error margin(ϵ) and through tuning our tolerance of falling outside that acceptable error rate. SVM has shown its great advantage in small sample learning, nonlinear classification and poor generalization ability and this is the reason it has offered precise and faster results compared to other algorithms.

The crops that have been chosen have similar training values of temperature, precipitation and elevation which was a challenge that we faced.

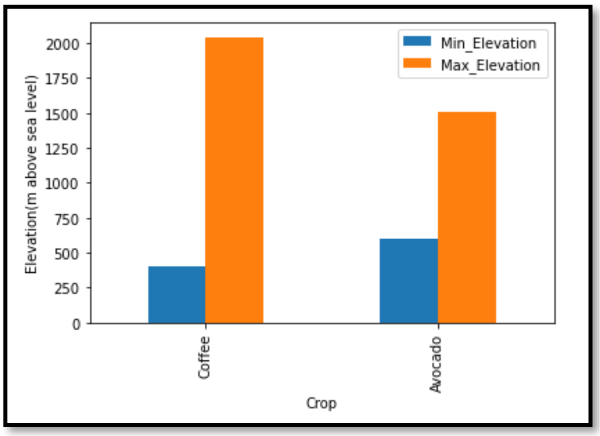
A complex RBF kernel is used in order to separate classes of coffee and avocado that are non-linear. As we have a non-linear dataset, by using the kernel trick we can make our non-linearly-separable data, linearly separable in a higher dimensional space. The trick does not actually transform the data points to a new, high dimensional feature space, explicitly. The kernel-SVM computes the decision boundary in terms of similarity measures in a high-dimensional feature space without actually doing the projection. The regularization parameter C that represents the misclassification/error term was reduced for better accuracy maximum error, ϵ (epsilon)

$$MIN\ \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{n}|\xi_i|$$

The SVCs aim to find the best hyperplane (also called decision boundary) that best separates (splits) a dataset into two classes/groups (binary classification problem). As we have a total of three features we have a 2D hyperplane. Mathematically, we can define the hyperplane as follows:

$$w_1x_1 + \ldots + w_dx_d + \beta_0 = 0,$$

where d is the number of features/ variables, xd is dth feature and wd as some weights, B0 is the bias term



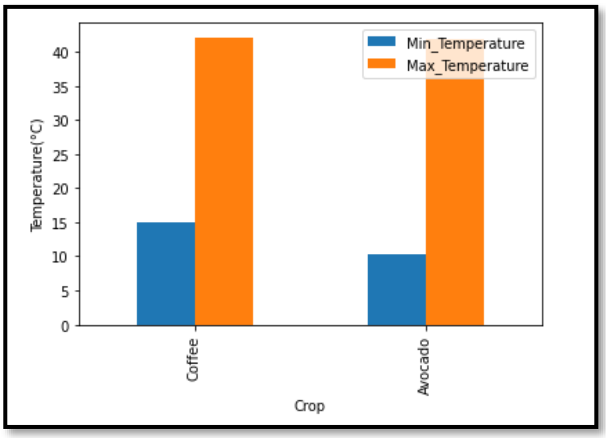


Fig. 4.2.7.1 Temperature ranges for Coffee and Avocado (from derived dataset)

Fig 4.2.7.2: Elevation ranges for Coffee and Avocado (from derived dataset)

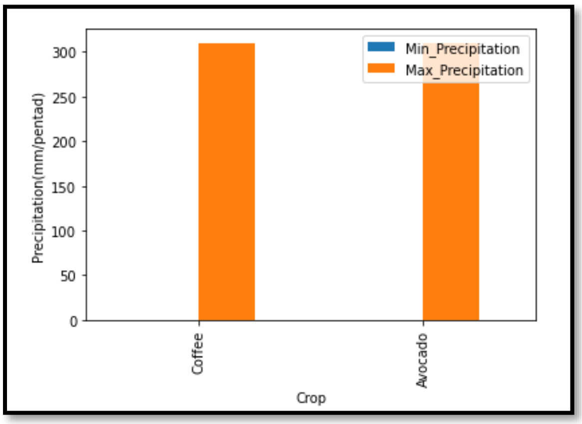


Fig. 4.2.7.3: Precipitation ranges for Coffee and Avocado (from derived datas

4.3 System 2- Market Price Prediction

Market prices of agricultural products are affected by many factors such as climate, supply and demand etc. The prediction is more complicated than commercial products. It is very difficult collect the data of impacting factors accurately and timely. This system aims at predicting the price of Coffee for a particular year, month and market that the farmer wishes to explore.

Dataset: Daily Market Prices Data of Coffee[18]

Attributes: State, District, Market, Commodity variety, arrival_date, min_price, max_price, model_price

Attributes chosen:

- Year
- Month
- Market

Markets:

- Kalpetta
- Sultan Bathry
- Pulpally
- Manathavady

Duration:

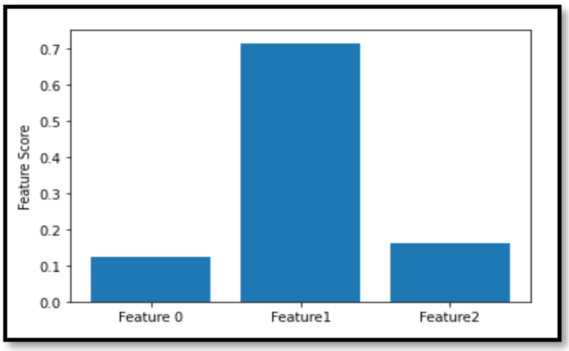
16 years(2003-2018)

4.3.1 Machine Learning Model

Algorithm selection: Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique which is based on the concept of ensemble learning, a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. Individual decisions operate as an ensemble and each individual tree in the random forest spits out a prediction value and the value with the most votes becomes the model's prediction. There is a direct relationship between the number of trees in the forest and the results it can get: the larger the number of trees, the more accurate the results. Random forest is also a very handy algorithm because the default hyperparameters it uses often produce a good prediction result. Understanding the hyperparameters is pretty straightforward, and there's also not that many of them.

One of the biggest problems in machine learning is overfitting, but most of the time this won't happen thanks to the random forest classifier. If there are enough trees in the forest, the classifier won't overfit the model. It's also easy to view the relative importance it assigns to the input features. Scores were calculated for each of the parameters to understand their importance in Regression.



Feature 0: Month

Feature 1: Year

Feature 2: Market

Fig.4.3.1: Features Score for Price Prediction calculated through Random Forest

The features of utmost importance were derived to be Year, Month and Market. These features were used to forecast prices for Coffee in Rs/quintal . The system will not only provide the user with the compatibility of the crop but also the price the farmer can expect if he plans to sell the crop at a particular time in a specific market.

A total of 1000 estimators have been used for the Random Forest Regressor

4.4 Dashboard

Fig 4.4.1 . About Us

Fig 4.4.1 is About Us section of the dashboard. It gives the user all the information about what we believe in and purpose behind creating the dashboard. On the left we have the navigation pane through which we can access different sections of the dashboard.

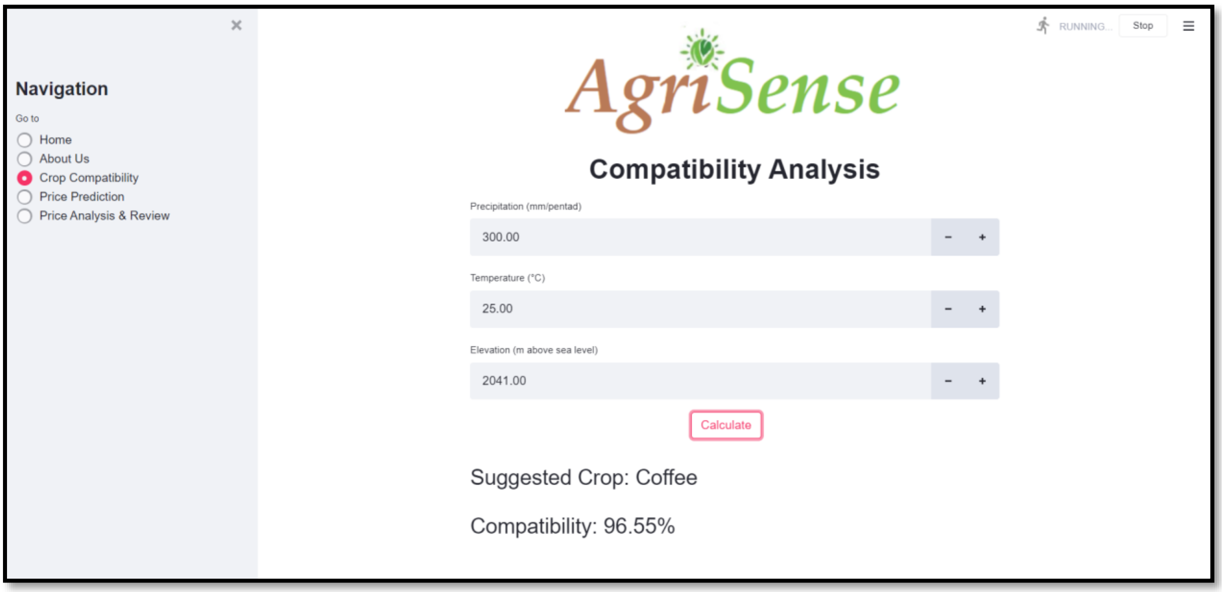


Fig 4.4.2. Compatibility Analysis Page

Fig 4.4.2 shows the Compatibility Analysis section of the dashboard. The user has to enter details like precipitation, temperature and elevation of the land and the region under consideration. The system give a compatibility percentage

Fig4.4.3. Price Prediction Page

Fig 4.4.3 is the price prediction section of the dashboard. The user has to input the year and month when he plans to harvest the crop and market in which he wants to sell it. The system will give a prediction of the price that the farmer can fetch for that crop at that particular time.

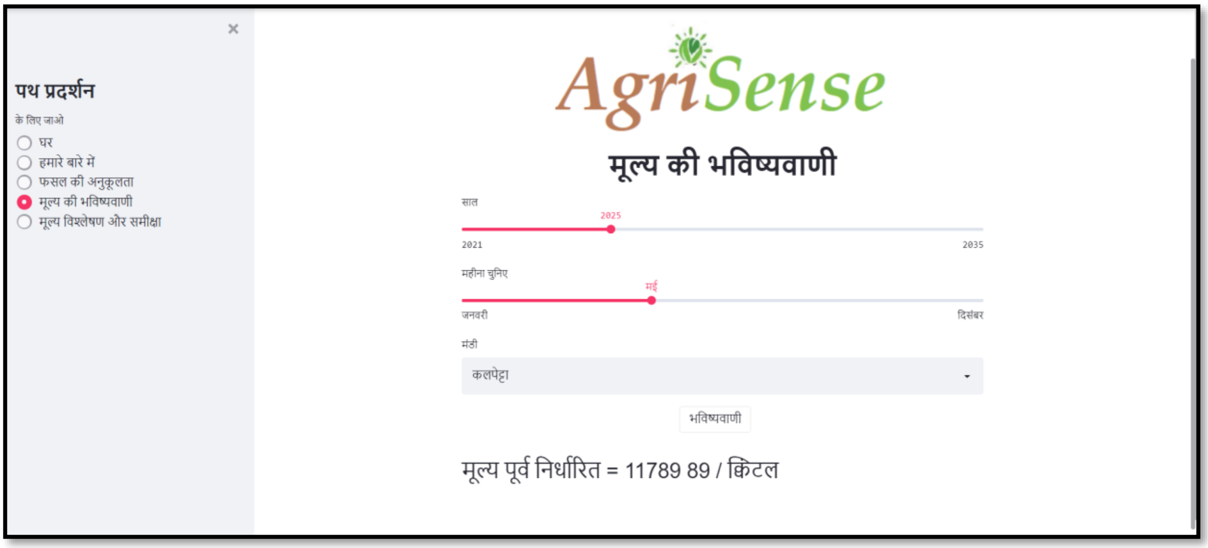


Fig 4.4.4. Price Prediction Page(In hindi)

Fig 4.4.4 shows the price prediction dashboard in Hindi for ease of use of farmers.
Integration of Vernacular languages. The Dashboard can be used in a variety of regional languages for the convenience of the user.

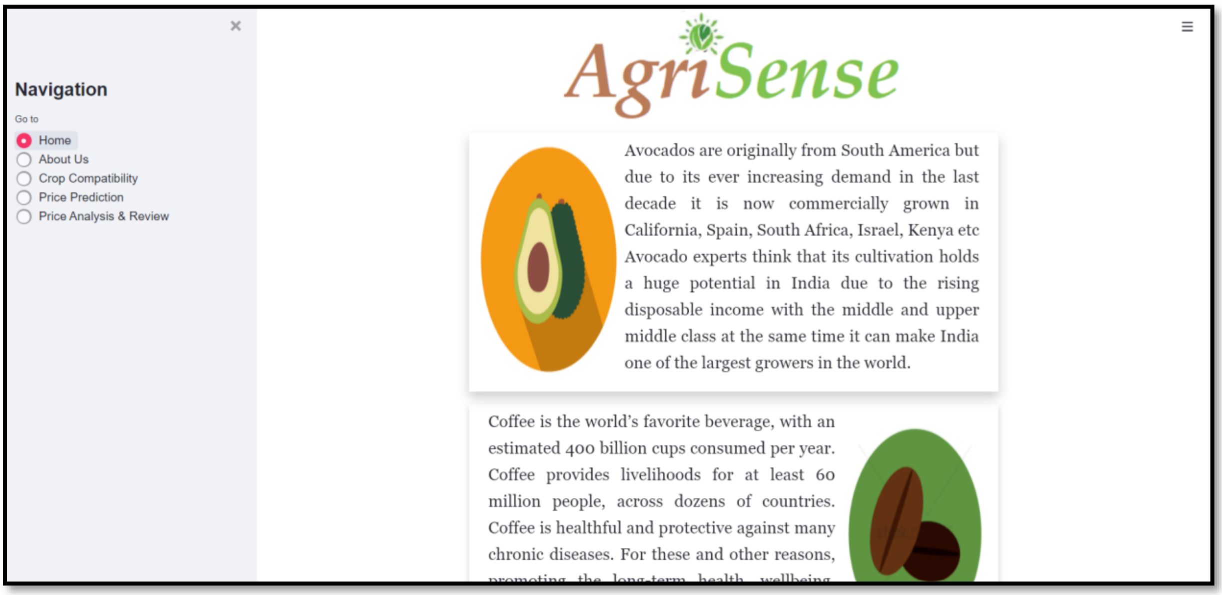


Fig 4.4.5. Home page

Fig 4.4.5 is the home section of the dashboard. It gives all the information about the crops under consideration like coffee and avocado and the regions they are grown in and the details about the market that is available to sell them.

Fig 4.4.6. Visualize and Review past data

Fig 4.4.6 is the Price analysis and review section of the report. It gives all the data that has been processed in an easy to view and understand form of graphs and charts. The user can view the complete data regarding the average modal price by year, month or the market.

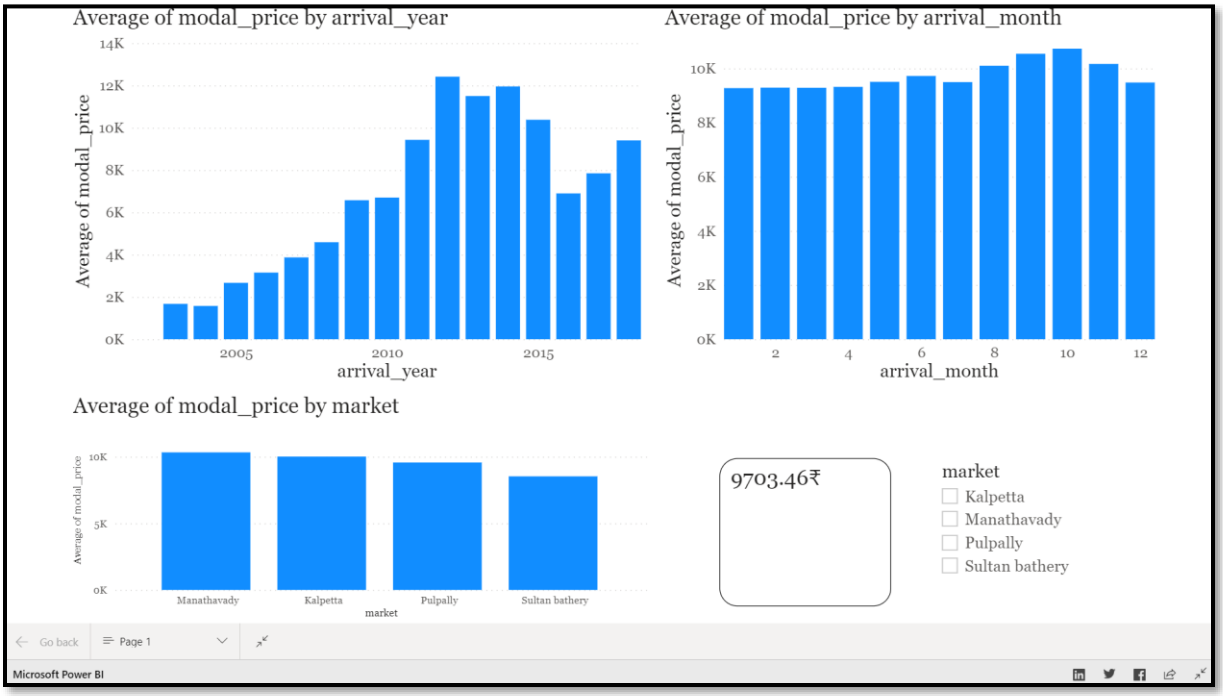


Fig 4.4.7 Visualize and Review past data

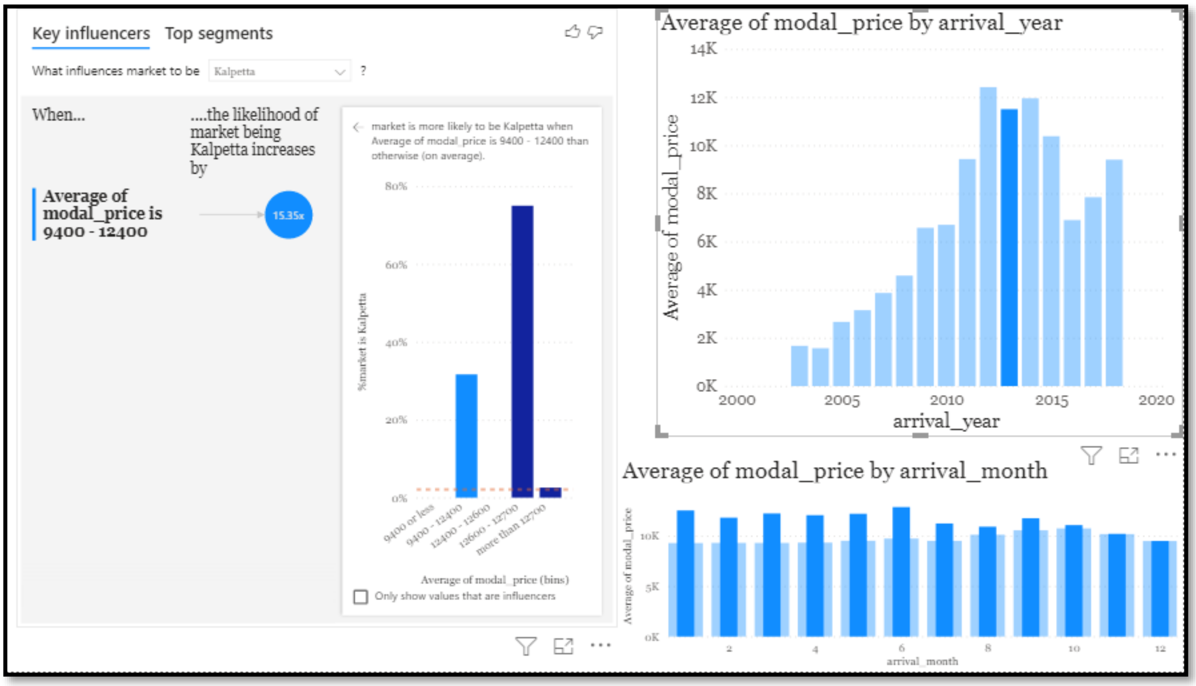


Fig 4.4.8. Visualize and Review past data

4.5 Results and Discussion

4.5.1 Analysis of results for Crop Compatibility

As of now, we have focussed on the two crops of Coffee and Avocado. Our system not only aims at providing the user with the suggestion of which of the crops to grow but also it gives the chances of how probable is the growth of that crop as compared to the other on the same piece of land. Our idea was to create a range of Temperature, Precipitation and Elevation which is more flexible than the ideal range of Coffee and Avocado which was the primary reason of using Remote Sensing to obtain climatic conditions of a region which is growing these crops all year round. This flexibility of values will give the farmers a greater possibility to explore the their piece of land with the suggested crop.

The classifications obtained through SVC using RBF kernel was approximately 79.14% accurate post which we have also provided with a score value that intends to quantify the compatibility of growing that crop. The model performance was evaluated using the following parameters:

Table 4.5.1: Model Performance of SVC

Metric	Model Performance
Accuracy	79.14%
MSE	0.338
RMSE	0.581
Precision	0.82
Recall	0.79

4.5.2 Analysis of results for Market Price Prediction

In addition to the crop suggestion, we wanted to provide a provision to the farmer/user where he/she can predict the market price of the crop for a year, month or a specific market. This functionality can help them analyse the kind of profit they might be able to make if they think of growing that crop. They can also review historic data of prices for a span of 16 years for 4 markets for better understanding. Random Forest Regressor(n_estimators=1000) was used to predict the prices. Overfitting was taken care of by Random Forest. Year followed by Market and Month were the most important features. The model performance was evaluated using the following parameters:

Table 4.5.2: Model Performance of Random Forest

Metric	**Model Performance**
Accuracy	96.01%
MSE	377746.44
RMSE	614.61
MAE	277.98
R^2	0.96

Chapter 5
Conclusion And Future Work

5.1 Conclusion

We have implemented the Region-Compatibility based Crop Recommendation System and Market Price Predictor" By studying the past research conducted and through this we gained thorough knowledge about the use of Google Earth Engine, how mapping could be done efficiently.. We further gained an understanding of the Machine learning algorithms used to develop a robust recommendation system.

We then did a comprehensive study on crops to understand the growing conditions of various crops to finalize our target crops. We studied various factors like temperature, rainfall, altitude, harvest time and elevation for a range of crops before selecting Coffee and Avocado.
We used various tools mentioned in the above chapters to prepare our database, pre-process it and give a visualisation.

We then studied the regions in which Coffee and Avocado are grown and how they have reacted to changes in factors such as precipitation, elevation and temperature over 12 years. We selected the target satellites for the remote sensing of this data. We then selected a particular region of interest for which we wanted to gather the data. Once we selected the desired ROI, we needed to extract data (precipitation, land-surface temperature and elevation) for that particular region. This was done by writing specific code for the particular parameter in the script box of GEE using either JavaScript or Python. After the successful execution of the codes we generated data for the phenological features. We collected the data for all the districts for each crop. After which we compiled and pre-processed data according to algorithmic needs.

We then trained our machine learning model to recognize various patterns over a set of training data, providing it an algorithm that it can use to reason over and learn from the phenological dataset. We selected the SVM Regression algorithm for crop classification. The accuracy of 79.14% was achieved by using a complex RBF kernel in order to separate classes of coffee and avocado that were non-linear.

Finally for the market price predictor we used the Random Forest algorithm. We used data for price of coffee and avocado sold in different markets throughout the country for over 16 years and were able to predict what is the expected price that a farmer would get for that commodity when sold in a market closest to him. We were able to achieve an accuracy of 96.01%

5.2 Future Work

- Dynamic remote sensing data: We plan to make use of dynamic data so that the user can get results based on the data which is collected in real time instead of the data which stored.
- We will be expanding the regions to cover the entire country: Currently we are only working on certain regions in the country to analyse and give results. In the future we want to able to be able to give results for any region under consideration.
- We will be adding data regarding more crops: We will be making this compatibility analysis and price prediction for more cash crops which are highly profitable and not just limit ourselves to coffee and avocado.
- Creating IOS and android app in Native languages of Farmers: We will be developing IOS and Android apps so that our platform is easily accessible to everyone. We have currently made part of our dashboard in Hindi but we plan on making the entire application in the native language of farmers so there is no language barrier.
- Using more parameters to get an accurate prediction: Currently we are only using limited parameters to give results like precipitation, temperature and elevation to give results but in the future we want to use more parameters like soil type and soil ph.
- Modelling seasonal crops: We also want to add data regarding seasons to get an accurate prediction regarding compatibility and price. This is because for a particular crop the yield changes massively depending on the season of harvest and sowing and the month and climatic conditions surrounding plantation.
- Consider unforeseen circumstances(Disasters/Calamities): Since this field heavily depends on nature we want to give an idea to the users regarding the effects an unseasonal rainfall might have on the yield or price. At the same time we want to give a probability of a calamity like a draught or flood depending on past data.
- Using real time data regarding prices from APMC markets: For Price analysis we want to be able to give real-time rates at which the crop is currently selling at any APMC market. This will help the user to get a better idea regarding the price prediction.

References

[1] A. Shelestov, M. Lavreniuk, N. Kussul, A. Novikov, and S. Skakun, "Large scale crop classification using Google earth engine platform," *Int. Geosci. Remote Sens. Symp.*, vol. 2017-July, pp. 3696–3699, 2017, doi: 10.1109/IGARSS.2017.8127801.

[2] O. Mutanga and L. Kumar, "Google earth engine applications," *Remote Sens.*, vol. 11, no. 5, pp. 11–14, 2019, doi: 10.3390/rs11050591.

[3] A. Mateo-Sanchis, M. Piles, J. Muñoz-Marí, J. E. Adsuara, A. Pérez-Suay, and G. Camps-Valls, "Synergistic integration of optical and microwave satellite data for crop yield estimation," *Remote Sens. Environ.*, vol. 234, no. October, p. 111460, 2019, doi: 10.1016/j.rse.2019.111460.

[4] M. Sabini, G. Rusak, and B. Ross, "Understanding Satellite-Imagery-Based Crop Yield Predictions," pp. 2–3, 2017, [Online]. Available: http://cs231n.stanford.edu/reports/2017/pdfs/555.pdf.

[5] X. Liu *et al.*, "Large-Scale Crop Mapping from Multisource Remote Sensing Images in Google Earth Engine," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 414–427, 2020, doi: 10.1109/JSTARS.2019.2963539.

[6] P. Rufin *et al.*, "Mapping cropping practices on a national scale using intra-annual Landsat time series binning," *Remote Sens.*, vol. 11, no. 3, pp. 1–26, 2019, doi: 10.3390/rs11030232.

[7] "Crops Research," [Online]. Available: https://www.agrifarming.in.

[8] S. Dimitriadis and C. Goumopoulos, "Applying machine learning to extract new knowledge in precision agriculture applications," *Proc. - 12th Pan-Hellenic Conf. Informatics, PCI 2008*, pp. 100–104, 2008, doi: 10.1109/PCI.2008.30.

[9] R. Katarya, A. Raturi, A. Mehndiratta, and A. Thapper, "Impact of Machine Learning Techniques in Precision Agriculture," *Proc. 3rd Int. Conf. Emerg. Technol. Comput. Eng. Mach. Learn. Internet Things, ICETCE 2020*, no. February, pp. 18–23, 2020, doi: 10.1109/ICETCE48199.2020.9091741.

[10] S. Sharma, G. Rathee, and H. Saini, “Big data analytics for crop prediction mode using optimization technique,” *PDGC 2018 - 2018 5th Int. Conf. Parallel, Distrib. Grid Comput.*, pp. 760–764, 2018, doi: 10.1109/PDGC.2018.8746001.

[11] J. Treboux and D. Genoud, “Improved machine learning methodology for high precision agriculture,” *2018 Glob. Internet Things Summit, GIoTS 2018*, pp. 0–5, 2018, doi: 10.1109/GIOTS.2018.8534558.

[12] Z. Doshi, S. Nadkarni, R. Agrawal, and N. Shah, “AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms,” *Proc. - 2018 4th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2018*, pp. 1–6, 2018, doi: 10.1109/ICCUBEA.2018.8697349.

[13] K. V. Divya, A. Jatti, P. R. Joshi, and S. D. Krishna, *Progress in Advanced Computing and Intelligent Engineering*, vol. 714. Springer Singapore, 2019.

[14] J. Padarian, B. Minasny, and A. B. McBratney, “Using Google’s cloud-based platform for digital soil mapping,” *Comput. Geosci.*, vol. 83, pp. 80–88, 2015, doi: 10.1016/j.cageo.2015.06.023.

[15] J. Mei, D. He, R. Harley, T. Habetler, and G. Qu, “A random forest method for real-time price forecasting in New York electricity market,” *IEEE Power Energy Soc. Gen. Meet.*, vol. 2014-October, no. October, 2014, doi: 10.1109/PESGM.2014.6939932.

[16] F. Optimization, “Forecasting Daily Stock Trends Using Random,” pp. 1152–1155, 2019.

[17] C. Luo, Q. Wei, L. Zhou, J. Zhang, and S. Sun, “Prediction of vegetable price based on neural network and genetic algorithm,” *IFIP Adv. Inf. Commun. Technol.*, vol. 346 AICT, no. PART 3, pp. 672–681, 2011, doi: 10.1007/978-3-642-18354-6_79.

[18] “Coffee Price Data,” [Online]. Available: https://data.gov.in/catalog/variety-wise-daily-market-prices-data-coffee?filters%5Bfield_catalog_reference%5D=92731&format=json&offset=0&limit=6&sort%5Bcreated%5D=desc.

Acknowledgement

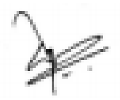
There are many individuals who have helped us in planning and executing this project by giving us innovative ideas, motivations and opportunities.

We would like to express our gratitude towards our mentor, Professor Kamal Mistry who guided and supported us throughout the course of this project. Her invaluable insights and guidance provided us with constant support, motivation and the ability to improve upon the project in every way possible. Her push for excellence has helped us achieve the goals and milestones that the project was planned with.

We would also like to extend our gratitude to the Head of Department of Computer Engineering, Dr. Pravin Srinath and Dean of Mukesh Patel School of Technology and Management for providing us with the opportunity and the resources to pursue this project and accomplish our goals.

We would like to thank our project coordinators and faculty members who helped us by scheduling all the project presentations and by providing us with constant feedback which helped us improve this project at every step of the way.

Signature:

Vidhi Agrawal	Darpan Bendre	Sadhika Bhatia	Darshan Daliya
B003	B008	B010	B018