

FIT5145
INTRODUCTION TO DATA SCIENCE

ANALYSIS ON RICE CROP GROWTH IN INDIA



Semester 2, 2019
Assignment 4 - Business and Data Case Study
Lecturer - Mahsa Salehi
Tutor - Mohsen Laali
Student - Dishi Jain
Student ID -30759307

Table of Contents

1.	Project Description.....	2
1.1	<i>Background</i>	2
1.2	<i>The Project</i>	2
2.	The Business Model.....	3
3.	Characterising Data and Data Processing.....	4
4.	Data Resources.....	6
5.	Data Analysis.....	6
6.	References.....	9

1. PROJECT DESCRIPTION

1.1 Background

Agriculture in India represents an important sector contributing to about 17% of the Indian GDP. As per 2018 reports, agriculture employed around 50% of the Indian workforce². Every individual in the nation is directly or indirectly affected by agriculture. Farmers who are looked upon as the backbone of the Indian economy, are still found to worry about their crops, profits, sales, debts and various other problems throughout their lives. There are many problems that a farmer has to face which the urban people do not even realise. Because of the difficulties, it is no surprise, that even the children of the farmers want to migrate to the cities and leave farming. We need to understand the extent of this issue otherwise a day will come where “food surplus” will convert to “food deficiency” in India.

Rice is considered as one of the main grains grown in India. The percentage growth of rice is about 39% over a decade. However there are many issues related to rice production and growth. Around 78% of the farmers do not have the right resources and knowledge. Often the crop suffers due to inconsistent soil moisture, less nutrient value and the use of old farming techniques¹. The issues also arise due to the lack of education. A huge percentage of Indian farmers do not attend schools and their children are seen to do the same. Due to the lack of education, the farmers are not able to solve their problems and make use of the new-age technology to the best fit.

Climatic conditions is another major difficulty that haunts every farmer day in and day out. For the farmers who want to build up their yields, climate change always acts as a gamble. Agricultural scientist,M S Swaminathan have already pointed out that just a mere 1 degree celsius rise in temperature can yield to a loss of 6 million tonnes per year². Performing analysis and visualisations on the climatic conditions during the different times of the year can help in the betterment of the crop supply.

Another major issue that is often overlooked is the storage management of the crop and the condition of the infrastructure around the agricultural land. Roughly around 15% of the yield is lost between the farm gate and the consumer³. This can be reduced if better roads are constructed. This will also allow a smooth working atmosphere for the farmers and the commuters.This can further augment the agricultural output and income for rural farming community.

1.2 The Project

This project aims to provide solutions and a business model to solve the issues seen in the cultivation of the rice crop in the farming community. Use of technology and data for carrying out analysis can help us to reduce the difficulties and ensure maximum productivity. Analysis can be carried out on the climatic conditions and the type of crop to grow at a period of time to secure the efforts put in by the farmers. Through the data collected over the years we can

also find out how lack of education impacts the agriculture and ways in which we can correct it.

2. THE BUSINESS MODEL

Data Science can play an important role in finding the best possible solutions for the problems faced by the farmers in today's world. Abundant amount of data has been collected which tells us how even a single decision while farming can cause a huge impact. What kind of soil to pick, which crop to grow, how much irrigation is required, what pesticides are less harmful, which climatic conditions yield best results and prediction of weather change can all be answered by looking at the history and the previous data. This project aims to provide a business model and solutions to such questions based on analysis and prediction.

Farmers around the country are beginning to adopt precision agriculture and agriculture technology through machines and equipment that use data analytics, Internet of Things and robotics to optimize the inputs and enhance the yields. Precision agriculture involves working on data about the crop type, climate, etc and then carrying out analysis on it to find major insights that will help the farmers. Based on the insights better decisions can be taken to increase the crop produce and the business. The goal of Precision Agriculture remains to optimize the profits while making a minimum use of resources. Precision agriculture can help predicting the climatic conditions and also estimate the potential losses. Accordingly the farmers can make adjustments to ensure minimum loss.

Precision technology is not only applied during pre-harvest but can also be applied to the post-harvest activities. Grading the quality of the produce and supply chain management can ensure a better productivity. It can allow the farmers to enhance the marketing of their products at a particular time of the year to a particular audience.

With the help of precision technology and the insights gathered from it we can assist farmers to make better decisions. The farmers can be educated regarding which crop to grow at what time to get the maximum profit. They can be taught about the climatic conditions and how majorly it impacts the growth of the crop.

The data for precision agriculture can be collected through various means. Sensors can be put into soils to collect and store the nutrient contents of the soil during different times of the year. A record of the pesticides used on different crops during different times can also provide us with good insights. The chlorophyll percentage in a crop can also be determined using a chlorophyll meter. This data can further help the farmers to pay more attention to the crops that have a low chlorophyll level. Climatic conditions can be monitored using tools like weather stations that measure the different variables like temperature, humidity, wind speed, rainfall, etc⁴.

Based on the data we can analyse and find patterns that will help in the better decision making for the farming sector. This will help in better crop productivity and will help farmers use their resources to the best use.

This business model will focus on analysing data to solve some major issues in rice cultivation. Rice is a type of a crop that requires varying temperatures to grow well. The temperature should be around 20-20 degrees celsius at the time of sowing, 22-25 degrees celsius at the time of growth and finally 25-30 degrees celsius during harvesting. To maintain these temperature conditions and to see how slight changes in temperature affect the crop, this business model will come into use.

3. CHARACTERISING DATA AND DATA PROCESSING

The data collection, processing, modelling and visualisation are the important factors of any data science project. This project will look forward for the following sections -

Data Collection - Collecting the data from the different sources using technologies like image processing and smart remote sensing.

Data Wrangling or cleaning - Cleaning the data and preparing it for further analysis.

Data analytics and modelling - Prototyping and preparing a basic model after analysis on the data.

Engineering - Data management, storage and maintenance.

Governance - Overall governance and working of the project to make it deliverable.

Operationalisation - Presenting the model to the company for value realisation.

The data for creating the model is characterised into different sections. It includes information about the velocity, variety, veracity and volume of the data. It includes the sources from where the data will be extracted. A detailed information about it is given in Table 1 below.

Table 1. Data Characteristics and Processing

Data Sources	Data can be collected via image processing, using devices like farm machinery and drone imagery, and information regarding crop field from devices inserted into the field.
Data Volume	The agriculture sector has shown increased volume of rice production. Rice is grown almost 40% of the total grain production in India. Analysis and data processing is hence considered as important factors in the field of rice.
Data Velocity	Rice production witnessed an increase of about 5.25% from 2010 to 2013. This pattern of increase is seen today too.

Data Variety	Multiple data sources give us multiple variety of data. This includes manual data about crop's properties, images from drones, values gathered about crop quality from sensors inserted into soil and values about chlorophyll level using a chlorophyll meter.
Data Veracity	There are many default/incorrect data sources online that provide us with incomplete and inconsistent information, which need to be cleaned.
Softwares	Softwares and ML techniques that are required to do the analysis include HDFS for storing of data, Python and R for the manipulation and cleaning of data and Spark.
Analytics	Analysis on data is to be carried out to using the best tools to determine which technique is most useful for prediction of crop production.
Processing	This includes the availability of skills to process the data. Understanding of languages like R, Python and libraries like matplotlib, etc is required for the processing of the data.
Capabilities	Major key requirements include the quality of crop, weather conditions, type of fertilizer being used and education amongst the farmers.
Security	Protection of the data is required. Checking to see that no manipulations are being done to the data.
Lifecycle	Continued analysis and increasing production of rice
Others	Analysis on the factors like knowledge of the farmers, their understanding of the analysis and strengthening their decision making power.

There are mainly three kinds of analyst/ professionals who will be involved in the business model.

- Data Analyst

They will primarily be responsible for gathering insights from the data. They will also focus on developing relationships between the variables in the data and find patterns within the dataset.

- Data Scientist

These professionals will mainly focus on developing models and products that in turn produce insights from the data. They will focus on developing models that will solve business problems.

- Data Engineers

These are the people that will primarily focus on managing the data infrastructure, automating the data processing and also deploying the data models. They will engineer the entire process and make sure the business model works efficiently.

4. DATA RESOURCES

There are different sources from where the data can be gathered for the analysis. Some of the sources include -

- Image processing using images from drones
 - We can use the new available technologies to increase the input in crop production. Using image processing many characteristics regarding the soil, crop, fertilizers, water irrigation level can be determined. This helps humans to not only analyse the visible images but also the invisible images such as /ultraviolet(UV), Near Infrared (NI), and Infrared (IR) rays. Images taken of the field at different times can be used as the data and can undergo image processing and analysis.
- Sensors inside soil to measure soil moisture value
 - Measuring the value of water in the soil at the root level is very important. It determines what kind of crops will be grown and increases profitability. The sensors should be planted in the soil correctly. This will store data like water level inside soil and fertilizer percentage absorbed in the soil
- Temperature Measurements using thermometers and weather stations
 - Rice is a crop that requires varying temperature for the proper growth. How and when changes in temperature could cause changes in the growth rate is hence an important concern. Proper measurements about the temperature can be done using tools like thermometers and getting data from weather stations. Automatic weather stations are used to measure a set of variables like humidity, wind speed, incoming solar radiation and rainfall.
- Remote Sensing-
 - Remote sensing techniques can be used to determine the different features of the earth's surface. It also helps to estimate the geo-biophysical traits with the help of electromagnetic radiation.

5. DATA ANALYSIS

The analysis is carried out on the data collected from the different sources after cleaning and wrangling. Analysis contributes to be an important factor as it forms the basis of any predictions or decision making. Analysis can be carried out using the techniques of Machine Learning which can help in determining what are the best conditions to grow rice in, what temperature is most suitable, what fertilizers to use and how much irrigation level is required. Good analysis results in good productivity of the crop and hence helps the farmers in decision making.

Using the data analysis techniques like image processing, linear regression and influence diagrams we can analyse and visualise the data to predict the best suitable conditions in

which we can grow rice in the agricultural fields. The data collected from weather stations give us an estimate of when to sow, grow, fertilize and harvest the crop. It helps to achieve the most productivity. It also educates the farmers to utilize their time and efforts in the best possible way.

Linear regression can also be used to plot the data values. This helps in understanding the condition of the crop in a more visual manner. Delivering the final presentation to the farmers hence become stronger and they will be able to relate more to the graphs and numbers rather than theoretical concepts.

Analysing the data -

The dataset for the analysis is extracted from the following link. [Click here.](#)

A sample of the dataset is given below -

Figure 1. Sample of dataset

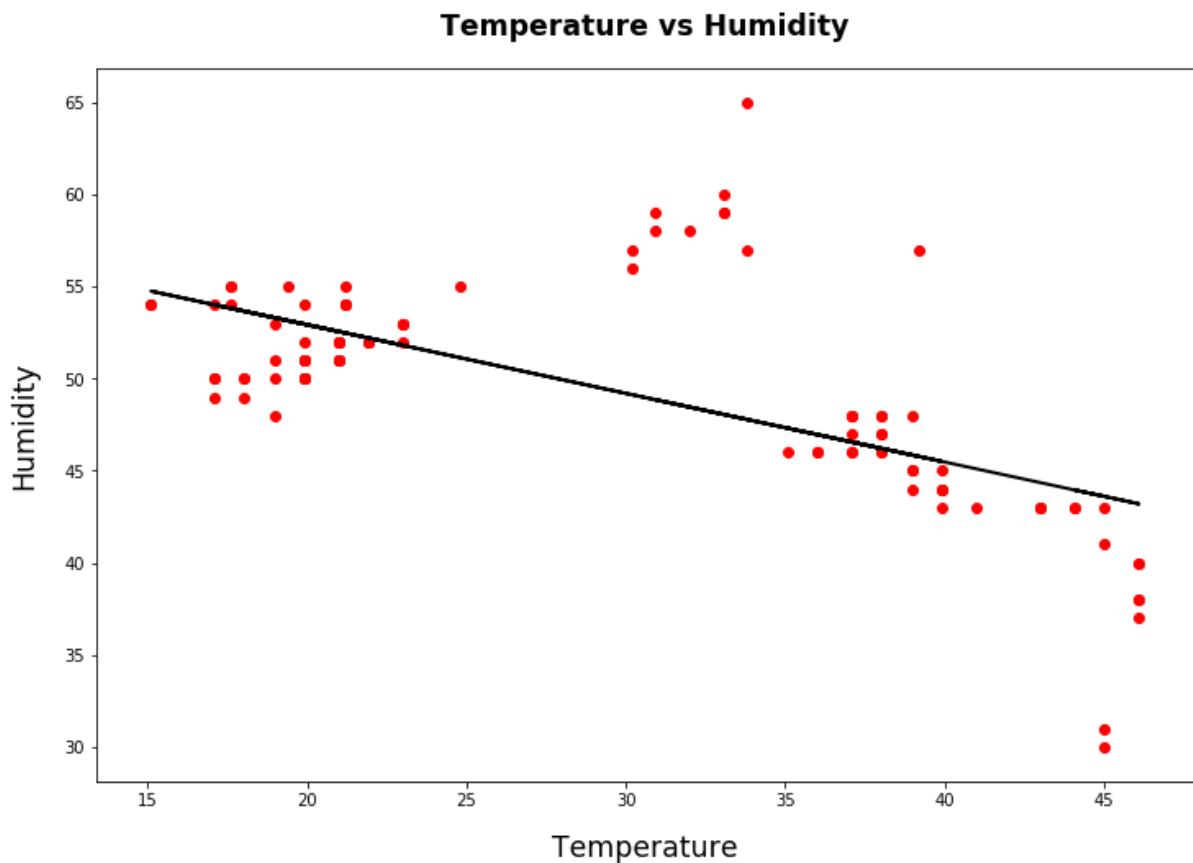
Temperature in Celsius	Soil Moisture Content	Humidity	Crop_Year	Area	Cost of Cultivation	Revenue Generated	Yield
45	42.1	30	2018	24574	1941.55	23076.74	91.59%
45	31	31	2018	25978	2172.46	1652.68	-31.45%
46.1	41	37	2018	25211	1898.3	1955.9	2.94%
46.1	81	38	2018	24097	3670.54	341765	98.93%
46.1	98	38	2018	17523	2975.8	2827.26	-5.25%
46.1	79.9	40	2018	24437	3539.47	3311.82	-6.87%
46.1	67	40	2018	18703	2003.76	5082.83	60.58%
45	95	41	2018	15034	4909.99	4475.72	-9.70%
44.1	76	43	2018	16875	2179.26	4207.44	48.20%

From the dataset we can derive relationships between the Year of growth, Temperature condition, Soil Moisture content, Humidity level and various other factors.

The major reason that was found in the wastage of rice was due to the temperature and humidity issues. The knowledge amongst the farmers about when to grow, cultivate and harvest the crop was missing. This lead to a lot of wastage and damage to the crop during these stages. Hence using the business model an analysis was carried out between the variables Temperature and Humidity.

After analysing and plotting the data using Linear regression as an ML technique, it was found that the best temperature conditions for rice to grow in are from 20-30 degree celsius. The humidity level required for the same crop to grow efficiently was above 50%. Hence sowing, growing and harvesting the crop at these temperature and humidity level conditions will give the max profitability.

Figure 2 - Linear Regression relationship between Temperature and humidity



With the help of image processing it was also found that the best month to grow rice in was during the month of June. During the months of June to October the images taken showed the maximum height of the crop, the best quality and the most soil nutrient content. During the months of March-April the crop was seen to be of a degrading quality with less nutrient value. Hence these months should be avoided. The above time was also calculated by using images from drones and sensors that were fitted into the soil.

The business model hence helps to determine the most suitable situations to grow rice in India. This can help in the reduction of wastage as well as reduce the unnecessary human efforts. Using the business model can help to determine the best climatic conditions and the best farm field condition to achieve the best results.

6. REFERENCES

1. <http://drdpat.bih.nic.in>Status%20Paper%20-%2009.htm>
2. https://en.wikipedia.org/wiki/Precision_agriculture
3. <http://imotforum.com/2017/11/biggest-problems-faced-farmers-india/>
4. <https://pdfs.semanticscholar.org/c181/1c5052740fe3213e4773fb36993ca2b508af.pdf>
5. <https://www.eo4idi.eu/eo4sd-knowledge-portal/6-field-data-collection/61-field-level-surveys/62-measurement-equipment/622>