COSC2669:
Case Studies in Data Science

Final Report

Group 21- Data Warriors

Swati Arora	s3796848
Vaishnavi Narayana Naik	s3797442
Aditi Sharma	s3795604
Maaz Shaikh	s3795603

Table of Contents

Introduction

From the overall area of 3,66,470 hectares of a sugarcane field, around 4.5 million tonnes of raw sugar are produced annually by crushing 30 million tonnes of sugarcane. Approximately 95% of the Australian raw sugar is produced in Queensland and around 80% is exported via various bulk storage ports. Sugarcane farming supports $1.1billion in economic activity every year. It provides over 9800 jobs and wages leading up to $379 million. History says that originally crops were brought in New South Wales where it failed due to weather conditions, paving the way for Queensland to become the cane growing powerhouse that it is today. Australia is one of the biggest suppliers of sugarcane in the world and Queensland is one of the highest sugarcane producing states for Australia. Queensland has 5 major cane growing regions. [2]

Sugarcane is giant tropical grass which takes between 9 to 18 months to reach the maturity level depending on the weather and multiple other conditions. The factors that are crucial for the successful growth of sugarcane crops are immense amounts of water (every year 1.5 metres of rainfall or accessibility to irrigation), intense sunlight, fertilised soil, fertilisers, and required chemicals for weed and often pest and control. Sugarcane has various growth stages with different requirements for each stage.[3]
Various stages are:

1. Germination and establishment phase
2. Tillering phase
3. Grand growth period
4. Maturation and Ripening phase

In the sunny weather of Queensland, usually, the sugarcane crop is yielded within this duration. Once harvested, to avoid sugarcane deterioration, sugarcane should be transported to the sugar mill within16 hours of being harvested. Also, to cut the cane, heavy-duty machines called cane-harvesters are used. Sugarcane harvesting in Queensland starts in May and ends by the mid-December.

Problem Statement

Importance of monitoring sugarcane fields

Sugarcane crops require regular monitoring and control during its various stages of growth. First, 60 days of growth is the germination and establishment phase. The buds in the germination process are sensitive to the temperature and are likely to grow at a slower rate when the temperature drops below 17 to 18 degree Celsius and increases if the temperature rises above 35 degree celsius. The amount of moisture present in the soil is directly proportional to the temperature outside.

According to Whitman et al. (1963) for a sugarcane crop in Australia, the ideal temperature is 30 degree Celsius, 16-22 degree celsius is considered to be significantly sustainable temperature and below that would be nil.[4]

Study shows that during the spring-summer season, the surface of the soil is intensively exposed to the sun. 60-150 days is the tillering and canopy development phase. Further phases are the grand growth phase and Maturing and ripening phase that takes place between 150- 360 days.

Nitrogen fertilizers have a large impact on the soil's pH value, and it can make the soil more acidic or more alkaline depending on the type of nitrogen fertilizer used. It is really important to monitor agriculture standards for sugarcane during the entire process which includes monitoring the major key factors such as soil pH, (Moisture Index) NDMI.[1]

Proposed solution

Overview

To make a farmer's life easier, we provide a solution which lets them monitor their sugarcane fields. The production of sugarcane crops depends on several factors like Moisture index (NDMI) and Soil pH value. These factors structure the overall health of the crop. For a farmer, it is suggested to understand these values before making decisions for the betterment of the sugarcane crop. Prior knowledge of likely climate and other key factors shared above can help the farmers prepare well in advance about the production, harvest and marketing efforts. Studying the crop health and the right harvesting time can help other stakeholders and decision-makers to prepare well in advance about the storage requirements, shipping schedules and customer allocations.
Our solution will provide information about the Vegetation index (NDVI) of the sugarcane crop that will indicate the crop health to the farmers. We also examine the Moisture index and pH value of soil suitable for the growth of the sugarcane.
For our research, a small area of Queensland was chosen. And the data was provided by Landsat 8. Landsat 8 is a satellite, recently launched, used for collecting data and images for agriculture purpose, education, business and government. Different bands provide various information.

Monitoring the health of the crops

To get the Vegetation Index from the Landsat satellite, we use the NDVI indicator. NDVI (Normalized Difference Vegetation Index) is a basic calculation from the different bands in the Landsat 8 satellite images that represents vegetation health and determines its overall condition. This process is done remotely without any hassle of visiting the fields and gathering the soil sample to monitor.

Allowing comparison

With the help of Vegetation Index, farmers can compare his field with the other fields and educate themselves about the methods and procedures used by other farmers to improve their crop's health. The standard way of measuring the vegetation index includes taking the soil sample from the field which will undergo the process of finding the vegetation index. It is time-consuming and needs more man-hours to complete the whole task.
Furthermore, the Vegetation Index can be beneficial to the government as it enables them to accurately compare different regions to identify which fields require additional assistance to increase the overall production.

Monitoring soil composition

To determine the moisture content of the soil, we use the normalised difference Moisture index. NDMI is used to calculate the moisture index of the soil. Moisture index is the portion of the total precipitation used to satisfy plant (vegetation) needs. Low moisture index can hinder the growth of the crop, hence it is important to monitor the moisture content.[1]
Soil pH is also an important factor and pH values in the range 5.5 - 7.5 are recommended for sugarcane production. Soil acidification is a natural process which is accelerated by nitrogen fertilisers and excessive acidification can cease the growth of the crop. Hence, It is important to monitor the soil pH as well.

Stakeholder Analysis

Critical stakeholders that have vital roles in sugarcane production. As Farmers are connected to sugarcane production from stage one to appropriately monitor and evaluate the health of the crop, the government has the potential to invest funds and technical help in the right direction by helping the overall industry. Some of the highlighted and discussed stakeholders and their responsibilities are mentioned below :

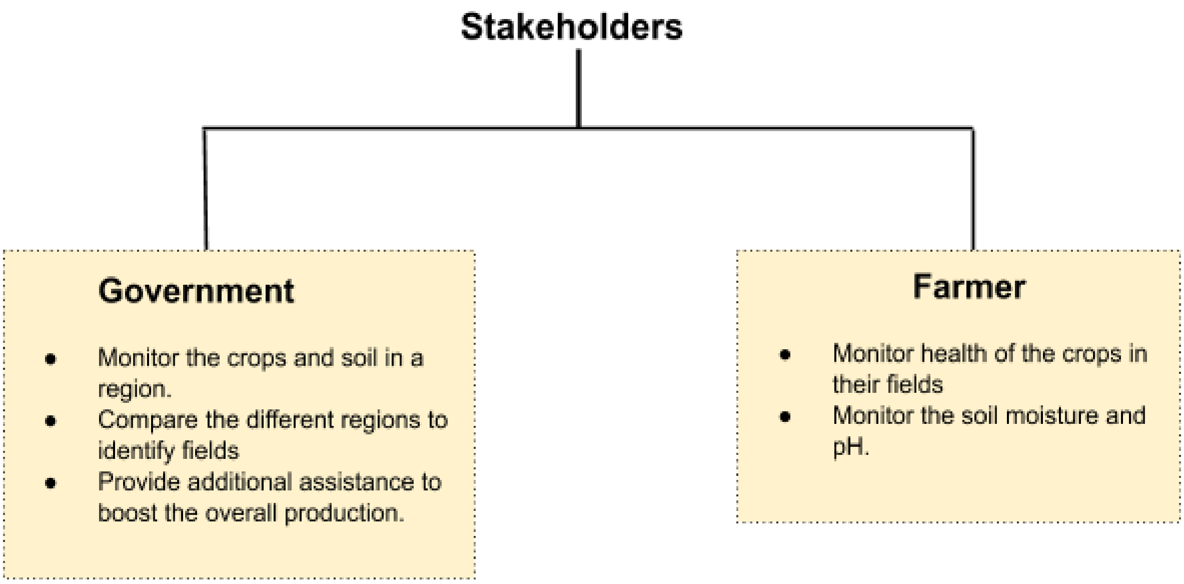

Fig 1. Stakeholder analysis

Methodology

Tools & Resources

Datasets

USGS Landsat 8 Collection 1 Tier 1 TOA Reflectance [5]

We use the Landsat-8 satellite images accessed from the Google Earth Engine data catalog. Of the different Landsat-8 datasets provided we use the calibrated top-of-atmosphere (TOA) reflectance dataset. Google Earth Engine API stores the images in the format ImageCollection which is a stack or sequence of images[6]. We use these ImageCollection objects to filter the Landsat-8 data catalog to generate the NDVI, NDMI and pH layer for the area of interest.

Property Boundaries

To aid the user in selecting their plots for analysis, we obtained the property boundaries for a few cities in the Whitsunday Region of Queensland from Queensland Spatial Catalogue[9]. The spatial data obtained is then overlaid on top of the map interface. This overlay process can be carried out for any region required by downloading the data from the catalogue.

Google Earth Engine API

Google Earth Engine is a geospatial processing service. With Earth Engine, you can perform geospatial processing at scale, powered by Google Cloud Platform[7]. To develop this project, we used the Earth Engine Python API[8] along with jupyter notebook and IPyleaflet.
Without the API, we would be required to manually download the data from Copernicus SciHub for a region or area we might evaluate. Integrating the latest satellite images becomes challenging, requiring repeated download manually whenever new data is available.
With Google Earth Engine API, the program automatically accesses the latest images and earth engine allows us to filter the dataset by date which helps automate the whole process, making it scalable. This way we can extend our project to analyse any area around the world.

IpyLeaflet

Ipyleaflet is a Jupyter / Leaflet bridge enabling interactive maps in the Jupyter notebook. We use IpyLeaflet to display Google Maps and add interactive components to it. The interactive components are shown below:

1. Users can draw polygons, rectangles, circles to select the area to analyse.

Fig 2. Polygon Drawing

2. The Layers Icon on the right allows users to toggle the various layers and the popups generated.

Fig 3. Layers

Cloud Masking

One of the challenges faced by remote sensing applications is the cloud cover in satellite images. To be able to provide an accurate analysis, we need to mask the cloud cover pixels in the images.
For scoring Landsat pixels by their relative cloudiness, Earth Engine provides a rudimentary cloud scoring algorithm in the ee.Algorithms.Landsat.simpleCloudScore() method. Using this method, we compute a mask of the cloudy pixels taking 20 as a threshold (cloud score is given in the range [0, 100]. Using this mask we update each image in the ImageCollection and generate a cloud masked image.
The following code snippet shows how we perform cloud masking when generating the NDVI layer.

```
# Calculate the NDVI for the given image while performing cloud masking.
def cloudlessNDVI(image):
    # Get a cloud score in [0, 100].
    cloud = ee.Algorithms.Landsat.simpleCloudScore(image).select('cloud');

    # Create a mask of cloudy pixels from an arbitrary threshold.
    mask = cloud.lte(20);

    # Compute NDVI.
    ndvi = image.normalizedDifference(['B5', 'B4']).rename('NDVI');

    # Return the masked image with an NDVI band.
    return image.addBands(ndvi).updateMask(mask);
```

NDVI

The Normalised Difference Vegetation Index is a basic calculation that represents the general vegetation health. From the Landsat-8 satellite, we use the bands 4(Red) and band-5(Near Infrared) to calculate the NDVI[10].

NDVI = (NIR - Red) / (NIR + Red)

The returned value from the equation ranges from -1 to 1, where values closer to 1 indicate healthy and ready to harvest plants, and closer to 0 represent dead or no plants planted.

NDMI

The Normalised Difference Moisture Index is used to calculate the moisture content in the fields. We use bands 5(Near-infrared) and band 6(SWIR-1) from Landsat-8 to calculate the NDMI[11]:

NDMI = (NIR - SWIR) / (NIR + SWIR)

The changes in vegetation water content and the spongy mesophyll structure are reflected by the SWIR band, while leaf internal structure and leaf dry matter content is reflected by the NIR band (https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/ndmi/)

Soil pH

To calculate the acidic-alkaline content present in the soil, we look at the pH indicator. We use bands 2(Blue) and band-6(SWIR-1) in a simple equation to get the pH value[12].

```
pH = 7.84 + (0.48 * B2) - (0.12 * B6)
```

Values less than 7 indicate an acidic soil state, while higher than 7 indicates a more alkaline soil state content.

Visualization

The WebApp displays a google map with the property boundaries filter overlaid over the map. With the polygon tool, users can select a field of their choice and create a polygon along the property boundary.

Fig 4. Plot creation

Once the polygon is created, the application will calculate the NDVI, NDMI and soil pH values and display a pop up over the polygon. The user can create as many polygons as they prefer. There is a time slider for the user to utilize based on which the NDVI, NDMI and soil pH values are shown for that date.

Fig 5. Multiple plots

The application also generates a time-series graph of the NDVI, NDMI and soil pH values from Jan 2014 to the latest date of the available data. The chart can also be downloaded in PNG format.

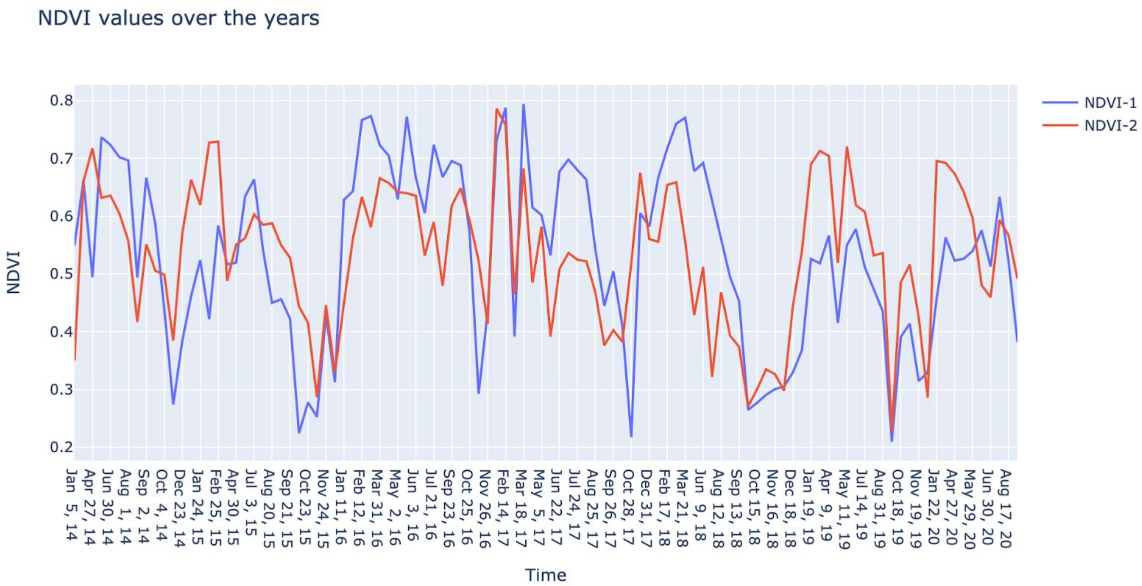

Fig 6. NDVI Time Series Plot

Impact and Significance of the Results

With the findings from our solution, our stakeholders (Farmers & Government) can gain valuable insight into the progress of specific crop fields across the state. Farmers can keep track of various factors affecting the growth of the crops to then evaluate the proper procedures required for maximum yield and production. The Government can use our solution to compare the progress and production rate of different regions to then maximize the country's production by helping out the regions in need using the same tactics used by successful fields.

Future Plans

Some of the additional features we aim to add in our research are :

- Monitoring the level of nitrogen and phosphorus present in the soil. They are the next important part of the fertility program after soil pH, and to ensure they are within the minimum agricultural standards to protect the environment. The mixture of phosphorus and nitrogen can affect the water quality if moved offsite into waterways.
- Provide an estimated crop yield forecast based on the current growth and past performance.
- Extend the web app support to more stakeholders and investors for high-level benefits.

Project Management

The project was divided into three phases, Planning, Execution and Presentation phase. In the planning phase, we were obliged to come up with the solution, complete milestone one requirements i.e. Use cases, Stakeholders & their requirements, future plans etc and do all the research & exploration for the solution. The execution phase required all of us to work on the development of our webApp which was divided into two stages 1) the map development which required us to code the map UI interaction and add the relevant filters to it, 2) WebApp development, is where we build the Web application and also processing and displaying the findings/input from the map. The final phase was for us to prepare for the two presentations we had and record our solution's demo. Below is a Gantt Chart for our project:

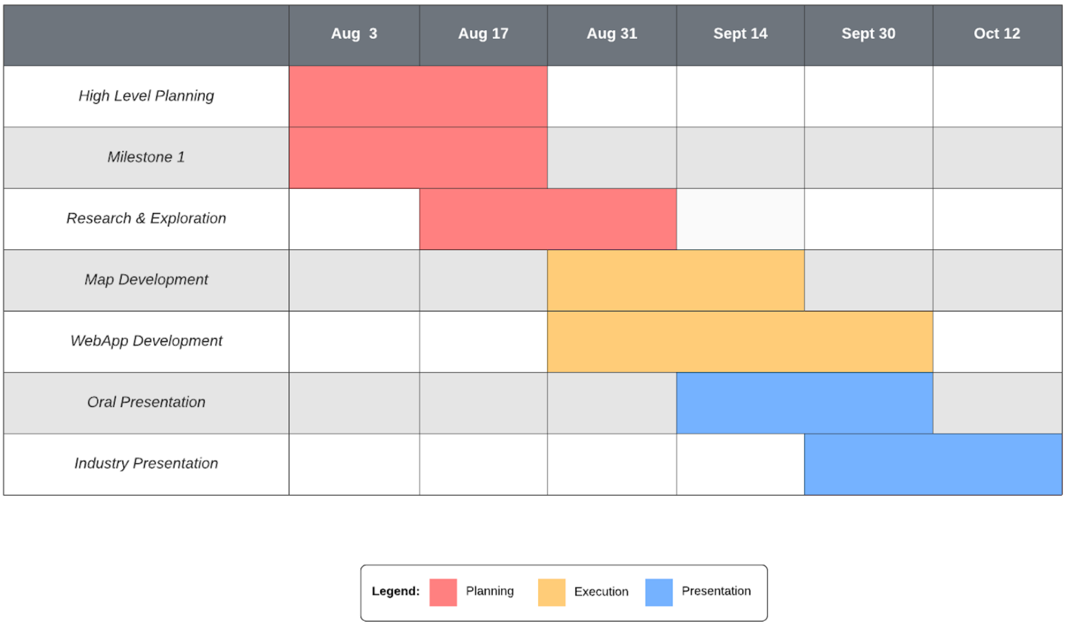

Fig 7. Gantt Chart

The team used slack as our main source of communication as well as a file-sharing platform. Weekly meetings were held to share information and resources as well as to keep track of each other. The contribution from the team was equal and the work done by each member was based on the strengths they offered to the team.

Name	Student ID	Contribution
Maaz Shaikh	s3795603	**25%**
Vaishnavi Narayana Naik	s3797442	**25%**
Aditi Sharma	s3795604	**25%**
Swati Arora	s3796848	**25%**

Conclusion

Since sugarcane is the powerhouse supporting various communities and businesses, using the metrics generated for the selected field(s), we can provide analytics to various stakeholders, resulting in healthier sugarcane crops and more business profit. This approach helps manage agriculture standards across the selected regions and developing future strategies for betterment.

The study above suggests that as the soil composition and weather conditions are very crucial for a sugarcane crop and can have a severe impact on its growth, it is of utmost importance to remotely monitor the regions that need more assistance to boost the overall production.

Our suggested application would make this difficult task easier and convenient to the farmers and the government and further aid them in making accurate decisions for the betterment of the crop.

References

[1]*Sugarresearch.com.au*,2020.[Online].Available: https://sugarresearch.com.au/wp-content/uploads/2019/06/Australian-Sugarcane-Nutrition-Manual-June-2019.pdf. [Accessed: 15- Aug-2020].

[2]*Elibrary.sugarresearch.com.au*,2020.[Online].Available: https://elibrary.sugarresearch.com.au/bitstream/handle/11079/15541/Chapter%204%20The%20Sugarcane%20Plant.pdf?sequence=5&isAllowed=y#:~:text=Five%20phases%20occur%20in%20the,slows%20down%20and%20sugar%20storage. [Accessed: 15- Aug- 2020].

[3] "Sugarcane growth stages | agropedia", *Agropedia.iitk.ac.in*, 2020. [Online]. Available: http://agropedia.iitk.ac.in/content/sugarcane-growth-stages. [Accessed: 15- Aug- 2020].

[4]2020.[Online].Available: https://www.researchgate.net/publication/226137423_Soil_temperature_in_a_sugar-cane_crop_as_a_function_of_the_management_system. [Accessed: 15- Aug- 2020].

[5]U. Reflectance and G. Engine, "USGS Landsat 8 Collection 1 Tier 1 TOA Reflectance", *Google Developers*,2020.[Online].Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C01_T1_TOA. [Accessed: 09- Sep- 2020].

[6]"ImageCollection Overview | Google Earth Engine | Google Developers", *Google Developers*, 2020.[Online].Available: https://developers.google.com/earth-engine/guides/ic_creating. [Accessed: 09- Sep- 2020].

[7]"JavaScript and Python Guides | Google Earth Engine | Google Developers", *Google Developers*, 2020. [Online]. Available: https://developers.google.com/earth-engine/guides. [Accessed: 09- Sep- 2020].

[8]"Python installation | Google Earth Engine | Google Developers", *Google Developers*, 2020. [Online]. Available: https://developers.google.com/earth-engine/guides/python_install. [Accessed: 12- Sep- 2020].

[9]"Queensland SpatialCatalogue- QSpatial: QueenslandGovernment", *Qldspatial.information.qld.gov.au*, 2020. [Online]. Available: http://qldspatial.information.qld.gov.au/catalogue/custom/index.page. [Accessed: 14- Aug- 2020].

[10]"Landsat Normalized Difference Vegetation Index", *Usgs.gov*, 2020. [Online]. Available: https://www.usgs.gov/core-science-systems/nli/landsat/landsat-normalized-difference-vegetation-index. [Accessed: 16- Sep- 2020].

[11]"Normalized Difference Moisture Index", *Usgs.gov*, 2020. [Online]. Available: https://www.usgs.gov/core-science-systems/nli/landsat/normalized-difference-moisture-index. [Accessed: 07- Sep- 2020].

[12] "Remote Sensing of Soil Alkalinity and Salinity in the Wu'er-Shuangyang River Basin, Northeast China", MDPI AG 2020. [Online]. Available: http://dx.doi.org/10.3390/rs8020163 [Accessed: 07- Sep- 2020].