SATELLITE INTELLIGENCE

Case Studies in Data Science

17 October 2019

CONTENTS

INTRODUCTION

Australia is a significant world producer of agricultural crops. They provide input for many industries, particularly the food processing industry. The agricultural sector in the country has had a turnover of $65 billion and a value add of $24 billion (Australianimmigrationbook.com.au, 2019).

Australia is the second largest supplier of raw sugar in the world market, worth $2 billion to the Australian economy each year. It provides the input for food and beverage industry. Sugarcane is predominantly grown in areas with tropical climate along the coastal regions, where there is heavy rainfall and water sources nearby. It is generally found along the North-East coast through Queensland and Northern New South Wales of which 95% raw sugar is produced from Queensland(AgriFutures Australia, 2019).

PROBLEM STATEMENT

Queensland government has a target to increase the food production, export and provide food for Australia's growing population. The farmers however, are struggling to meet up to these expectations as they are unable to productively use their own land under the legislation.

Smaller areas in coastal regions of Queensland is taken so as to analyse the growth of sugarcane through the years. The Sentinel 2A satellite provided images 10-20 days apart that can be analysed for further predictions.

30 JUNE 2019

20 JULY 2019

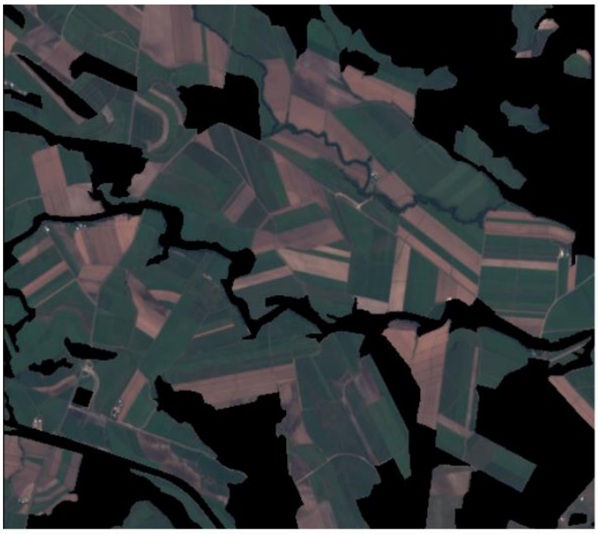
The images depict lands that are barren, land that contains the sugarcane and land that contains other crops. The greens were classified according to different period of their growth cycle and the quality of the crop. These images were used to understand the harvest trend of sugarcane. The different shades of green was noted to explain the maturity and quality of the crop. Once the cycle of the sugarcane crop was understood it would be easier to predict the harvest and production of the raw product. The stakeholders such as the farmers and sugar mills were informed of the time of harvest, on how to improve the yield of the crop or the start of raw product production(Canegrowers.com.au, 2019).

The sugar obtained can be fermented to produce Ethanol. The Biomass that is left after the juice is extracted can be used to generate steam and electricity. This could be an alternate biofuel source that can uplift the socioeconomic status of the country and the sustainability of natural resources. The analysis of the sugarcane fields would help stakeholders such as government to help identify new barren land for assisting the Biofuel industry(Talukdar et al., 2017).

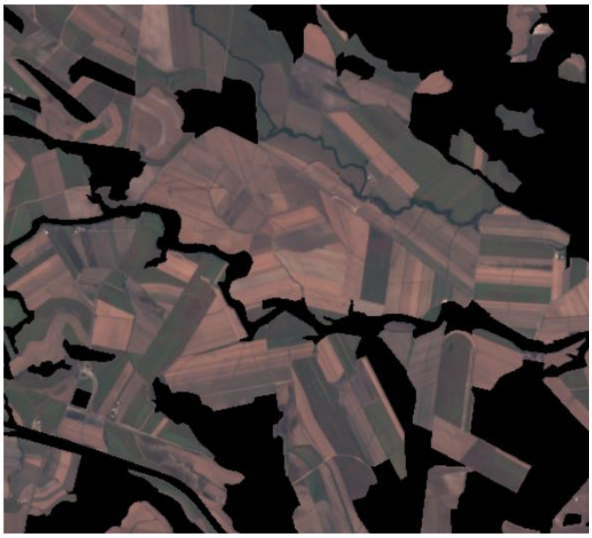
METHODOLOGY

DATA PREPARATION

As the first step in data preparation, the mask provided along with the dataset was applied on the TCI images. This helped identify the areas that cultivate sugarcane as compared to other crops. Below are the TCI images on which sugar region mask is applied.



20 JULY 2017



28 SEPTEMBER 2017

The json file provided was converted into an excel containing 74 rows and 33 columns. This file contains information pertaining to various properties of the images such as the location of the images, the snow cover, cloud cover, date of the image capture, season, centroid coordinates etc.
There are three extra json files(rows in excel) for following dates, 2017-12-02, 2018-01-01, 2019-04-26.

From this data, the images that contained a cloudCover more than 30% were ignored as these images could negatively impact the results gained from the data modelling.

DATA EXPLORATION – TIME SERIES EXPLORATION

- Land pixels of each TCI image is found using the formula green/ (green+red+blue) and then by masking those regions having threshold less than 0.32

- Following is a TCI image in which land pixels are masked

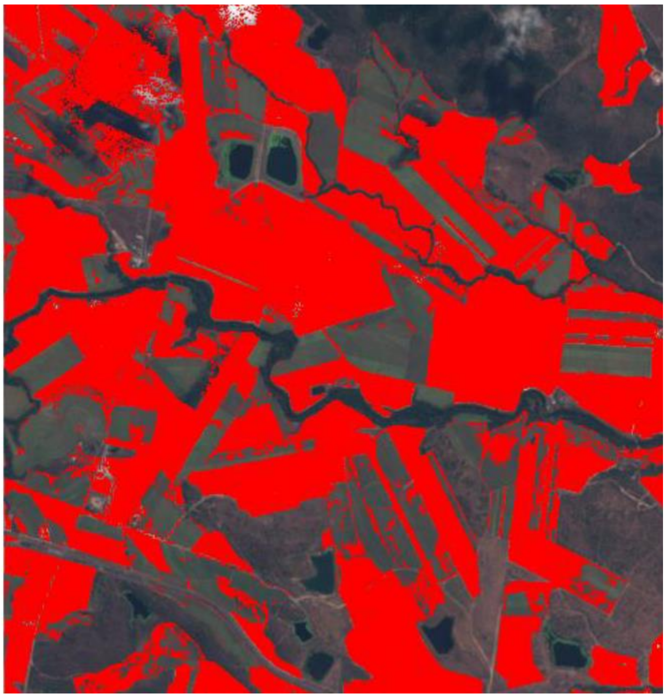
BEFORE MASKING

Image of Proserpine region as seen on 24th August 2018

24TH AUGUST 2018

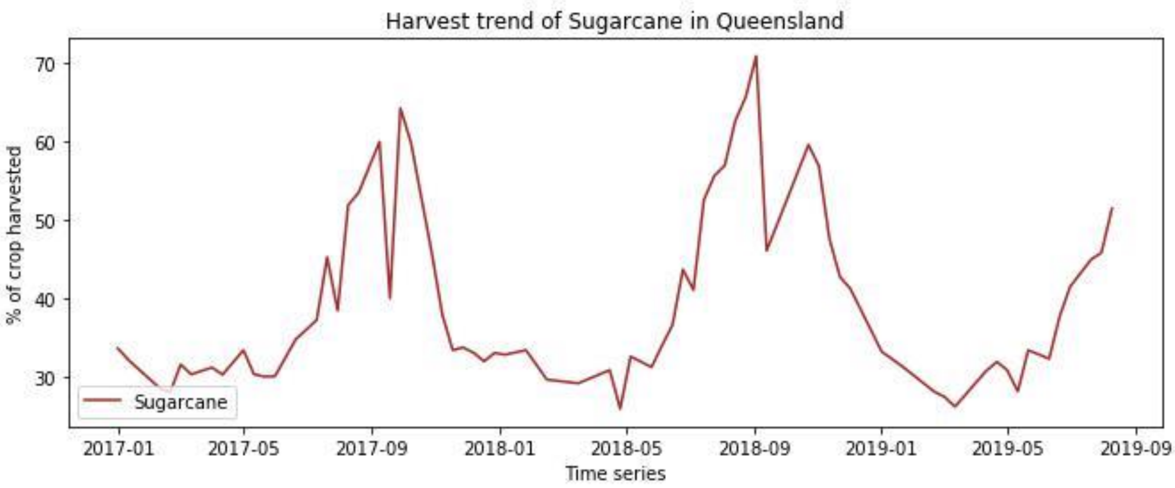
AFTER MASKING

Image of Proserpine region as seen on 24th August 2018



24TH AUGUST 2018

- To understand the trend in sugarcane harvest over a period of time, a time-series has been plotted .



- The graph shown depicts the harvest trend of sugarcane in Proserpine area of Queensland from a time period of January 2017 - September 2019
- By observing the graph, it can be inferred that most of the sugarcane is harvested between the months of July and October, which is winter season in Queensland.

DATA EXPLORATION – NORMALIZED DIFFERENTIATION VEGETATION INDEX

What is NDVI?

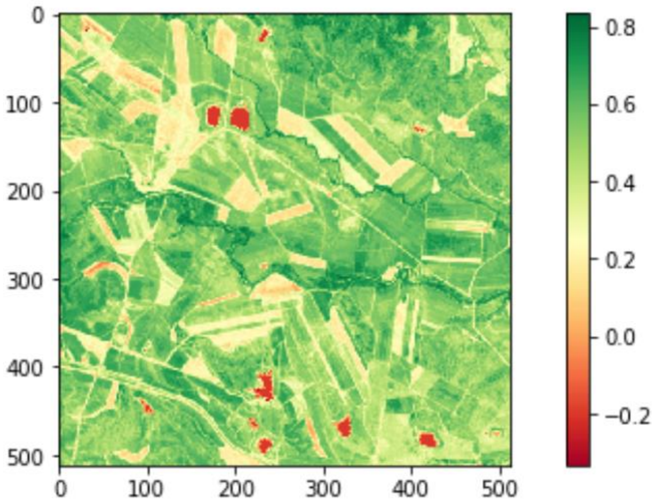
The Normalized Differentiation Index or the NDVI is used to indicate the health of the vegetation at a given time for a particular area of interest (GIS Geography, 2018). The hypothesis that is being tested here is to determine the period of greenest cover and inform the stakeholders – farmers and sugar mill owners that the crop is ready for harvest. In order to calculate the NDVI, the red band (Band 4) and the near-infrared band (Band 8) from the Sentinel2A images were taken. These were used to calculate NDVI using the formula below:

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$$

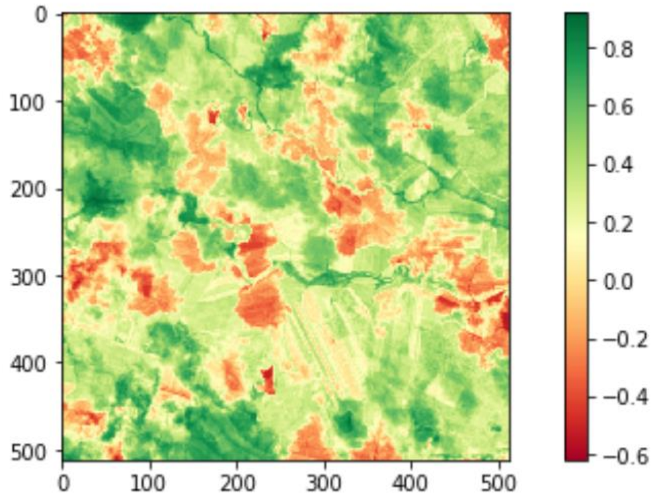
Where NIR indicates the value for Near-Infrared and Red indicates the value for the Red band of the image in consideration. NDVI values generally range between -1 and +1. A variety of data is captured by satellite sensors and one such type of data specifically measures wavelengths of light absorbed and reflected by green plants. Dense vegetation reflects a lot of near-infrared light (not visible to the human eye) as compared to the visible red light, which means that the NDVI value is closer to +1. The reverse happens in case of sparse vegetation, indicating the NDVI value is closer to -1. Thus, as a plant canopy changes from early spring growth to late-season maturity and senescence, these reflectance properties also change (Pandey, 2018).

NDVI Calculation for Sentinel2A Images:

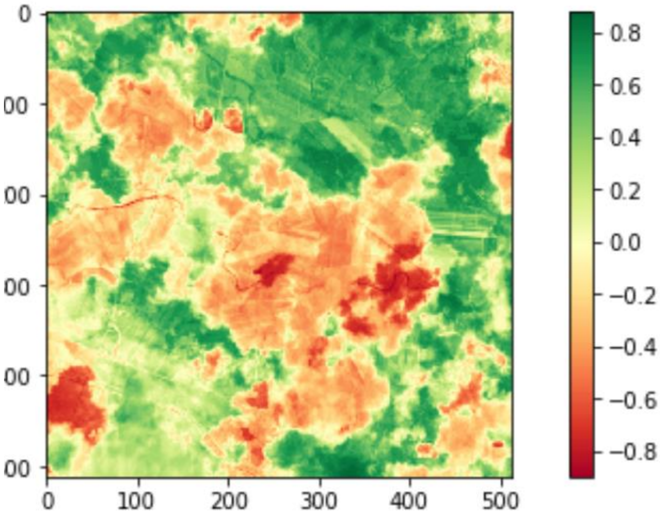
This concept applied to the TCI images and the mean NDVI values of each image were taken. Below are few of the TCI images after NDVI calculation



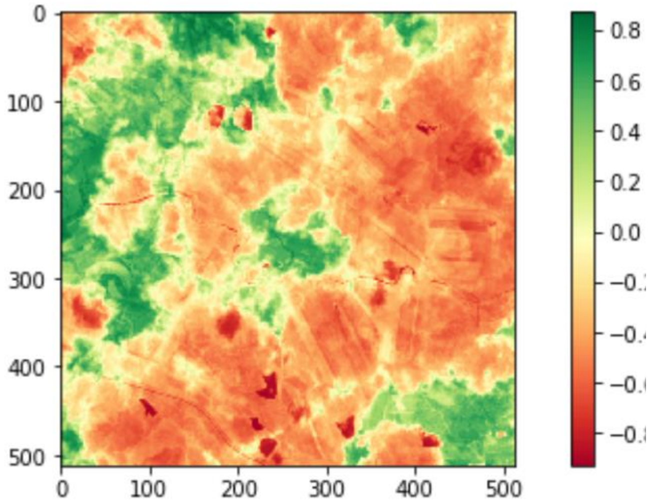
12TH MARCH 2019



11TH MAY 2019



15TH JUNE 2018



3RD SEPTEMBER 2018

Results & Analysis:

The images depict how the harvest cycle in the Proserpine area of Queensland changes over the months. It can be observed that during the period between March and May, most of the areas are reflected in green, meaning the crops are ready to harvest. As we move forward from June to September, there is a higher reflection from the red area, indicating that there is more land available to be planted.

Applications of NDVI:

The results of these analyses can be useful to inform the stakeholders (farmers) to conduct Precision Farming that can help them correspondingly prepare for either planting or harvest. Another very useful application of NDVI is that it can help indicate drought. When the NDVI values are very low (closer to -1), it indicates that the water levels in the surrounding regions are restricting the growth of the crop. This leads to a decline in the vegetation and thus a drought-like situation. This application of NDVI can also be used to inform the stakeholders to prepare for any adverse or untoward incidents.

DATA MODELLING: PREDICTION OF NDVI LEVELS USING IMAGE CLASSIFICATION (CNN)

In order to build an image classifier model that uses satellite imagery of the sugarcane field to predict NDVI levels, the following techniques were used.

The sugarcane - region mask image, which contained the possible sugarcane growth area was overlaid on each TCI image. The final image after applying the mask is given below:

TCI Image   Sugarcane region mask   Final image after applying mask

These masked images are used as source data. In order to prepare the target class, the mean NDVI values were calculated for each image using their respective band images such as B4 and B8. Furthermore, the mean NDVI values are classified into 5 labels given below:

- Label “1”: for mean NDVI <0.15. Signifies low overall vegetation
- Label “2”: for mean NDVI between 0.15-0.2. Indicates crops are in early phenological phase
- Label “3”: for mean NDVI between 0.2-0.3. Signifies crops are growing
- Label “4”: for mean NDVI between 0.3-0.5. Indicates crops are mature and ready for harvest
- Label “5”: for mean NDVI >0.5. Indicates high overall vegetation

The data containing masked images was considered as the source and their respective NDVI labels as target. This was arranged based on their date of capture and split into 2 parts with ratio 8:2. The first 80% of the data was used for training while the rest was used for NDVI level prediction. The initial data is split sequentially as the sugarcane field's NDVI values alter from low to high during each plantation and high to low during harvest cycle throughout a single year provides a rich training data.

The training data is then fed into Convolutional Neural Network (CNN) model developed using keras. Then, train_test_split is used for random split of training data by random state 42. The input image pixel size, while reading, was reduced to 256 from 521 for faster evaluation.

CNN Model Architecture:

The model contains convolution layers, polling layers, flatten and dense layers. The detailed architecture is given below:

Layer (type)	Output Shape	Param #
conv2d_57 (Conv2D)	(None, 254, 254, 256)	7168
max_pooling2d_30 (MaxPooling)	(None, 127, 127, 256)	0
conv2d_58 (Conv2D)	(None, 125, 125, 64)	147520
max_pooling2d_31 (MaxPooling)	(None, 62, 62, 64)	0
conv2d_59 (Conv2D)	(None, 60, 60, 64)	36928
max_pooling2d_32 (MaxPooling)	(None, 30, 30, 64)	0
conv2d_60 (Conv2D)	(None, 28, 28, 256)	147712
max_pooling2d_33 (MaxPooling)	(None, 14, 14, 256)	0
conv2d_61 (Conv2D)	(None, 12, 12, 512)	1180160
flatten_17 (Flatten)	(None, 73728)	0
dense_33 (Dense)	(None, 512)	37749248
dense_34 (Dense)	(None, 6)	3078

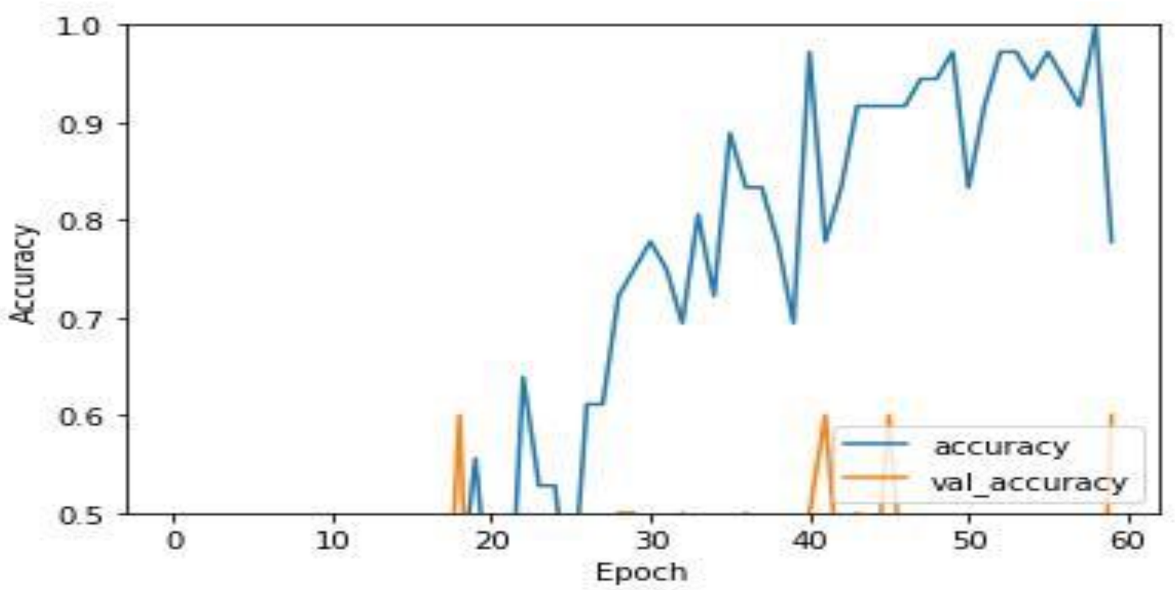
Total params: 39,271,814
Trainable params: 39,271,814
Non-trainable params: 0

The training data is then passed onto 60 epochs. A sample epoch from the model is given below:

Epoch 60/60
36/36 [==============================] - ETA: 2s - loss: 0.4270 - accuracy: 0.84 - 24s 679ms/step - loss: 0.6574 - accuracy: 0.7778 - val_loss: 2.6775 - val_accuracy: 0.6000

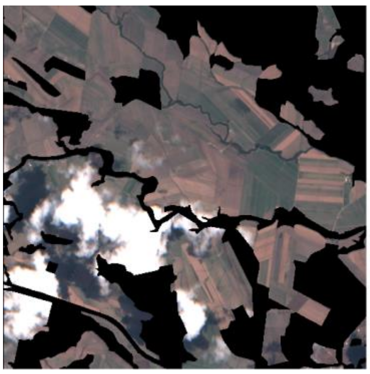
Model Evaluation:

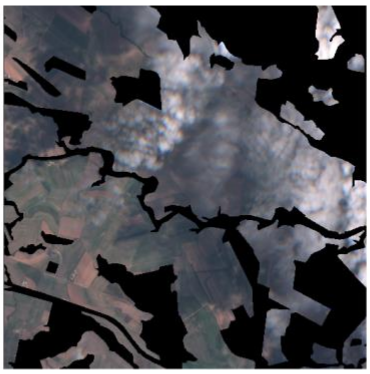
The resultant accuracy obtained from this model is 60% (Accuracy: 0.6000000238418579). The error rate is mostly due to imbalanced class labels in training. The below graph indicates the accuracy rates during each epoch.

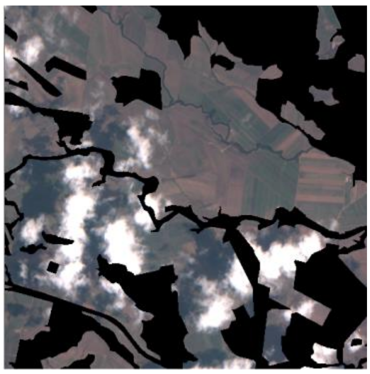


NDVI level prediction:

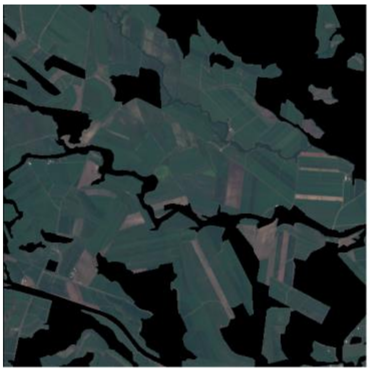
The rest of the 20% data separated from the initial stage is used for prediction. The prediction results obtained from the Image classifier (using CNN) model is given below:



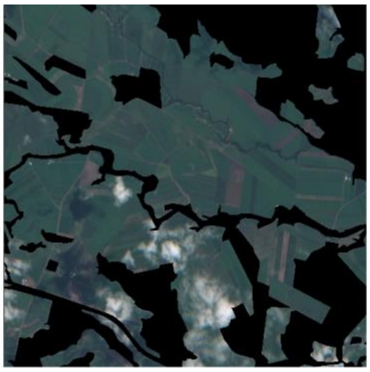




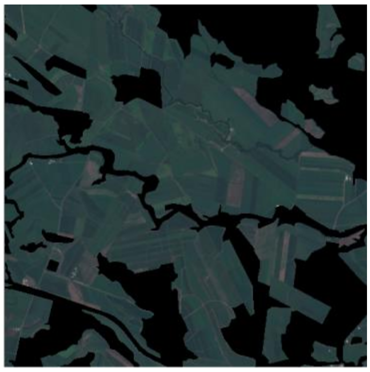
Actual: 3 Predicted: 3 Actual: 1 Predicted: 1 Actual: 3 Predicted: 4



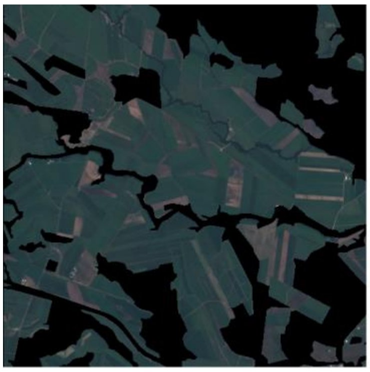
Actual: 4 Predicted: 4



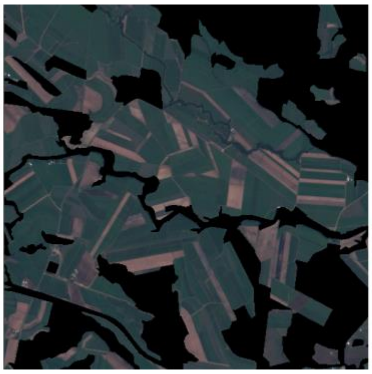
Actual: 4 Predicted: 4



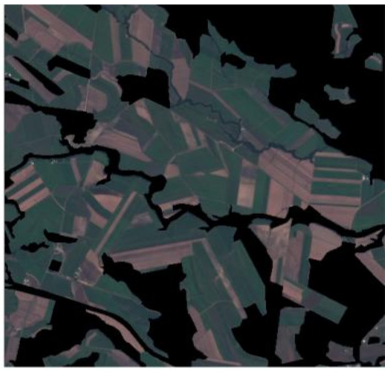
Actual: 4 Predicted: 5



Actual: 4 Predicted: 4



Actual: 4 Predicted: 4



Actual: 3 Predicted: 4

It could then be concluded that the images that were classified above the NDVI label of 3 and more contained majority of areas that were ready to harvest and the images that were classified under the NDVI label of 3 consisted of areas that could potentially grow sugarcane.

CONCLUSION

Insights obtained from Image Classification Algorithm:

1. **Farmers –** The image classification algorithm predicts the NDVI levels in 5 different categories of 1 to 5. The farmers would like early prediction of when the crops can be harvested. Also, an understanding of when they can expect the land to not have any vegetation would be useful to them. This categorization would guide the farmers to decide when the crops are to be harvested or when it would be a perfect time to plant new sugarcane crops.

2. **Sugar Mills –** The sugar mill owners await the arrival of sugar canes. An NDVI level of 4 and 5 will notify them that they can expect the sugar cane any time soon. This would in turn help them in better planning of their activities.

ALGORITHMS CURRENTLY IN PIPELINE

1. **Time Series Clustering**

   Goal for this algorithm

   The goal of this algorithm was to obtain a pattern of NDVI values across different time periods and predict the relationship for the future. The historical data obtained for NDVI level during different time periods will have pattern denoting the various cycles of the NDVI levels and the temperature of that area during this time period. The algorithm should train the model based on this analysis and predict the NDVI levels for a future time period. This will enable the farmers to understand the vegetation of that region and plan accordingly as to what needs to be cultivated in the land or when it would be the right time to plant the crops.

   For example: If the model trained predicts the NDVI level to be low during the period of August to September, the farmers can decide if they want to harvest new crops during this phase or let the land fertilise and prepare itself for the next season.

   What is Time Series Clustering

   Clustering is the practice of finding hidden patterns or similar groups in data. It is one of the most common methods for unsupervised learning, where a classification is given to every data entry without predefining the different classes. (Anon, 2019). Clustering time-series data has been used in diverse scientific areas to discover patterns which empower data analysts to extract valuable information from complex and massive datasets. In case of huge datasets, using supervised classification solutions is almost impossible, while clustering can solve this problem using unsupervised approaches. In this research work, the focus is on time-series data, which is one of the popular data types in clustering problems. (Beta.vu.nl, 2019).

Application of Time Series Algorithm to the satellite images of sugarcane regions-

The NDVI levels obtained from the Image Classification algorithm will be used to establish a pattern for the NDVI levels during different time periods. This pattern will be used to determine the variations in NDVI level that can be expected in the future.

For example: A pattern of high NDVI levels are observed during the period of January to March for the data provided. We expect a high NDVI prediction for the same period while training the model for Time series Algorithm.

2. Tree Prediction using Deep Learning

Goal for this algorithm

Many government agencies would be interested in determining if the land would be good enough for future investment. They would like to know about a particular type of land in the region rather than the vegetation or harvesting patterns of the land. An accurate percentage of the particular type of land coverage in that area or a highlighted view of only those particular regions would be of interest to them. The algorithm will help to extract this information.

What is Tree Prediction algorithm-

This algorithm examines the vegetation cover of a region with the help of satellite data.

The satellite imagery data can be analyzed over a period of time to understand the variation in the vegetation. Prediction of hurricanes, droughts and floods are areas where analysis of tree prediction algorithm is being extensively applied. (Medium, 2019).

Application of Tree Prediction Algorithm to the satellite images of sugarcane regions

The algorithm is applied to bifurcate regions with different types of vegetation. For example, it will highlight only the barren land, area with larger amount of vegetation etc. Using this highlighted area, the percentage of that land cover can be cultivated for further analysis.

REFERENCES

1. Australianimmigrationbook.com.au. (2019). An Overview of Australian Agriculture | The Australian Immigration Book. [online] Available at: http://www.australianimmigrationbook.com.au/business/overview-australian-agriculture [Accessed 17 Oct. 2019].
2. AgriFutures Australia. (2019). Sugarcane | AgriFutures Australia. [online] Available at: https://www.agrifutures.com.au/farm-diversity/sugarcane/ [Accessed 17 Oct. 2019].
3. Talukdar, D., Verma, D., Malik, K., Mohapatra, B. and Yulianto, R. (2017). Sugarcane as a Potential Biofuel Crop. Sugarcane Biotechnology: Challenges and Prospects, pp. 123-137.
4. Canegrowers.com.au. (2019). Vegetation amendments put brakes on sugarcane growth. [online] Available at: http://www.canegrowers.com.au/page/media/media-releases/2018/vegetation-amendments-put-brakes-on-sugarcane-growth [Accessed 17 Oct. 2019].
5. GIS Geography. (2018). What is NDVI (Normalized Difference Vegetation Index)? - GIS Geography. [online] Available at: https://gisgeography.com/ndvi-normalized-difference-vegetation-index/ [Accessed 17 Oct. 2019].
6. Pandey, P. (2018). Satellite Imagery Analysis with Python. [online] Medium. Available at: https://medium.com/analytics-vidhya/satellite-imagery-analysis-with-python-3f8ccf8a7c32 [Accessed 17 Oct. 2019].
7. [1] Medium. (2019). What the NDVI index is and how it makes a farmer's life easier. [online] Available at: https://medium.com/onesoil/what-the-ndvi-index-is-and-how-it-makes-a-farmers-life-easier-d6e900d91c9f [Accessed 17 Oct. 2019].
8. [2] Wang, C. and Xi, Y. Convolutional Neural Network for Image Classification. [online] Available at: https://pdfs.semanticscholar.org/b8e3/613d60d374b53ec5b54112dfb68d0b52d82c.pdf [Accessed 17 Oct. 2019].
9. [3] Beta.vu.nl. (2019). [online] Available at: https://beta.vu.nl/nl/Images/ stageverslag-roelofsen_tcm235-882304.pdf [Accessed 17 Oct. 2019].
10. [4] Anon, (2019). [online] Available at: https://umexpert.um.edu.my/file/publication/00005678_120457.pdf [Accessed 17 Oct. 2019].
11. [5] Medium. (2019). Tree Cover Prediction with Deep Learning. [online] Available at: https://medium.com/dataseries/tree-cover-prediction-with-deep-learning-afeb0b663966 [Accessed 17 Oct. 2019].