Agriculture Loan Verification using GIS and Artificial Intelligence

Mr. Aviruk Chakroborty
MS Computational Engineering, Ruhr University, Germany

Abstract: Agriculture is the pillar of any developing economy which is all the truer in case of India. Agriculture sector plays a strategic role in the economic life of the Indian society. Despite having so much importance, it is increasingly facing various problems, which arise from inadequate technologies, indifferent attitude of the government, socio-economic condition of farmers and the credit systems prevalent in India. For more than 100 years now, the Central Government and the Reserve Bank of India have been making efforts to enhance institutional credit in rural areas particularly to assist in agricultural operations. But certain inherent challenges and risks such as high number of non-performing assets (“NPAs”) are posing problems. To solve these problems, it is important that certain substitute approaches as extension to monitor agriculture activities and fair assessment mechanisms for deducing loan viability should be practiced. This paper suggests using several Agriculture Indices and their extension on various Artificial Intelligence enabled innovative technology for decision making if an agriculture loan should be disbursed and making inferences on the growth of the crop in the catchment area.

Key Words – Agriculture Loan, NDVI, GIS, Artificial Intelligence

Introduction

Agriculture is the pillar of any developing economy which is all the truer in case of India. Agriculture sector plays a strategic role in the economic life of the Indian society. In the Indian economy agriculture contributes one-third of the national income. Sixty percent of the export directly or indirectly originates from agriculture sector. It provides employment to 67 percent of the work forces. It plays a decisive role in economic development and planning and provides numerous to the industrial and service sector. The requirements of finance in agricultural sector, very few farmers will have capital of their own to invest in agriculture. Therefore, a need arises to provide credit to all those farmers who require it. Even if we look into the expenditure pattern of the farmer families, they have hardly any savings to fall back on. Therefore, credit enables the farmer to advantageously use seeds, fertilizers, irrigation, machinery, etc. farmers must invariably search for a source, which supplies adequate farm credit. For more than 100 years now, the Central Government and the Reserve Bank of India have been making efforts to enhance institutional credit in rural areas particularly to assist in agricultural operations. But certain inherent challenges and risks such as high number of non-performing assets (“NPAs”) are posing problems.

The impact of the unseasonal rains brings a lag, as NPA recognition policies for agricultural loans (one or two crop seasons past due) differ from those of corporate or retail loans (90 days past due). With delinquencies in the agri-loan portfolio likely to rise, they will add to the already stressed assets of banks translating to impact profitability of banks. As a reason, this paper suggests an alternative method and draws an implementation road-map for decision making if an agriculture loan should be disbursed and making inferences on the growth of the crop in the catchment area using several Agriculture Indices and their extension on various Artificial Intelligence enabled innovative technology.

Literature Review

The literature review consisted of exploring various open data sources on prevailing remote sensing methodologies used for agricultural data capture. The data sources published by ISRO and NASA websites were instrumental in refining the methodology and approach towards creating an integrated Agriculture Loan Automation Process. The following papers and various Open Sources provided the data points needed to create a new robust process. The second part of the research direction was on hardware needed to capture and communicate latitude and longitude in a real-time fashion. For this we evaluated technical specifications of various commercial instruments used in the land survey industry. This led to creating our own hardware with radical innovations in the field of aggressive cost optimization so that it will be affordable and expendable and hence can be used by the poorest farmers in the world.

Researchers like Peter Strasse of Institute for Surveying, Remote Sensing & Land Information (IVFL), Vienna have explored various remote
sensing applications in agriculture. The reasoning and logics he presents lucidly draws on how use of remote sensing can significantly contribute monitoring prospects of the agriculture sector and this indeed can enhance growth of investments with less risk. Over here the paper is scoping the application of Remote Sensing and GIS data for reducing risks for banks while disbursing loans for farmers (in India and all over the world). The process also encompasses use of a hardware tool modified with Artificial Intelligence.

Agriculture Loan Verification Process
We ask the farmer to fill out only 3 details:

1. Biometric or Aadhar data
2. Land Size
3. And place a device 1 inch cube size in the farm (This device costs only INR 200 and is reusable by the bank for other Agricultural loan customers as well) (It is a 4 channel GPS receiver) [or his Lat Long if he/she has a smart phone]

The rest is all automated till the disbursement of loan.

A biometric finger print on smart phones which will be available to bank employees will be used to take thumb prints of uneducated framers. These thumb prints will be used to access the Database and using the Aadhar database we access the corresponding file in CIBIL e - KYC and auto fill the user details in the bank loan form.

The underlying framework can be built on top of a block chain in phase 2 for added security and making it more distributed in nature.

From a Farmer’s biometric or Aadhar we will auto fill out the details by using CIBILs EKYC protocol and then the location of the land will be cross verified with land registry (if its digitized in that state) and the Lat Long for farmers who have no phones will be given by the 1 inch cube GPS Receiver. Using this location, we will run a NDVI via Landsat 8 data and classification algorithm of Google earth satellite data.

This gives us a probability score on past and future performance which when embedded in our master algorithm via logistic regression and ANN gives me a probability of Default score and this can be used a key decision matrix for Agriculture loans. Also, we will use NDVI for monitoring of the land post loan disbursement.

The Aadhar Card will also be used to integrate the loan forms with UIP so that the loan amount can directly flow into the verified designated Krishi personal account using UIP Framework.

Land registries in India and UAE are fully digitalized in selected states. Hence these land registries will be accessed to verify land ownership pre-loan disbursal. Pre-loan disbursal there will also be multi stage data validation using multiple databases including voters’ ID. We will also integrate this platform with various rural platforms such as FASAL and E-Chowpal.

We also will distribute a 4 channel GPS receiver for verification of land location and cross reference it with digitized land records.

We use a logistic regression based Default Probability Model for rating proposals pre-loan disbursal. These ratings are a part of loan approval decision mechanism.

NDVI and Google Earth, LSAT - We use Remote Sensing using Bhuvan (Indian Government's Remote Sensing Platform) and Land SAT data for estimating the probability default pre-loan disbursment and also as a monitoring tool post loan disbursement. This process is analogous to our PDLR model but in case of social media sentiment which we use for our urban personal loans, we use a hybrid parameter which is created by amalgamating NDVI, along with a image classifier super imposed on the Land SAT data of the farmer's field on which the loan is being taken. The machine learning image classifier looks at the cropped image and then classifies it into green, fallow, habituated, reaped, fertile and various other categories. this model also looks at the NDVI and growth pattern of that specific land of the past three years and also monitors growth of the crop in the designated farm land over the next Agricultural cycle post loan disbursement.

If we look at past balance sheets of organizations to judge their capabilities and to give them loan, why don’t we objectively quantify past performance of a farmer using unbiased satellite data. This is more necessary as still farming is a predominantly cash economy and quantifies financial results are not always available.

Chakroborty Agriculture Loan Water Risk Index (CALWRI) - This index quantifiably rates the risk of the banks Agriculture Loan Portfolio in relationship to water. This index is created by categorizing exposure of Agriculture Loan into elegant classifications: a.) Lands which uses irrigation and b.) rain dependent lands.

These zero to hundred scaled indices use NDWI to quantify water risk in irrigated lands and a cost function proportional to the premium in a HDD.
weather option for auto hedging and mitigating rain delay risk.

This is what Wikipedia says about NDVI Index:

**The normalized difference vegetation index (NDVI)** is a simple graphical indicator that can be used to analyze remote sensing measurements, and assess whether the target being observed contains live green vegetation or not.

Live green plants absorb solar radiation in the photosynthesis active (PAR) spectral region. Leaf cells have also evolved to re-emit solar radiation in the near-infrared spectral region (which carries approximately half of the total incoming solar energy), because the photon energy at wavelengths longer than about 700 nanometers is not large enough to synthesize organic molecules. A strong absorption at these wavelengths would only result in overheating the plant and possibly damaging the tissues. Hence, live green plants appear relatively dark in the PAR and relatively bright in the near-infrared. The pigment in plant leaves, chlorophyll, strongly absorbs visible light (from 0.4 to 0.7 µm) for use in photosynthesis. The cell structure of the leaves, on the other hand, strongly reflects near-infrared light (from 0.7 to 1.1 µm). The NDVI is calculated from these individual measurements as follows:

\[
\text{NDVI} = \frac{(\text{NIR} - \text{VIS})}{(\text{NIR} + \text{VIS})}
\]

where VIS and NIR stand for the spectral reflectance measurements acquired in the visible (red) and near-infrared regions, respectively. This spectral reflectance are themselves ratios of the reflected over the incoming radiation in each spectral band individually, hence they take on values between 0.0 and 1.0. By design, the NDVI itself thus varies between -1.0 and +1.0. It should be noted that NDVI is functionally, but not linearly, equivalent to the simple infrared/red ratio (NIR/VIS). The advantage of NDVI over a simple infrared/red ratio is therefore generally limited to any possible linearity of its functional relationship with vegetation properties (e.g. biomass). The simple ratio (unlike NDVI) is always positive, which may have practical advantages, but it also has a mathematically infinite range (0 to infinity), which can be a practical disadvantage as compared to NDVI. Also in this regard, note that the VIS term in the numerator of NDVI only scales the result, thereby creating negative values. NDVI is functionally and linearly equivalent to the ratio \(\text{NIR} / (\text{NIR}+\text{VIS})\), which ranges from 0 to 1 and is thus never negative nor limitless in range. But the most important concept in the understanding of the NDVI algebraic formula is that, despite its name, it is a transformation of a spectral ratio (NIR/VIS), and it has no functional relationship to a spectral difference (NIR-VIS).

We use Landsat 8 data band 4-5-6 for NDVI calculations and use the Hyper spectral sensor in NASA's EO1 satellite called Hyperion to calculate over 17 variables which are used to calculate the yield probability and to monitor the farm output - some of the parameters are listed below:

- **Vegetation Condition Index (NDVI)** = \(100\%\times(\text{NDVI}-\text{NDVI}_{\text{min}})/\text{(NDVI}_{\text{max}}-\text{NDVI}_{\text{min}})\)
- **Vegetation Condition Index (NDWI)** = \(100\%\times(\text{NDWI}-\text{NDWI}_{\text{min}})/\text{(NDWI}_{\text{max}}-\text{NDWI}_{\text{min}})\)
- **Temperature Condition Index** = \(100\%\times(\text{LST}_{\text{max}} - \text{LST})/(\text{LST}_{\text{max}} - \text{LST}_{\text{min}})\)
Agriculture Loan Indices

In the field of agriculture to have an advanced technology implementation we can use Hyperspectral mapping for detecting and prediction of soil characteristics, vegetation and crop parameters, water, chlorophyll index, environment prediction and stress index for all agriculture parameters which acts as an aid for improved and field wide applications.

As hyperspectral mapping gives the image resolution of vegetation, soil and environment for crop yielding it helps in identifying more likeable area for investing in agriculture field.

Besides this, hyperspectral imagery also provides spatial data which anticipates with GIS models and gives crop responses to nutrients, water, vegetation vitality and pesticides to generate application prescription maps.

With GIS, available data from satellites we can aggregate and model field data and crop status data collected for real time which will promote increased efficiencies and reduced producer costs and impact on the environment.

There are 15 variable indexes which are primarily necessary for this implementation which are listed below.

List of 15 agriculture indices:

1. Normalized Difference Vegetation Index (NDVI) = \( \frac{(\rho_{NIR} - \rho_{RED})}{(\rho_{NIR} + \rho_{RED})} \) for Landsat 8 = \( \frac{NIR-RED}{860-1640} \)

Variables are as follows: RED = [670, 50, 30] and INR = [800; 10; 10] and explanatory variables RED = 700 nm and NIR = 800 nm

Wavelengths = 670; 50; 30 and 800; 10; 10

Spectral range: 450 – 12500 for Landsat and 349.896-2582.28 for Hyperion

Satellites: Hyperion, Landsat 8

Bands: 8 for Landsat 8 and 242 for Hyperion

Applications: 1. for agriculture
2. For crop parameters, crop yield and land management
3. Forestry
4. for vegetation
5. For vegetation biomass, cellulose calculation, lignin calculation, starch, stress ratio, vitality and water index

2. Modified Triangular Vegetation Index (MTVI) 1 = 1.2*[1.2*(\(\rho_{800} - \rho_{550}\)) – 2.5*(\(\rho_{670} - \rho_{550}\))] = 1.2*[1.2*(B45 – B20) – 2.5* (B 32 - B 20),] for Hyperion =1.2*[1.2*(NIR – GREEN) – 2.5* (RED - GREEN)]

Applications: 1. Vegetation
Wavelengths: 550, 670, 800
Spectral range: 450 – 12500 for Landsat and 349.896-2582.28 for Hyperion
Bands: 8 for Landsat and 242 for Hyperion

Applications: for vegetation

3. Modified Triangular Vegetation Index (MTVI) 2 = \( \frac{1.5*[1.2*(B45-B20)-2.5*(B32-B20)]}{\sqrt{(2*B45+1)^2-(6*B45+5+\sqrt{B22})-0.5}} \) ( for hyperion) = \( \frac{1.5*[1.2*(NIR-GREEN)-2.5*[RED- GREEN]]}{\sqrt{(2*NIR+1)^2-(6*NIR+5+\sqrt{RED})-0.5}} \) ( for Landsat )

Wavelengths: 550, 670, 800
Spectral range: 450 – 12500 for Landsat and 349.896-2582.28 for Hyperion
Bands: 8 for Landsat and 242 for Hyperion

Applications: for vegetation
4. Renormalized Difference Vegetation Index (RDVI) = \(( \rho_{800} - \rho_{670} )/ \sqrt{ \rho_{800} + \rho_{670} } = (B45 - B32)/ \sqrt{B45 + B32}\) for Hyperion = \((NIR - RED)/ \sqrt{NIR + RED}\) for Landsat

Applications: vegetation, vegetation-chlorophyll

5. Modified Chlorophyll Absorption in Reflectance Index 1 (MCARI) 1 = 1.2*\[2.5*( \rho_{800} - \rho_{670} ) - 1.3*( \rho_{800} - \rho_{550} )\] = 1.2*\[2.5*( \rho_{800} - \rho_{670} ) - 1.3*( \rho_{800} - \rho_{550} )\] for Hyperion = 1.2*\[2.5*( NIR - RED ) - 1.3*( NIR - Green )\] (for Landsat)

Applications: vegetation, vegetation chlorophyll

6. Modified Chlorophyll Absorption in Reflectance Index 2 (MCARI) 2 = 1.5*\[1.2*\( \rho_{800} - \rho_{550}\) - 1.3*\( \rho_{670} - \rho_{550}\)] / \(\sqrt{2*(\rho_{800} + 0.5*\sqrt{\rho_{670}}) - 0.5}\) (for Hyperion) = 1.5*\[1.2*\(\sqrt{\rho_{NIR} - RED}\) - 1.3*\(\sqrt{\rho_{GREEN}}\)] / \(\sqrt{2*(\rho_{NIR} + 0.5*\sqrt{\rho_{RED}}) - 0.5}\) (for Landsat 8, 7 etm+)

Applications: vegetation, vegetation chlorophyll

7. Soil Adjusted Vegetation Index (SAVI) = (1 + L) * \(\rho_{800} - \rho_{670}\) ) / (\(\rho_{800} + \rho_{670} + L\) ) = (1 + L) * \(B45 - B32\) ) / (\(B45 + B32 + L\) ) (for Hyperion) = (1 + L) * \(NIR - RED\) ) / (\(NIR + RED + L\) ) (for Landsat 7 etm+, 8)

Here, value of L = 0, 5 and between (-0, 9 and 1.6)

Applications: 1. Vegetation- biomass

2. Agriculture - crop parameters
3. Soil
4. Vegetation

8. Transformed Soil Adjusted Vegetation Index (TSAVI) = a(NIR - aRED - b)/(NIR+aRED - b) where, a= soil and b= intercept of soil = B(B95 - B35 - A)/(B35+35+B95-A) + X(1+B^2) (for Hyperion) = B(NIR - B. Red - A)/(Red+B(NIR - A) + X(1+B^2)) (for Landsat)

Applications: soil, vegetation, vegetation-biomass, vegetation – LAI

9. Modified SAVI with self-adjustment factor L (MSAVI) = \(\frac{1}{2} \left[ 2* \rho_{800} + 1 - \sqrt{2 * \rho_{800} + 1}^2 - 8 * (\rho_{800} - \rho_{670}) \right] = \frac{1}{2} \left[ 2* \sqrt{\rho_{NIR} + 1} - \sqrt{2 * \rho_{NIR} + 1}^2 - 8 * (\rho_{NIR} - RED) \right] \) (for Hyperion) = \(\frac{1}{2} \left[ 2* \sqrt{\rho_{NIR} + 1}^2 - 8 * (\rho_{NIR} - RED) \right] \) (for Landsat 7 etm+, 8)

10. Optimized Soil Adjusted Vegetation Index (OSAVI) = \((1 + Y) * (\rho_{800} - \rho_{670}) / (\rho_{800} + \rho_{670} + Y)\) here we take Y = 0.16 = \((1 + 0.16) * (\rho_{800} - \rho_{670}) / (\rho_{800} + \rho_{670} + 0.16)\) for general formula,

\(= (1 + 0.16) * \frac{(B45 - B32)}{(B45 + B32 + 0.16)}\) (for Hyperion) = \((1 + 0.16) * (\rho_{NIR} - RED) / (\rho_{NIR} + RED + 0.16)\) (for Landsat 7 etm+, 8)

Applications: 1. Vegetation

11. Photo chemical reflectance index (PRI) = Leaf Chlorophyll Index PRI1 = \(\frac{(\rho_{528} - \rho_{567})}{(\rho_{528} + \rho_{567})}\)
PRI = \( \frac{(\rho_{531} - \rho_{570})}{(\rho_{531} + \rho_{570})} = \frac{(850 - 710)}{(850 + 680)} = \frac{B71 - B36}{B71 + B33} \) (for Hyperion)

<table>
<thead>
<tr>
<th>Wavelengths</th>
<th>Application</th>
</tr>
</thead>
</table>
| 680, 710, 850 | 1. Agriculture 
2. Vegetation 
3. Vegetation – cellulose |

12. Cellulose Absorption Index (CAI) = \( 0.5*(\rho2000 - \rho2200) \) – \( \rho2100 = 100(0.5*(2030nm - 2210nm) - 2100nm) = 100(0.5*(B188 - B206) - B195) \) (for Hyperion)

<table>
<thead>
<tr>
<th>Wavelengths</th>
<th>Application</th>
</tr>
</thead>
</table>
| 680, 710, 850 | 1. Agriculture 
2. Vegetation 
3. Vegetation – cellulose |

13. Normalized Difference Water Index (NDWI) = \( \frac{(\rho_{860} - \rho_{1240})}{(\rho_{860} + \rho_{1240})} = \frac{B72 - B109}{B72 + B109} \) (for Hyperion)

<table>
<thead>
<tr>
<th>Wavelengths</th>
<th>Application</th>
</tr>
</thead>
</table>
| 860, 1240 | 1. Vegetation 
2. Vegetation-water |

14. Short wave Infrared water stress Index (SIWSI) = \( \frac{(\rho_{6} - \rho_{2})}{(\rho_{6} + \rho_{2})} = \frac{800 - 560}{1060 + 680} = \frac{B45 - B20}{B151 + B33} \) (for Hyperion)

<table>
<thead>
<tr>
<th>Wavelengths</th>
<th>Application</th>
</tr>
</thead>
</table>
| 550, 680, 800, 1660 | 1. Agriculture 
2. for crop parameters 
3. for crop management precision 
4. Vegetation stress 
5. for vegetation water calculation 
6. for vegetation water stress calculation |

15. Plant Water Index (PWI) (Simple Ratio, Water Band Index (WBI), Water Index (WI)) = \( \frac{\rho_{970}}{\rho_{900}} = \frac{B83}{B76} \) (for Hyperion)

<table>
<thead>
<tr>
<th>Wavelengths</th>
<th>Application</th>
</tr>
</thead>
</table>
| 970, 900 | 1. Vegetation 
2. for vegetation water calculation 
3. for vegetation water stress calculation |
GPS location ping for Agricultural Loan

In most 3 world country farmers are not rich enough to afford a smart phone. Hence, confirming the location of the farm becomes a tedious process we have automated the agriculture loan using remote sensing data whose basic input parameter is latitude & longitude of field using which we capture the image of the field in 3 spectrums:

1. Visual inspection. i.e., picture- we use the same data source Google Earth users and using an Artificial Intelligence algorithm we classify that area into the following buckets:
   - Water
   - Urban landscape, field
   - Field with vegetation
   - Forests
   - Ice

2. Land sat data analysis where we have information in a 30 * 30 meter resolutions in 7 bands of spectral data using which we calculate various parameters like NDVI and chlorophyll content of the field which gives us an indication of the amount of possible yield from that field.

3. Our 3rd layer of AI algorithm analysis is on data derived from the Hyperion sensor from the EO1 satellite of Nasa which has hyperspectral imaging using 240 distinct bands. this gives us a higher amount of information which we can use for analysis.
   - GSM Module
   - GPS receiver
   - Micro controller
   - 10 USD for 5000 units

References


- Baijal Arpita, “High Ratio of Agriculture NPAS In Priority Sector Lending By Public And Private Banks In India – Reasons, Suggestions And Discussions”, International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

- E01 – Nasa Datasets
- Google Earth Images