AI Driven Credit Scoring Model for Farmers

Problem Statement: Credit scoring for farmers is a lot more different and demanding when compared to credit scoring of financial institutions or businesses. Challenges are many due to low level of financial inclusion of farmers and their limited access to credit. This is compounded by lack of structured data and related information with respect to farmers. This project aims to address this problem by first identifying relevant data that can help in coming up with a roadmap of quantifying their credit risk. Subsequently, appropriate machine learning techniques will be utilized to come up with a credit scoring model for farmers. This study is also expected to provide insights on promoting financial inclusion to farmers.

Major Challenges:

1. **Data Verification:** The data provided by them is valid or not is one of the biggest concerns and this needs to be verified by other available relevant sources.
2. **Natural Disaster:** This is tough to overcome but can be mitigated by using meteorological prediction.
3. **Misuse:** One could misuse by taking advantage of various schemes. This needs check.
4. **Digitalization:** Though digitalization is a boon, adapting to new technologies may be tough for farmers in the initial phase.
5. **Lack of Education:** Lack of education may be a major impediment. This requires continuous training.

Broad Data Heads: To come up with a credit score, we need to understand various aspects about the farmers and related details. For that, we need information about the following:

1. Demographics
2. Farming details
3. Capital Strength
4. Additional Information

Data Requirements: On each of the above points, the data required is elaborated below:

Demographics:

- Age
- Marital status
- Family Background- Number of members in the family, their educational background
- Farmer's standing or reputation
- Experience as a farmer
- Address
- Years of stay

Farming Details:

- Land area- Bigger the land area, more will be the production.
- Type of soil- Better the type of soil, better will be the production. Also, different types of crops grow in different type of soil and their selling prices are also different. This will directly impact the production value.
- Location-This will help us to determine the nature of the region.
- Production history- Better the production history, better is the cash flow which is directly proportional to the credit score.
- Water access- An essential ingredient for crops to grow.
- Machineries and technologies available-This will help us determine the way the production is done.
- Crops grown by the farmer
- Selling price of crops

Capital Strength:

- Average monthly income
- Bank statement
- Loan frequency
- Loan amount
- Reason for taking loan
- Available assets valuation

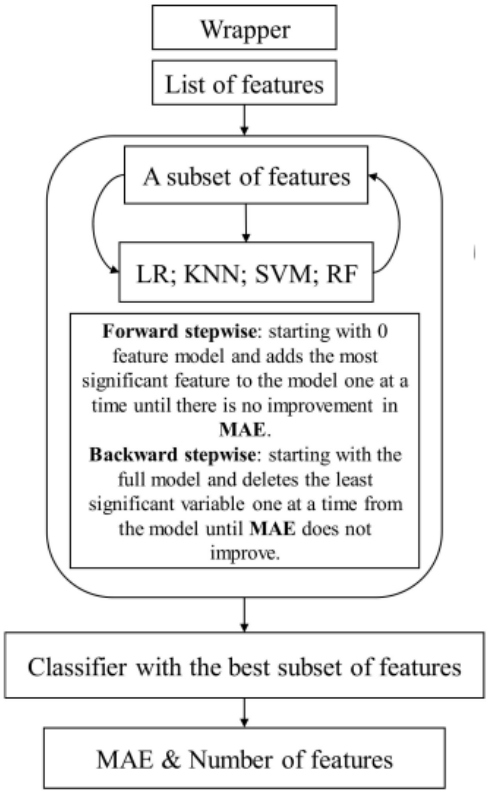
- Recently sold any asset or not

Additional Information:

- Government incentives
- Yield prediction
- Future price prediction of the selected crop
- Details of Farming Partners
- Health issues
- Other income sources
- Average monthly expenditure

Techniques that could be applied:

1. Logistic Regression
    - Forward
    - Backward
    - Stepwise
2. KNN : K Nearest Neighbours
3. Support Vector Machines
4. Ensemble Methods
    - Bagging
    - Boosting
    - Random Forests



Solution Approaches: Credit scoring is a classification problem between the identification of defaulters and non-defaulters. Both SVM and RF are black-box models and are sensitive to hyperparameters. Researchers proposed a modified harmony search random factor that is more robust in terms of performance, explain ability, and computational time. However, it is not completely known that AI-ML algorithms will not cause bias especially against minorities like small and marginal holdings, women communities, etc., Hence, to reduce the impact of human biasness it is necessary to train the data appropriately so that more robust AI-ML algorithms can be formed. Referring to the below table of the comparative analysis of credit scoring techniques, it can be stated that the hybrid

model (whether it is AI-ML with AI-ML-based or AI with any other method of credit scoring) could be the best fit for credit score assessment, whereas logistic regression yields a lesser impact on credit scoring. This analysis has been performed by assigning the weightage (sum of weightage is one) to the below given four parameters: accuracy (0.3), performance (0.3), robustness (0.2), and volume of data (0.2). This weightage has been assigned on the basis of the importance of these parameters. Accuracy and performance are the two main features of any ML-based model, then comes the almost equal weightage of robustness and the size of the data handling. The rating has been assigned from 1 (very low performance) to 5 (very high performance), depending on how these models performed and what the results have been interpreted as after having the empirical studies of the existing literature.

Parameters	Comparative Analysis—Credit Scoring Techniques												
	Weights	ANN		SVM		RF/XG Boost		Logistic Regression		GA		Hybrid Model	
		Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score
Accuracy	0.30	4	1.2	4	1.2	5	1.5	3	0.9	4	1.2	4	1.2
Performance	0.30	4	1.2	3	0.9	5	1.5	3	0.9	4	1.2	4	1.2
Robustness	0.20	3	0.6	3	0.6	3	0.6	3	0.6	4	0.8	5	1
Volume of Data	0.20	3	0.6	3	0.6	3	0.6	2	0.4	3	0.6	5	1
Total	1.00		3.6		3.3		4.2		2.8		3.8		4.4

Clustering Farmers:

The data could be analysed to look for similar behaviour of farmers by applying clustering technique. This will help in getting insights about offering different loan offers for different groups of farmers. This eventually can be used to update their credit score.

Available Data Sources on the web:

- Data on Drought in different areas: https://farmer.gov.in/Drought/Droughtreport.aspx
- Data on soil fertility in different areas: https://farmer.gov.in/soilfertilitymaps.aspx
- Farmer's Data:
  1. Tripura: Tripura.xlsx - Google Drive
  2. Meghalaya: Meghalaya.xlsx - Google Drive
  3. Mizoram: Mizoram.xlsx - Google Drive
  4. Nagaland: Nagaland.xlsx - Google Drive
  5. Manipur: Manipur.xlsx - Google Drive

Deliverables:

By analysing the data using appropriate machine learning techniques, we can come up with an efficient credit scoring model that assigns a credit score to farmers or rank the farmers as per their potential to repay the loan thus mitigating the risks involved in lending.