

LTF Challenge- Farmer Income Prediction

Launch date – July 23, 2025

Problem Statement: Predicting Farmer Income in India

1. Problem Background

Access to credit is the key to attaining full economic potential. People employed in farming and allied activities struggle to get loans due to insufficient or non-existent credit histories. As a result, this population is often exploited by untrustworthy lenders.

L&T Finance (LTF) aims to broaden financial inclusion for the unbanked population by providing a positive and safe borrowing experience. To ensure a safe borrowing experience, LTF utilizes several data sources to assess the creditworthiness of the population. In addition to traditional credit data vendors, we leverage weather monitoring datasets, soil quality data, crop yield, commodity price dynamics, and land records to predict their client's ability and willingness for repayment.

Although LTF uses state-of-the-art models that leverage statistical and machine learning models along with emergent GenAI literature to profile and predict the creditworthiness of the farming population. We challenge economic modelling community to help us unlock the full potential of our data assets and raise our methods of assessing creditworthiness to a level where the loan application of a farming professional worthy of repayment should not get rejected. We also encourage the teams to acquire and use alternate public or private data sources that may help improve the farmer income prediction model.

2. Potential Data Resources

We have a farmer income dataset which we are collecting from our existing customers, the dataset is mapped with farmer demography, land holding, local climatic profile and indicator to present local living index. We do not claim that this is the exhaustive information related to farmers, but it can be enriched using alternative data resources. For the help of the participants, we provide here a list of example data sources that can help improve the model performance which participants are free to utilize to improve the predictive power of their empirical models.

- a. Farmer income data
- b. Weather: Temperature, humidity, rainfall, sunshine hours, wind speed, thunderstorm, satellite data
- c. Land holding and soil quality: Soil-seed selection, Soil nutrients selection (fertilizers, manures, soil culture)
- d. Crop: Area planted, crop type, yield
- e. Market Price: Price dynamics of major crops in the market

- f. Farmer Information: Location, age, education, farm size, market access, farmer psychometrics
- g. Public intervention: Govt policies, macroeconomic indicators
- h. Miscellaneous: Market crash, Crop failure, Tariff barriers to trade, Any other data

3. Team Composition and Participation (can be adopted to the format which hackathon organizer are recommending)

The challenge participants are tasked with predicting the income of farmers in India based on the above data sources. The participants are free to form their collaborative groups as the group participation is allowed in the competition. There is no participation fee for the competition. The questions related to competition can be addressed to pgdba.conclave@iimcal.ac.in.

4. Competition Ethics and Code Ownership

The challenge participants are encouraged to come with their original ideas and LTF do not take ownership of any copyright violations, plagiarism allegations and conflict of interests against any of the participants of the competition. The participants are instructed to reference the article whenever they are importing ideas from the scholarly articles, blogs, tutorials etc. At max five submission per team is allowed. Participant are suggested to use **Python** as their preferred language although we welcome original ideas in any other coding environment that can be reproduced at our end. The evaluation of submitted solution will be objectively based on **MAPE (mean absolute percentage error)** LTF will own the copyright of the empirical methods developed by participants once they accept the winning amount. Kindly read the terms and conditions document for further instructions.

5. Data Description and Submission Criteria

Download the dataset, which includes:

- **Training Data:** Farmer demography, landholding, local climatic conditions, and living index indicators.
- **Dataset for Prediction:** For testing your model.
- **Data Dictionary:** Describing all variables in detail.
- **The link to the dataset:** https://drive.google.com/drive/folders/1Wp1eGu2eOkXZdKvzBQIF-AfTC0XxabpD?usp=drive_link
- **The sample prediction file** has been provided for your reference.

Submission Requirements

We recommend submitting the following:

- A **working python code (.py)** to predict farmers' income.
- A **presentation** outlining your methodology and modeling steps in **PDF format** (max. **10 slides** excluding introductory and Conclusion slides)
- The **model's performance** on the test dataset in the prescribed format, (which will be evaluated using the **Mean Absolute Percentage Error (MAPE)** metric)

Note: The primary objective is to evaluate your thought process and problem-solving skills. While model accuracy is important, it is secondary to your approach and methodology.