IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

Indian Farm Distress Early Warning System – Phase 1

Venumadhava M, Bareera Sadaf, J Mythri, Sabah Iffath, Shaik Shaguftaj
Department of Computer Science - Artificial Intelligence and Machine Learning
Proudhadevaraya Institute of Technology, Hospet, Karnataka, India
Corresponding author: sabahiffath04@gmail.com

Abstract : *Agricultural distress in India can arise from the combined effects of rainfall variability, changes in crop production, fluctuations in agricultural market prices and difficulties associated with agricultural credit. Although information relating to these factors is available through different public institutions, the data are maintained in separate systems and may differ in format, coverage and frequency of update. This Phase I paper proposes the design and development of an Indian Farm Distress Early Warning System that will integrate four categories of public data: rainfall information from the India Meteorological Department (IMD), crop production and yield information from the Ministry of Agriculture, market-price information from AGMARKNET, and agricultural credit information from NABARD. The proposed workflow will prepare and combine the selected indicators, calculate a District Distress Score for the districts included in the study, apply K-Means clustering (with four clusters) for risk grouping, and use a Random Forest model to forecast next-season risk. Outputs are planned to be presented through a Streamlit web dashboard and a Power BI analytical report. This paper focuses on the problem context, objectives, related work, proposed data sources, system design and evaluation plan. It does not report implementation results or predictive-performance values; these will be examined in Phase II.*

Keywords: agricultural distress; early warning system; district-level analysis; K-Means clustering; Random Forest; India..

I. INTRODUCTION

Agriculture is an important source of livelihood in India. Farmers may experience economic pressure due to irregular rainfall, drought, crop losses, changes in market prices and difficulties in agricultural credit repayment. These factors may interact. For example, inadequate rainfall can affect crop production, while unfavourable market prices can reduce the income available to meet cultivation expenses and repayment obligations. Monitoring only one factor may therefore provide an incomplete understanding of agricultural risk in a district.

Several public institutions publish information relevant to these conditions. The India Meteorological Department (IMD) provides rainfall information; the Ministry of Agriculture publishes crop production and yield information; AGMARKNET provides agricultural market-price information; and NABARD publishes agricultural credit-related information. However, these sources are maintained separately and may use different formats, update schedules and geographical coverage. A common analytical framework could help bring selected indicators together for district-level monitoring.

The project will also consider the Farmers' Distress Index (FDI) developed by ICAR-CRIDA. As described in the project study, this approach uses information from newspapers and social media, telephonic surveys and multiple dimensions of distress. The proposed system is intended to complement such assessment by exploring an automated approach based on selected public datasets. It will not replace field surveys or direct consultation with farmers.

This Phase I paper presents the proposed approach for developing an Indian Farm Distress Early Warning System. The project is planned to combine four categories of government data, derive a district-level distress score, group districts

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

using K-Means clustering, forecast next-season risk using Random Forest, and present the analysis through accessible dashboard interfaces.

II. PROBLEM STATEMENT

Information that may indicate agricultural distress—including below-normal rainfall, changes in crop yield, market prices falling below the Minimum Support Price (MSP), and agricultural credit stress—is available through different public data sources. However, the project identifies a gap in the availability of an integrated, automated and publicly accessible district-level platform that combines these signals. Separate data formats and update frequencies can make it difficult for users to examine the indicators together and identify districts that may require closer attention. The proposed project will address this gap by designing a data-driven platform for integrated monitoring and early warning.

III. OBJECTIVES

- To identify and study key indicators associated with agricultural distress in India.
- To collect selected public datasets relating to rainfall, crop production or yield, market prices and agricultural credit.
- To prepare and integrate the selected datasets at district level, subject to data availability and coverage.
- To design a District Distress Score using the selected indicators.
- To apply K-Means clustering with four clusters to group district-level observations.
- To develop a Random Forest-based approach for forecasting next-season risk.
- To design a Streamlit web dashboard and a Power BI analytical report for presenting the planned outputs.

IV. REVIEW OF RELATED WORK

Reddy, Bhattacharya and Reddy (2021) presented a Farmers' Distress Index intended to support an action plan for reducing vulnerability in the drylands of India. The project report describes this approach as using telephonic surveys of 640 dryland farmers and seven dimensions of distress. This work is relevant because it demonstrates the multidimensional nature of farmer distress. The proposed project will explore a complementary approach based on the integration of public datasets.

Ramesh and Vishnu Vardhan (2017) examined agricultural data using data-mining techniques and applied K-Means clustering to district-level rainfall data in Andhra Pradesh. Their work provides context for using clustering to identify groups of agricultural observations with similar patterns. The proposed project will examine K-Means clustering across selected distress-related indicators.

Sujatha et al. (2024) studied machine-learning-based crop-yield prediction in South India and included Random Forest in their model analysis. Chlingaryan, Sukkarieh and Whelan (2018) reviewed machine-learning approaches for crop-yield prediction and nitrogen-status estimation in precision agriculture. These studies provide background for considering machine-learning techniques with agricultural and environmental data.

Patel and Patel (2022) discussed statistical machine-learning techniques for agricultural crop-yield prediction. The project report refers to data normalisation in connection with combining indicators measured in different units. The proposed work will examine an appropriate normalisation procedure during data preparation.

V. PROPOSED SYSTEM AND METHODOLOGY

Proposed Data Sources

The project proposes to use four categories of public data. The final selection of datasets will depend on their availability, completeness, time period and district-level coverage.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

TABLE I: Proposed Data Sources

Proposed source	Information to be examined	Purpose in the system
India Meteorological Department (IMD)	Rainfall information and rainfall departure	Examine climatic stress signals.
Ministry of Agriculture	Crop production and yield information	Examine changes in agricultural output.
AGMARKNET	Agricultural market prices and relevant MSP comparisons	Examine possible market-price stress.
NABARD	Agricultural credit-related information, including available NPA indicators	Examine credit-related stress.

Data Collection and Integration

The selected datasets will be obtained from the respective public sources. The data will be examined for missing values, duplicate records, inconsistent district names, differences in units and variations in reporting periods. Python libraries such as pandas and NumPy are planned for data loading, cleaning and transformation. District identifiers and time periods will be aligned as far as the source data permit. Records with insufficient coverage will be documented, and the treatment of missing values will be specified in the detailed implementation.

District Distress Score

A District Distress Score will be designed to combine the selected indicators into a common district-level measure. Since the indicators may use different units and scales, the project will examine the use of normalisation, including MinMaxScaler. The score construction will define the indicators, their direction, any weights applied and the method used to handle missing data. The final formula and its justification will be documented after the data and feature definitions have been finalised.

K-Means Clustering for Risk Grouping

K-Means clustering with K = 4 is proposed to group districts according to similarities in their selected distress indicators or derived score-related features. The resulting groups will be examined to understand the patterns represented by each cluster. The final interpretation and risk labels will be assigned only after reviewing the cluster profiles; the clusters will not be assumed to represent fixed severity levels without supporting analysis.

Random Forest for Next-Season Risk Forecasting

A Random Forest model is proposed for forecasting next-season agricultural distress risk. The project will identify an appropriate target variable and predictor features based on the available historical data. The data will be divided into training and testing sets using a procedure appropriate to the time structure of the observations. The model will then be trained and evaluated using suitable metrics. The target definition, feature set, model settings and validation method will be documented as part of the implementation.

Dashboard and Visualisation

The project plans to use Streamlit to develop a web-based interface for displaying the system outputs. Power BI Desktop will be used to prepare an analytical report, planned as a five-page dashboard. The interfaces are intended to make district-level information easier to explore through visual summaries and relevant filters. SQLite is planned for database management, and joblib may be used to save and load the trained machine-learning model.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-41002

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

VI. PROPOSED TECHNOLOGY STACK

TABLE II: Proposed Technology Stack

Technology / tool	Planned use
Python 3.12	Core programming language for data processing and machine learning.
Visual Studio Code	Development environment for Python scripts.
pandas and NumPy	Data loading, cleaning, transformation and numerical operations.
scikit-learn	K-Means clustering, Random Forest and data normalisation.
Streamlit	Development of the web application.
Power BI Desktop	Preparation of the analytical dashboard.
SQLite / DB Browser for SQLite	Database storage, viewing and querying.
Matplotlib and Seaborn	Charts and statistical visualisations.
joblib	Saving and loading the trained model.

VII. PROPOSED EVALUATION PLAN

In Phase II, the completed system will be evaluated using the prepared district-level dataset and the implemented models. The evaluation will examine data coverage and quality, the distribution and interpretability of the District Distress Score, the characteristics of the four K-Means clusters, and the predictive performance of the Random Forest model. If the forecasting task is formulated as classification, suitable measures such as precision, recall, F1-score and accuracy will be considered. If it is formulated as a regression task, measures such as mean absolute error and root mean squared error may be appropriate. The final choice will depend on the target variable and model design. The dashboard will also be checked for whether users can access and interpret the intended district-level information.

VIII. EXPECTED OUTCOMES

A prepared dataset combining selected rainfall, crop, market-price and agricultural-credit indicators for the districts covered by the study.

A documented method for calculating a District Distress Score.

Four K-Means clusters that can be examined for similarities in district-level distress indicators.

A Random Forest model designed to forecast next-season risk, subject to available historical data and validation.

A Streamlit interface and a Power BI analytical report for presenting the system's outputs.

These are planned deliverables, not achieved results. Their completion and performance will be reported in the Phase II paper.

IX. SCOPE AND ANTICIPATED LIMITATIONS

The proposed study will focus on the districts and years for which the selected public datasets provide sufficiently compatible information. Differences in geographical coverage, reporting frequency and data quality may limit the number of districts or periods that can be included. District-level analysis may also hide variation between villages, farm sizes, crops and individual households. The system is therefore intended as a decision-support and monitoring tool. Any warning will require careful interpretation and, where possible, validation through field-level information and expert assessment.

X. CONCLUSION

This Phase I paper has outlined the proposed design of the Indian Farm Distress Early Warning System. The project will explore the integration of rainfall, crop production or yield, agricultural market-price and agricultural-credit information to support district-level monitoring. It proposes to develop a District Distress Score, apply K-Means clustering with four clusters, and use Random Forest for next-season risk forecasting. Streamlit and Power BI are planned as the presentation layers. The next phase will focus on implementation, model evaluation, analysis of district-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

level outputs and reporting of the actual results. The system is intended to complement existing agricultural monitoring and field-based assessment rather than replace them.

ACKNOWLEDGMENT

The authors acknowledge Dr Venumadhava M, the Department of Computer Science – Artificial Intelligence and Machine Learning and Proudhadevaraya Institute of Technology for their guidance and support.

REFERENCES

[1]. Singh, A., & Trivedi, P. (2015) studied the agrarian crisis in India and identified major reasons for farmer distress, including increasing cultivation costs, farmer indebtedness, declining public investment, and unfavourable government policies.

[2]. Ramesh, D., & Vishnu Vardhan, B. (2017) analysed district-level rainfall patterns in Andhra Pradesh using the K-Means clustering algorithm. They grouped rainfall data into four clusters to identify rainfall patterns and their relationship with crop production.

[3]. Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018) reviewed various machine learning techniques for crop yield prediction and nitrogen status estimation in agriculture. Their study highlighted the effectiveness of ensemble methods such as Random Forest in analysing weather, soil, and other agricultural data.

[4]. Nishant, P. S., Venkat, P. S., Avinash, B. L., & Jabber, B. (2020) proposed a crop yield prediction system using machine learning techniques such as Lasso, Kernel Ridge, and Elastic Net. They also used stacking regression to improve the accuracy of crop yield predictions.