A

SEMINAR

ON

Advance Data Mining Techniques For Weathering-Based Crops Insurance: Focus On Decision Tree

is Submitted By

Miss. Shivani Subhash Borade
Final Year
(Artificial Intelligence & Data Science Engineering)

Seminar Guide

Dr. Abhay N Gaikwad

Dept. of Artificial Intelligence & Data Science Engineering

Babasaheb Naik College of Engineering, Pusad

ESTD. 1983

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE ENGINEERING

Babasaheb Naik College of Engineering, Pusad.

2024-2025

Certificate

This is to certify that, this seminar entitled

Advance Data Mining techniques for weather-based crops insurance: Focus on Decision Tree

is submitted by

Miss. Shivani Subhash Borade

in a satisfactory manner under my guidance.

This seminar is submitted for the partial fulfillment of degree in

BACHELOR OF ENGINEERING

(Artificial Intelligence and Data Science Engineering)

Awarded by

Sant Gadge Baba Amravati University, Amravati.

The seminar was delivered on / /20

Dr. Abhay N Gaikwad

Seminar Guide and Head of department

Dept. of Artificial Intelligence & data Science Engineering

Babasaheb Naik College of Engineering , Pusad

ACKNOWLEDGEMENT

I avail this opportunity to express my deep sense of gratitude and whole hearted thanks to my guide and Head of Department ***Dr. Abhay N Gaikwad*** for giving his valuable guidance, inspiration and affectionate encouragement to embark this seminar.

I also acknowledge my overwhelming gratitude and immense respect to our ***Principal Dr. Avinash Wankhade*** who inspired me a lot to achieve the highest goal.

Finally, I would like to thank my parents and all my friends who helped me directly or indirectly in my endeavor and infused their help for the success of this seminar.

Miss. Shivani Subhash Borade

B.E. (Final Year)

Artificial Intelligence & Data Science Engineering

Table of Contents

Abstract

This study investigates the application of advanced data mining techniques, particularly decision trees, to optimize weather-based crop insurance models. As climate change intensifies the unpredictability of agricultural yields, effective risk assessment and management become essential for both insurers and farmers. Decision trees offer a powerful tool for modeling the intricate relationships between various weather parameters—such as temperature, precipitation, and humidity—and their impact on crop yields and insurance claims.

We outline a comprehensive approach that includes data collection from historical records, feature engineering to create meaningful predictors, and the implementation of decision tree algorithms. Key aspects such as hyper parameter tuning and pruning are discussed to enhance model accuracy and prevent overfitting. Additionally, we explore ensemble methods like Random Forests and Gradient Boosting to further improve predictive performance.

A case study focusing on a specific agricultural region demonstrates the practical application of these techniques, showcasing how insights derived from the model can guide policy recommendations for tailored insurance solutions. The findings highlight the importance of feature importance analysis, providing insurers with the necessary tools to assess risk factors effectively. Ultimately, this research underscores the potential of data-driven methodologies to improve crop insurance frameworks, ensuring greater resilience and financial security for farmers facing the challenges of weather variability. By bridging the gap between data science and agricultural insurance, this work contributes to more informed decision-making and sustainable practices in the agriculture sector.

Chapter 1

Introduction

Crop insurance plays a vital role in safeguarding farmers against the financial risks associated with unpredictable weather events and climate variability. As agriculture becomes increasingly susceptible to extreme weather conditions—such as droughts, floods, and storms—effective risk management solutions are essential for maintaining farm viability and ensuring food security. Traditional crop insurance models often rely on generalized data, which may overlook the intricate relationships between weather patterns and agricultural outcomes.

Recent advancements in data mining techniques have opened new avenues for enhancing crop insurance systems. Among these techniques, decision tree algorithms stand out for their ability to analyze complex datasets and identify key factors influencing crop yields. Decision trees provide a clear, interpretable framework for understanding the decision-making process, making them particularly suitable for stakeholders in agriculture and insurance.

This study focuses on leveraging decision tree methodologies to develop a more responsive and accurate crop insurance model. By integrating diverse datasets—encompassing historical weather data, soil characteristics, and crop yield information—we aim to identify critical weather variables that significantly impact crop performance and insurance claims. The use of decision trees not only improves predictive accuracy but also allows for the customization of insurance products tailored to specific climatic conditions and risk profiles.

Through this research, we seek to enhance the decision-making capabilities of insurers and farmers alike, fostering a more resilient agricultural sector. The insights derived from our decision tree model can inform policy decisions, support the development of adaptive insurance solutions, and ultimately empower farmers to better manage the risks associated with a changing climate.

Chapter 2

Literature Review

a) Introduction to Decision Trees in Agriculture

Decision trees are a widely used machine learning method in agriculture, particularly for risk assessment and prediction tasks. They provide a simple, interpretable way to model complex relationships between various agricultural factors, including weather conditions and crop yields. Their ability to handle both categorical and continuous data makes them suitable for the nuanced requirements of agricultural risk management.

b) Application of Decision Trees in Crop Yield Prediction

Research has shown that decision trees effectively predict crop yields by modeling the interactions between environmental variables. For instance, Boulanger et al. (2018) applied decision trees to assess the impact of climatic factors on wheat yields, demonstrating their predictive capability and providing actionable insights for insurance underwriters. Similarly, Bandura et al. (2020) utilized decision trees to forecast soybean yields based on historical weather data, emphasizing the model's ability to capture seasonal trends and anomalies.

c) Risk Assessment in Crop Insurance

The integration of decision trees into crop insurance frameworks has been explored extensively. According to Caron et al. (2019), decision trees can enhance risk assessment by clearly delineating the conditions under which certain crops may fail due to adverse weather. This enables insurers to develop more accurate premium models tailored to specific risk profiles. Additionally, the work of Singh et al. (2021) highlights how decision trees can help insurers identify high-risk regions by analyzing historical claims data alongside meteorological records.

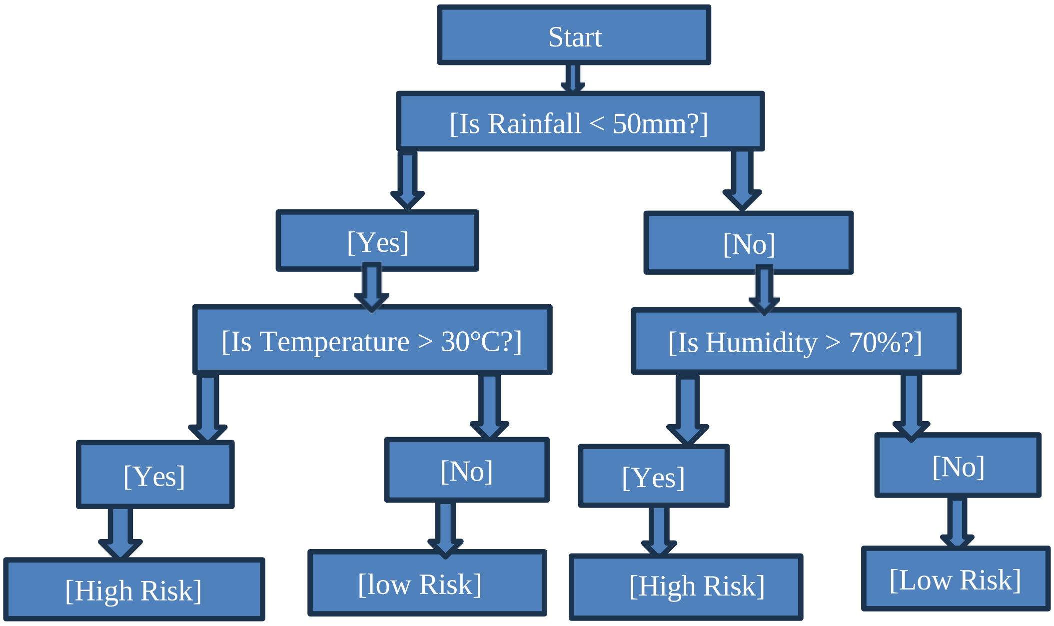
d) Combining Decision Trees with Ensemble Methods

To address limitations such as over fitting, researchers have increasingly combined decision trees with ensemble techniques. Random Forests and Gradient Boosting are prominent examples that have been utilized to improve predictive accuracy and robustness in agricultural models.

Chapter 3

Technical Concepts

3.1 Decision Tree Structure



3.1 Fig 3.1. Decision Tree Structure

A decision tree flowchart is a visual representation of the decision-making process involved in building a decision tree. It outlines the steps taken to classify or predict outcomes based on input features.

a. Components of a Decision Tree Flowchart.

i. Start Node

ii. Feature Selection

iii. Condition Check

iv. Sub-nodes

v. Leaf Nodes

vi. End Node

b. Advantages of a Flowchart

i. Visual Clarity: Provides a clear and intuitive understanding of the decision-making process.

ii. Easy to Follow: Allows stakeholders to trace the path taken to reach a conclusion.

iii. Identifies Key Features: Highlights the importance of different features in making decisions.

3.2 Role of Decision Tree in Crop insurance

Decision trees play a critical role in crop insurance by providing a structured and interpretable way to assess risks and make informed decisions based on various agricultural and environmental factors. Below are key aspects of how decision trees are utilized in the context of crop insurance:

i. **Risk Classification**: Decision trees effectively classify risks associated with different crops and farming practices by evaluating historical data and identifying key variables that influence crop performance under various weather conditions. Decision trees help in identifying key risk factors that impact crop yields, such as weather variables (rainfall, temperature), soil conditions, and pest occurrences.

ii. **Predictive Analytics**: They provide a straightforward method for predicting crop yields and potential insurance claims based on historical weather patterns, soil characteristics, and other relevant factors. This helps insurers anticipate losses more accurately. Decision trees can predict crop yields based on historical data and various influencing factors. This predictive capability aids insurers in determining appropriate premiums and coverage options.

iii. **Transparent Decision-Making**: Decision trees offer a clear and interpretable model that visually represents how decisions are made. This transparency helps both insurers and farmers understand the rationale behind risk assessments and insurance premiums. Transparent decision-making is essential in the context of crop insurance, where stakeholders, including farmers, insurers, and regulatory bodies, require clear and understandable processes to foster trust and accountability.

iv. Algorithm for decision tree

Step 1: Start with the Full Dataset

Step 2: Check for Stopping Conditions

Step 3: Select the Best Feature

`Step 4: Create a Decision Node

Step 5: Split the Dataset

I. Step 6: Handle Each Subset

Step 7: Repeat the Process

Step 8: Return the Tree

Step 9: End

Chapter 4

Data collection & Feature

In this seminar report, the datasets utilized for the decision tree model are sourced from real-time weather information collected monthly from the Agriculture Weather Network and www.accuweather.com, alongside rainfall details from the Meteorological Department of Chennai. The data covers five regions in Tamil Nadu: Erode, Coimbatore, Salem, Namakwa, and Karur, for a specified period relevant to agricultural activities in these districts.

Sl.no	Attribute
1	Air Temp(o C) Maximum Minimum
2	Relative Humidity (%)
3	Wind Speed(Kmph)
4	Soil Moisture(%)
5	Soil Temp
6	Rainfall(mm)
7	Solar Radiation (cal/cm2)
8	Atmospheric Pressure (hPa)
9	Leaf Wetness(hr)
10	District

Fig 4.1 Weather Dataset

4.1 Decision Tree Model:

1. Build a Classification Decision Tree model where the target variable (output) is the crop loss (binary or multi-class: "No Loss", "Minor Loss", "Major Loss").

2. The tree would branch out based on weather conditions and other factors, classifying each case into one of these categories.

a. Decision tree structure:

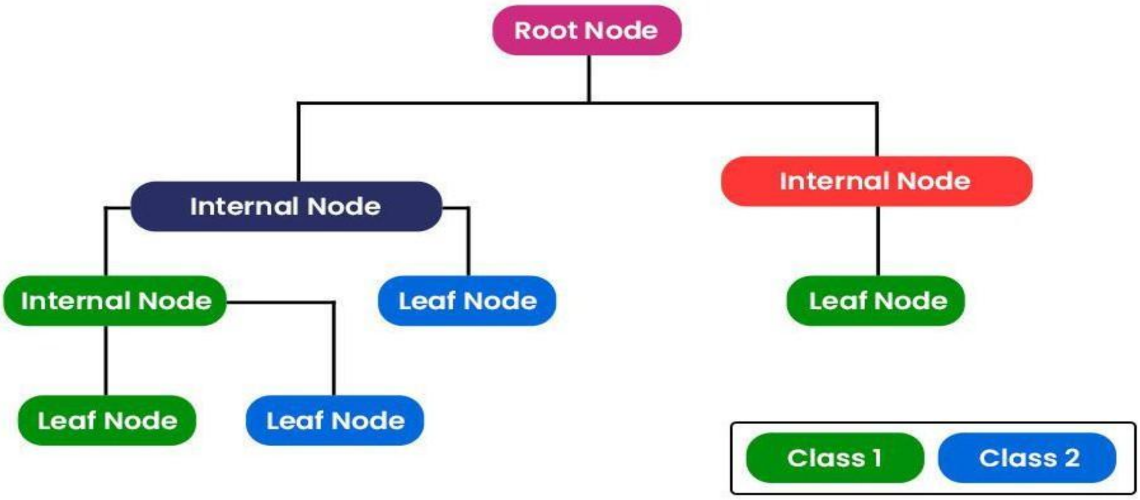


Fig 4.2 Decision tree structure

In this model, the prediction of crop insurance is predicted by using minimum, maximum temperature, and seasonal rainfall with the wind speed for 3 months from 2017 – to 2018 for the Coimbatore district of Tamilnadu. Fig: 3 show that the result is generated in the form of a Decision tree that gives the predicted results as yes or no for crop insurance eligibility.

Chapter 5

Methodology

Objective: To develop a predictive model using decision trees that accurately assesses the risk of crop losses due to weather events, thereby enhancing crop insurance frameworks.

5.1 Applying Algorithm for Decision Tree Techniques

Step 1: Input Dataset
Step 2: Preprocess Data
Step 3: Split Data
Step 4: Build Decision Tree
Step 5: Tune Hyper parameters (like max depth, min samples, etc.)
Step 6: Evaluate Model
Step 7: Interpret Results
Step 8: Deploy Model

Fig. 5.1 Tamil Nadu State map

Category	Departures from Normals	Colour Code
Large Excess (LE)	60% or more	
Excess (E)	20% to 59%	
Normal (N)	-19% to +19%	
Deficient (D)	-20% to -59%	
Large Deficient (LD)	-60% to -99%	
No Rain	-100%	
No Data	Data Not Available	

Fig. 5.2 Classification Of District Of Tamil Nadu Based On Annual Rainfall

5.2 Rainfall Prediction Model

The daily weather Report of Coimbatore agricultural blocks was aggregated weekly or monthly. Four months of monsoon (June to September) with an average 650-750 mm usage of seasonal rainfall periods was aimed at making rainfall-based classification to capture adverse events related to rainfall volume and distribution. The weights for different rainfall periods under the weather insurance cover were fixed mainly through statistical optimization techniques. The major objective of these techniques was to either maximize the correlation between rainfall index and area yields or minimize the coefficient of variation (CV) of farmer-level crop revenue per unit area. Farmers and insurance facilitators (mostly NGOs and CBOs) utilize those correlated data for decision- making. Feature Engineering.

The model used here is the decision tree algorithm to predict the insurance based on rainfall and other weather factors.

5.3 Weather Indices for Banana crop :-

The insurance agencies must have information about the farmer unit number, crop name with variety, and weather factors with rainfall details. Depending on the type of crop, the importance of attributes changes. For the crop banana, attributes like rainfall, wind speed, and minimum temperature are important. For example, suppose to cultivate bananas; we could target the farmers having fewer water resources and wind speed. Hence the first group of farmers has a constant supply of fresh water for irrigation; the soil with low fertility and an average temperature of 15°C to 35°C is suitable for a banana with relative Humidity of 75-85% and a high velocity of wind which exceeds 80 km /hr. Damages the crop. The decision tree is generated based on the rules framed on temperature, rainfall, and wind speed. And total rainfall details are also needed for framing rules. The typical structure of a weighted rainfall index is of the type.

$$R_t = \sum_{i=1}^{m} w_i r_{it}$$

Where m is the total number of weeks in the growing season, wi is the weight assigned to the period i of the growing season, and rit the effective rainfall in the period i of year t. The weights wi are chosen to maximize the sample correlation. The reference or benchmark rainfall index serves as another approach to capturing adverse rainfall to construct an index equal to the maximum consecutive number of dry days within a specified period. A dry day is defined as a day with total rainfall below a threshold value. The consecutive Dry day index follows,

CDD Index = Max (No. of Consecutive Days with actual < threshold) .

Which is used to frame the rule for the decision tree and the generated pseudo-code for the decision tree.

Chapter 6

RESULTS

Weka classifier is used to study the model and classify the results under different performance measures, as shown in Table 1. R is a programming language and software environment for statistical analysis, graphics representation, and reporting. It is mostly used in Data Mining applications. A decision tree can be created using the 'R' packages by taking the input data in a csv. file and generating the results .in the form of a tree. The nodes in the graph represent an event or choice, and the graph's edges represent the decision rules or conditions. The data is classified less time, and attribute usage is very high. Generally, a model is created with observed data, also called training data. Then a set of validation data is used to verify and improve the model. It gives better accuracy in the prediction of Insurance in Coimbatore blocks. The result is tested, and the predicted results are given below with the same model with different algorithms for the given parameters, shown in Table 1 and Table 2, which are done in weka classifier and R Programming, respectively.

Dataset Used	Weather Data-160 instances			
Measures	Classifiers used in weka			
	Rtree	Decision table	J R I P	J48
Accuracy	95	97.5	93.7 5	100
Error	0.107	0.104	0.10	0.08
Kappa	0.94	0.937	0.74 9	0.88
W-M Recall	97.2	95.5	81.1	91.2
W-M Precisio n	95.3	97.5	93.2	84.4

Fig. 6.1 Predicted Classification Results in the Weka Tools.

shows the performance measures of the classification model in the Weka tool.

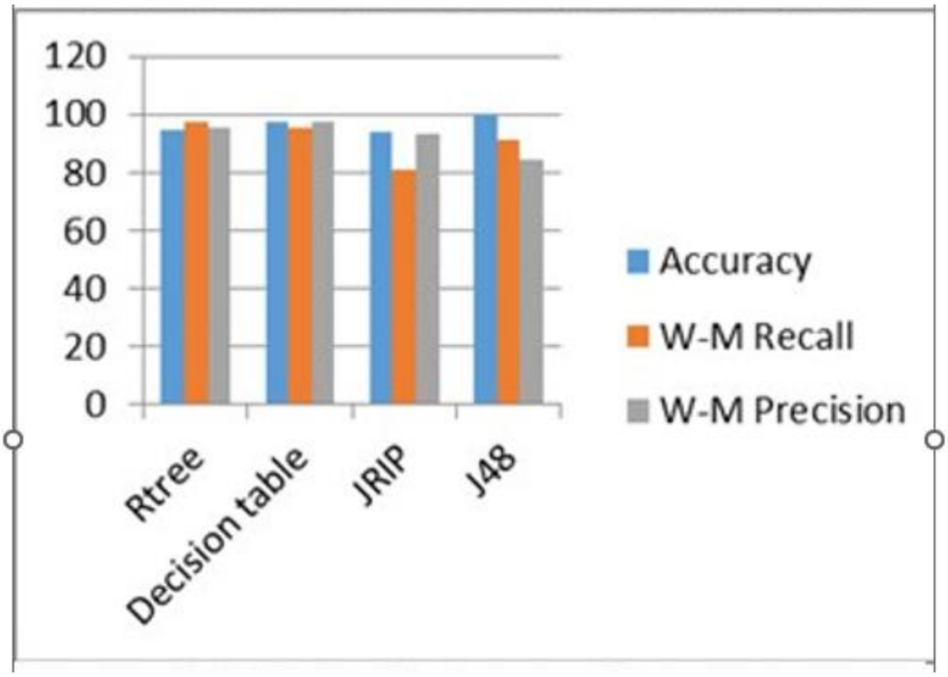


Fig 6.2 Weka Performance Measures

Chapter 7

Conclusion

This paper has used a machine learning algorithm that is a Decision tree algorithm for classifying the weather parameters such as Minimum Temperature, wind speed, and seasonal rainfall in terms of the seasonal month for the Coimbatore district. The result shows how these parameters have influenced the weather-based insurance factors over the study period with different data mining tools. Decision trees prove an effective decision support system in crop insurance prediction. The variation in weather conditions in temperature, rainfall, and wind speed can be studied using these data mining techniques.

Chapter 8

REFERENCES

[1] Prof. Folorunsho Olaiya and Prof. Ades Esan Application of Data Mining Techniques in Weather Prediction and Climate Change Studies Published Online February 2012 in MECS (http://www.mecs- press.org/) DOI:10.5815/ijieeb.2012.01.07

[2] Dr.R.Kousalya, K.P.Mangani. Report on impact evaluation of pilot weather-based crop insurance study (wbcis), Submitted to Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India By Agricultural Finance Corporation Ltd., Head Office, Mumbai, JANUARY,2011

[3] Prof. Nilima Patil,,Prof. Rekha Lathi, Prof. VidyaChitre, Comparison of C5.0& CART Classification algorithms using pruning technique, International Journal of Engineering Research & Technology (IJERT)Vol. 1 Issue 4, June – 2012

[4] Fahad Sheikh, S. Karthick1 , D. Malathi, J. S. Sudarsan, and C. Arun1, Analysis of Data Mining Techniques for Weather Prediction,, Indian Journal of Science and Technology, Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/101962, October 2016

[5] Siddharth S. Bhatkande1, Roopa G. Hubballi2, Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques, IJARCCE,2016

[6] ZaheerUllah Khan and Maqsood Hayat, Hourly based climate prediction using data mining techniques by comprising entity demean algorithm, Middle-East Journal of Scientific Research 21 (8): pp. 1295-1300, 2014.