International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Designing Weather Based Crop Insurance Payout Estimation Based on Agro-Meteorological Data using Machine Learning Techniques

K.P.Mangani , R.Kousalya

Abstract: ***Data mining techniques have been extensively used to mine up-to-date information from agricultural databases. In Agriculture, the Loss Assessment and Estimation in Crop insurance can be done on various factors like yield-based, crop-health based and weather-based variations. Weather-based variations are taken into account to design the insurance payout classifier model for the selected crop within the selected agricultural blocks of Tamilnadu. Then the weather attributes that undergone feature selection are given as input to the model with the rule-based classification algorithm implementing the neighboring approach with a sequential covering strategy named as CBKNN-PAYRULE which is statistically higher than other state-of-the-art rule-based classification algorithms. This model is proposed to classify the agricultural blocks based on the Area-wise Assessment of adverse temperature for the groundnut crop from their nearest neighbor. Then By combining the classified neighboring approach with the threshold factors the Rule-based classifier is done to generate the rules to estimate the insurance payout value as per policymakers for the selected agricultural blocks. Then decision-making techniques are applied to predict the insurance with the possibility of product basis risk, which covers the deviations in weather indices with the risk profile factors for the notified agricultural blocks for the specified crop. Thus the proposed technique can support the simultaneous prediction of the insurance payout to be paid in case of adverse weather factors of the selected crop for five districts with high accuracy and the correlation analysis of weather factors with the payout concerning to each district is also made. The Experimental results show that the proposed work enhances the accuracy in insurance payout prediction of the groundnut crop of the selected districts.***

Keywords: ***Data mining, Sequential Covering Algorithm, Agriculture, Crop Insurance Payout, Class-based-KNN, classification.***

Manuscript published on 30 September 2019
* Correspondence Author
K.P.Mangani*, Ph.D scholar, Department of Computer Science, Dr.N.G.P Arts and Science College, Coimbatore, India, Email: mangani.kp@gmail.com
Dr.R.Kousalya, Professor and Head, Department of Computer Applications, Dr.N.G.P Arts and Science College,, Coimbatore, India, Email: kousalya@drngpasc.ac.in
Insituition Communication No: DrNGPASC 2019-20 CS006

I. INTRODUCTION

The Farmers have shown a substantial interest in Weather Index Insurance (WII) with the basis risk that remains a key challenge for making the crop insurance payout and premium effective. There is a great necessity for crop insurance to provide economic support to farmers depends upon the crop to soothe their farm profits and persuade them to apply technology in agriculture, reduce indebtedness and decrease the need for relief measures. The farmer should have a better option to insurance his crop and transfer the hazard to the insurer. The insurer can estimate the weather risk factor and the payout to be paid to the farmer. Risk factors include Geographical basis risk, product basis risk, and product design basis risk. Product basis risk covers risk arising from deviations in parametric weather indices. The Risk could be high for rainfall, temperature and moderate for others factor like frost, heat, humidity, etc. Crop insurance provides a safety-net for farmers to mitigate losses arising from adverse climatic conditions and also encourages them to continue to invest in inputs and technology to record the loss assessment and increase the yields. Many practical data mining systems are used for predicting the insurance payout for the specified crop based on average temperature and rainfall that deal with building the classifier model of the system. This approach may promote agricultural growth and this is achieved by considering only the weather data that can be done by selecting the weather parameters and processed by the CART algorithm and recommended the crop insurance program in Coimbatore district [2]. Faster claims settlement can be done by the Government's financial details and the liabilities could be budgeted up-front and close-ended, as it supports the premium subsidy [16]. This work describes an interactive system for area-wise detection of the weather data and classifies the adverse weather parametric agriculture blocks which are eligible for index insurance and the insurance payout to be paid are predicted. This paper proposes a hybrid data mining algorithm CBKNN-PAYRULE rule based algorithm to predict the insurance payout value to be paid as per policymakers and recommend the eligible insurance areas that uses the Agro-meteorological data to make the crop insurance model very efficient.

It is done by an aggregation of districts as defined by State Government to diversify/ spread the risk and cover high/low/medium risk districts/area equally in such a way that each group/cluster contains a mix of districts with different risk profiles. Instead of using only rainfall based index an agro-meteorological index like average temperature and relative humidity that impacts the crop growth should be considered for weather-based crop insurance in India[17][20].

A. Background and Motivation

Insurance Claims processing should be made strictly as per the insurance term sheets, payout structure, and the scheme provisions. There are many aspects that influence the weather index insurance subject to the guidelines mentioned by the government .meanwhile this weather index insurance covers only weather triggers and automatically the payout calculation will be released by the government for the specified crop and Area with their respective Automatic weather station [16]. The additional factors to be considered are yield/loss assessment, risk profile, management practices by the farmers and the record of already eligible and received payout list from the government which is cross-checked in the field-work. This work considers the weather parameters with the respective crop/area and the risk factors like the high, medium, low-risk area of insurance declared by the government and predicts the payout to be paid by the bank/Insurance Agencies. Data Mining simplifies the findings and correlations in the given agriculture data.

This paper is organized into six sections. Section I presents the previous researched related to crop insurance payout in India and the data mining technologies. Section II describes the data collection and feature selection. Section III explains the proposed methodology and crop insurance payout estimation with data mining algorithms. Section IV presents the proposed work with the hybrid approach. Section V illustrates the experiments done and their results. Section VI shows the performance effectiveness of the proposed algorithms and Section VII concludes the research work.

RELATED WORKS

In the Survey paper the authors Aman Vohra, Nitin Pandey & S.K. Khatri provide a brief review of a variety of Data Mining techniques that have been applied to model data about the agricultural domain. The Data Mining techniques applied to Agricultural information include k-means, bi-clustering, and K nearest neighbor, Neural Networks (NN), Support Vector Machine (SVM), Naive Bayes Classifier and Fuzzy C-means. This survey summarizes the application of data mining techniques and the predictive modeling application in the agriculture field [1].

The author Ramzan and his co-authors have made the Analysis and investigation using data mining techniques by examining the changing patterns of weather parameters which include minimum temperature, maximum temperature, wind speed, and rainfall. After preprocessing of knowledge and outlier analysis, K-means clustering algorithm and Decision Tree algorithm were applied. Two clusters were generated by using K-means Clustering algorithm with the lowest and highest of mean parameters. The result attained with the smallest error (33%) was selected in the test data set. While for the number of rules generated of the given tree was selected with a minimum error of 25%. The results exposed that these techniques can be used for weather analysis and climate change studies [4].

Baojing Sun, ChanghaoGuo and G. Cornelis van Kooten has identified the potential weather variables that impact corn yields and to research the potency of weather-indexed insurance beneath varied thresholds for payouts. Statistical relationships between climate variables and crop yields area unit accustomed construct weather-indexed insurance that permits a farmer to hedge against adverse precipitation outcomes. Mean root square loss is employed to match the potency of varied weather product. It provides a method for calculating the premium for an insurance product that provides a payout if Cumulative Rainfall in a growing season is too low [6].

Abdallah Alashqur has made a set of classification rules which is constructed on top of naïve Bayesian classifier named Rule-based Naïve Bayesian Classifier (RNBC).Here the dataset is scanned only once, at the time of building the classification rule set. Subsequent scanning of the dataset , is avoided. The need to scan the dataset with each classification instance is avoided. In customary naïve Bayesian classification, for every time a new record is to be classified, the entire dataset needs to be scanned and a set of equations needs to be applied. Thus RNBC is considered an enhancement over existing naïve Bayesian classifiers [3].

Lebloisa and Philippe discussed the most advanced projects that have taken place in developing countries using these types of crop insurances are described. Subsequently, the methodology that has been used to design such projects in order to choose the meteorological index, the indemnity schedule and the insurance premium, is described. In specific, more research is needed on implementation, assessment of benefits, spatial variability of weather, deal with climate change and interactions with other hedging methods are discussed [14].

Xavier Gine et al [5] uses historical rainfall data to estimate the distribution of payouts on a rainfall index insurance product developed by the general insurer ICICI Lombard of southern India and discussed about geographically segmented insurance contracts. More research works is needed to improve the types of weather shocks against which rural household consumption is not well insured.

The author Rajesh Kumar discussed a model using a decision tree that has been proposed by the author to predict events like fog, rain, and thunder by inputting average temperature, humidity and pressure, which can be used by farmers in makings intelligent decisions [7].

This Research paper [8] focuses the clustering task for grouping the weather data in season-wise which can be used to analyze the reliability factor of Temperature, Humidity and Rainfall data by the given weather details from different regions in Tamil Nadu District based on correlation method. It provides explicit services to the assessment of pollution impacts from various industries and thermal power plants. The atmospheric correlations play a significant role in determining the climate trend. The author presents an effective method named NRS-CART, which is a hybrid method by combining neighborhood rough set (NRS) and classification and regression tree (CART).

Moreover, the innovative approach was used to detect and classify water-inrush possibilities [9].

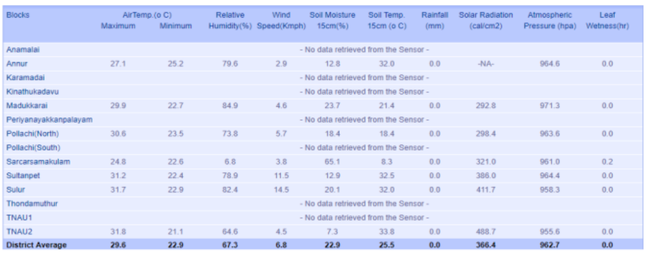
Johan Huysmans et al [10] proposed the advantage of the new algorithm Minerva and its ability to extract a set of rules from any type of black-box model. Experiments show that the mined models perform well in comparison with various other rule and decision tree learners.

II. DATA COLLECTION AND FEATURE SELECTION

In this Research work, the datasets are taken from the real-time weather datasets collected monthly from TamilNadu Agriculture Weather Network and www.accuweather.com and rainfall details From the Meteorological Department of Chennai for five regions such as Erode, Coimbatore, Salem, Namakkal ,and Karur. The data covered for the specified period in Tamilnadu districts and the case data is shown in **Figure: 1(a) & (b)**. The Features needed to Extract are Minimum and Maximum Temperature, Rainfall and Humidity are taken from Tamilnadu Agriculture Weather Network. In the network, 10 types of agricultural-related weather parameters from 385 AWS (automatic weather stations) are collected at the hourly interval and hosted on this website. Using this information the Agricultural officers will develop weather-based agro advisories at district and block level for the farmers. Even the farmers can estimate the payout based on their climatic conditions.

A. *Preprocessing*

The weather dataset used for this research work is collected and preprocessed by removing the noisy features and handled the missed and null data by KNN Algorithm[15].At every stage of the research, the standard procedures were established like Data Cleaning, Data Selection, Data Transformation, and Data Mining. And the machine learning algorithms are applied that are used to analyze the Agro-meteorological datasets.



Blocks	AirTemp.(o C) Maximum	Minimum	Relative Humidity(%)	Wind Speed(Kmph)	Soil Moisture 15cm(%)	Soil Temp. 15cm (o C)	Rainfall (mm)	Solar Radiation (cal/cm2)	Atmospheric Pressure (hpa)	Leaf Wetness(hr)
Anamalai	- No data retrieved from the Sensor -									
Annur	27.1	25.2	79.6	2.9	12.8	32.0	0.0	-NA-	964.6	0.0
Karamadai	- No data retrieved from the Sensor -									
Kinathukadavu	- No data retrieved from the Sensor -									
Madukkarai	29.9	22.7	84.9	4.6	23.7	21.4	0.0	292.8	971.3	0.0
Periyanayakkanpalayam	- No data retrieved from the Sensor -									
Pollachi(North)	30.6	23.5	73.8	5.7	18.4	18.4	0.0	298.4	963.6	0.0
Pollachi(South)	- No data retrieved from the Sensor -									
Sarcarsamakulam	24.8	22.6	6.8	3.8	65.1	8.3	0.0	321.0	961.0	0.2
Sultanpet	31.2	22.4	78.9	11.5	12.9	32.5	0.0	386.0	964.4	0.0
Sulur	31.7	22.9	82.4	14.5	20.1	32.0	0.0	411.7	958.3	0.0
Thondamuthur	- No data retrieved from the Sensor -									
TNAU1	- No data retrieved from the Sensor -									
TNAU2	31.8	21.1	64.6	4.5	7.3	33.8	0.0	488.7	955.6	0.0
District Average	**29.6**	**22.9**	**67.3**	**6.8**	**22.9**	**25.5**	**0.0**	**366.4**	**962.7**	**0.0**

(a) Sample weather data for Coimbatore district

Weather Data for Aravakurichi
Daily mean till 8:30 AM
District Karur Block Aravakurichi
Current Last Day Last 7 Days Last Month

Summary	AirTemp (o C) Maximum	Minimum	Relative Humidity(%)	Wind Speed(Kmph)	Soil Moisture 15cm(%)	Soil Temp. 15cm (o C)	Rainfall (mm)	Solar Radiation (cal/cm2)	Atmospheric Pressure (hpa)	Leaf Wetness(hr)
01-06-2019	37.1	26.8	0.4	6.0	17.2	26.3	0.0	615.5	986.2	0.0
02-06-2019	38.7	27.8	0.4	7.2	15.9	27.8	0.5	680.7	986.1	0.0
03-06-2019	35.9	26.5	0.4	8.5	15.0	26.7	0.0	450.2	986.2	0.0
04-06-2019	36.4	27.8	0.5	10.6	14.5	27.9	0.0	630.9	986.1	0.0
05-06-2019	38.0	25.4	0.4	6.4	14.0	28.7	1.0	627.0	985.8	0.0
06-06-2019	37.6	24.6	0.4	4.9	17.7	27.6	8.0	644.0	985.7	0.8
07-06-2019	36.3	27.3	0.4	4.9	18.0	25.9	0.0	647.9	985.7	0.0
08-06-2019	37.8	25.8	0.4	5.7	16.8	26.9	0.0	540.3	986.3	0.0
09-06-2019	37.9	27.6	0.4	5.1	16.2	28.4	0.0	682.1	985.8	0.0
10-06-2019	37.6	26.4	0.4	4.1	15.6	28.5	0.0	530.8	985.6	0.0
11-06-2019	34.5	26.8	2.6	5.8	15.0	27.7	0.0	518.1	986.1	0.0
12-06-2019	35.6	26.9	0.4	9.8	14.4	27.8	0.0	563.0	986.1	0.0
13-06-2019	34.7	28.3	0.4	15.0	13.9	27.5	0.0	571.5	986.3	0.0
14-06-2019	35.7	28.4	0.4	16.6	13.3	27.9	0.0	642.2	987.0	0.0

(b) Daily weather precipitation data for the block
Aravunkurichi of Karur District.

Figure:1(a) & (b) Retrieved from "http://tawn.tnau.ac.in/General/HomePublicUI.aspx", Tamilnadu Agriculture Weather Network by TNAU,2019 [21]

B. Feature Selection

The classification of data depends more upon the attributes, finding those relevant attributes is called attribute selection and the wrapper methods are used to measure the subclasses of attributes according to their utility to a given predictor. The method searches for an appropriate subset of attributes using the learning algorithm. Here the minimum temperature, maximum temperature with their respective month, block and district is selected by Rapid Miner [12].

III. PROPOSED METHODOLOGY

After Feature Selection, in the proposed methodology, the classification of the adverse temperature of the agriculture blocks of each district is done. For each Rise in Mean temperature (1^{o} Celsius) the mean temperature is classified as High/Very High temperature, Low/Very Low Temperature, Moderate/Low Moderate Temperature as shown in Table I. The mean Temperature values for three months of January, February, and March are taken from the given dataset, which has to be grouped according to the payout structure. The requirement of temperature by the crop plant may also vary fortnightly (i.e. 1st Jan, 2nd Jan, 1stFeb, 2nd Feb, 1st Mar, and 2nd Mar) or monthly or that the State Government decides. Consequently, the assessment of weather parameters should be evaluated for each fortnight/month according to historical data. The pay-out trigger(s) should be fixed strictly according to demonstrated correlation with the requirement of weather parameters to the crop at each critical stage [16]. The proposed work describes the classifier model for groundnut crop and the corresponding insurance payout to be paid to the farmer when their areas falls under the threshold values by applying the data mining algorithms with their experiments.

A. *Applying Data Mining Technique- Class-Based K-NN classifier (CB-KNN)*

In classification, many datasets have the problem in the difference of their class sizes, implication that one class will have too many instances while others have too few instances. Class-Based k-NN(CB-KNN) was developed by Voulgaris and Magoulas for unbalanced datasets and selecting very few instances when finding the classification based upon the K nearest neighbor[11]. The mean temperature values are chosen from the defined weather dataset and the value of K is automatically selected to maximize the degree of certainty of the classification, this selection is done by the classifier to decrease the influence of the most distance instance and categorize the temperature as Low, Very Low, Moderate, Low Moderate, Very High , and High as shown in

Table-I: Classification of Mean Temperature

Rise in mean temperature	Mean Temperature
1^{o} Celsius	Very Low
2^{o} Celsius	Low
3^{o} Celsius	Low Moderate

4^{o} Celsius	Moderate
5^{o} Celsius	High
6^{o} Celsius	Very High

Then the harmonic mean of the distances of Mean Temperature values is calculated as shown in equation (1).It is done by finding the inverse of the mean of the inverses of distances.

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \quad (1)$$

Finally, these harmonic means calculated are compared with the other sample sets and the class that gives the lowest value is chosen for the classification [13] [15].

B. *Learning Rule-based classification by Sequential Covering strategy*

Machine learning techniques that produce rules by covering algorithms works by adding the rule and test the data samples and trying to create a rule with maximum accuracy. The decision tree rule chooses associate attribute to maximizee the separation between the categories (using info gain criterion), the covering algorithm chooses an attribute-value pair to maximize the probability of the desired classification [18] [19]. There are two categories in this dataset: "YES" and "NO". From this data set, the system learns the set of rules that can be used to classify records. The rules are of the form:

R_x:(Condition1∧ Condition2∧ … ∧Condition n) → Class (2)

In (2), R_x is the rule-id and the left- hand side (LHS) of the rule represents conditions on the attributes in the dataset except for the class label. An example classification rule of figure 2 is

R_I : (Temperature=V.High) ∧ (payout>0.0) ∧ (Risk =High)→ (insurance=yes) (3)

In (3), R_I covers 12 records in the dataset which denotes that the rule R satisfies 12 records defined in the given dataset RIPPER is another algorithm that uses the same covering approach with the hill-climbing strategy. It applies an information gain criterion to frequently add conditions until the rule covers no negative examples.

C. *Crop Insurance payout prediction Model*

The Crop Insurance payout Prediction Model was done with the monthly weather Report of five districts is aggregated at a weekly or monthly frequency with the definition of Standard Meteorological Weeks (SMW). The evaluation of seasonal rainfall periods and temperature was aimed at constructing weather-based crop insurance to understand for farmers to capture the adverse events related to temperature is released by the government with the help of Insurance companies. Agricultural Insurance company of India (AIC) have released the payout term sheet for the farmers to implement the Heat or Rise in mean temperature with the corresponding payout table is shown in Figure 2.

Adverse Temp vs. Payout Table:

	STANDARD SOWING PERIOD					
Period (Fortnight)	Jan-1st FN	Jan-2nd FN	Feb-1st FN	Feb-2nd FN	Mar-1st FN	Mar-2nd FN
Fortnightly Trigger Temp. (°C) →	12.86	13.59	14.95	16.31	19.72	22.91
Rise in Fortnightly Mean temp (°C)	Payout (Percentage of Sum Insured)					
1.0	0.00	0.00	0.00	0.00	0.00	0.00
2.0	0.00	0.00	0.00	3.82	4.31	4.31
3.0	0.00	0.00	0.00	6.76	6.57	6.57
4.0	0.00	3.99	3.53	9.92	8.39	8.39
5.0	4.66	5.70	4.92	12.68	9.52	9.52
6.0	6.60	7.04	9.20	15.17	10.78	10.78

Source:Agricultural Insurance Company of India(AIC)

Figure 2.Heat or Rise in Mean Temperature, Operations by WBCIS, Agricultural Insurance Company of India(AIC), Retrieved from " http://www.aicofindia.com/ /AICEng/General_Documents/Product_Profiles/WBCIS_FAQ.pdf "[22]

Then weather data is preprocessed and then apply feature selection process The selected Dataset I attributes such as min and max temperatures, Threshold trigger temperature and rise in mean temperature were collected for different Fortnightly temperature trigger periods under the weather insurance cover area were fixed through standard statistical techniques and class-based KNN approach is made to get the optimized results. Dataset II attributes deals with Risk profile dataset which is released by the government to the insurance companies based on crop loss [17] as shown in Figure 3. High risk denotes the crop loss percentage to be higher and Medium Risk to be little lower and low risk to be lower and the payout structure varies accordingly.

Sl. No.	District Name	Loss Cost	Risk Level
1	District 1	7.80%	High Risk
2	District 2	8.60%	High Risk
3	District 3	5.40%	Medium Risk
4	District 4	3.20%	Low Risk
5	District 5	4.60%	Low Risk

Source: Department of Agriculture, Cooperation and Farmers Welfare, operational guidelines, PMFBY, India

Figure 3. Reprinted from "Risk coverage levels -Operational Guidelines - PMFBY", by Department of Agriculture ,Co-operation and Farmer, Ministry of Agriculture & Farmers Welfare, 2016 ,India, [17]

Both the datasets of processed weather data and risk profile data are considered for the classifier model. The model helps to classify the adverse temperature areas by CBKNN and then apply the sequential covering strategy to predict the insurance payout and decision rules to recommend the insurance. This model was built by constituting the hybrid algorithm of class-based KNN and rule-based algorithm for crop insurance payout prediction which is shown in Figure 4.

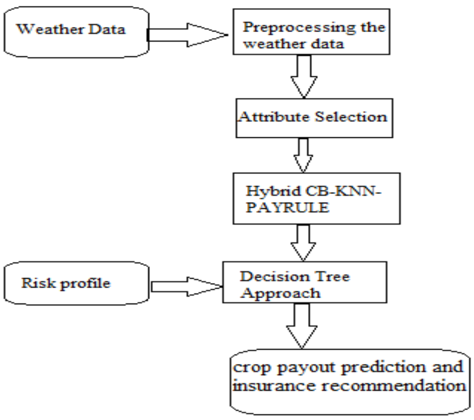


Figure 4: Prediction Model for Crop Insurance Payout

IV. PROPOSED WORK

Data mining helps to classify the adverse temperature blocks of selected district and analyze those temperatures with the specified Fortnightly Trigger Temperature and predict the payout. The proposed algorithm consists of two sections.
1) First section deals the CB-KNN classifier which is applied in the given dataset and categorize the temperature as shown in Table I.
2) Second section deals with building a rule-based classifier with sequential covering algorithm to generate the rules and payout values and measuring the accuracy of the classifier.
3) Third section deals with the decision tree approach that builds the decision tree with the given risk factors and recommends the eligibility of crop insurance areas by yes or no.

A. *Proposed CBKNN-PAYRULE Pseudocode*

Step 1) Load the Pre-Processed trained weather dataset
Step 2) Apply Feature selection on the dataset to select the appropriate attributes such as Minimum and Maximum temperature for the respective agriculture blocks for the specified month of each district.
Step 3) Compute the Mean temperature of each agriculture block of each district.
Step 3) Add the Fortnightly Trigger temperature point (threshold factor) defined for the data collected as shown in Figure 2.
Step 4) Collect and add the risk profile factors as defined in Figure 3 for the trained weather dataset.
Step 5) Apply Hybrid CBKNN-PAYRULE algorithm on the datasets that classifies the mean temperature for every Rise in mean temperature (in degree Celsius) as shown in Table I and calculates the payout values by applying rule-based classification algorithm and check the classification accuracy
Step 6) Calculate the payout prediction accuracy and its rank of the classifier, which measures the performance of the classifier in terms of model size and number of rules generated to classify the given samples.
Step 7) Apply decision tree approach with the risk file factors and classified temperature with their estimated payout values to recommend the insurance eligible areas.

B. *CBKNN-PAYRULE Algorithm*

Input: *T //Training data/tuples*
Temperature information like min, max temperatures for chosen blocks of Tamilnadu districts

Output*: the set of IF-THEN rules and the expected payout price*
//predicted crop insurance payout price of selected blocks of Tamilnadu district is done and rule set to describe the conditions for payout structure

// Get the minimum and maximum temperature values of first and second fortnight temperature of three months particularly January, February and March as per payout table
Collect the min, max temperature data's of the chosen agriculture blocks of five districts

// Calculated Mean Temperature is given as input as y
Compute the Mean Temperature as y
// Fortnightly Trigger temperature point is considered as threshold x ,Mean temperature as Attribute y.

Initialize training dataset D, Threshold x, Attribute y, category label set T

For each y belongs to D do
Calculate the distance D(y, x) between y and x using Harmonic Mean distance technique and categorize them
End for
Select the subset N from the dataset D, the N contains k training samples of the test sample x
Calculate the category of x;

$$T_X = \arg\max \sum_{y \in N} I(t=class(y)),\ t \in T$$

// category T_x classifies the temperature as High/V.high, Low/V.Low and Moderate/Low Moderate

Rule_set= { }; // initial set of rules learned is empty
Initialize model_size;
// Give the categorized temperature T_x as input to the Rule_One
For each class A do
Repeat
Rule_One = Learn-OneCrop-Rule (D, Att_vall, T_x, A);
Remove tuples covered by Rule form D;
Decrement model_size;
Until termination condition;
Rule_set=Rule_set+Rule; // add a new rule to rule-set
End for
Return Rule_Set, model size;
// returns the rules generated according to the dataset size and the payout value according to rule.

Generate the Rule Set for Groundnut Crop of chosen agricultural blocks of selected districts;
Predict their payout value;

Apply decision tree approach with CART with the risk profile factors and classified temperature to recommend the insurance area.

V. EXPERIMENTS AND RESULTS

In this section the performance of the proposed Hybrid approach is analyzed by the results generated. The results have the rules that are generated by the hybrid approach is shown in Table II.

(Month = 1-Jan) and (mean-temperature >= 26.5) and (rise<0)=> temperature=Normal (8.0/0.0)
(Month = 1-Jan) and (mean-temperature >= 26.5) and (rise=1) => temperature=Very Low (8.0/0.0)
(Month=2-Jan) and (mean-temperature>=27.5) and (rise=2) => temperature=Low (4.0/0.0)
(Month = 1-Feb) and (mean-temperature >= 29.5) and (rise=3) =>temperature=very moderate (28.0/0.0)
(Month=2-Feb) and (mean-temperature>=31.5) and (rise=4) => temperature=moderate (22.0/0.0)
(Month = 1-Mar) and (mean-temperature >= 34.5) and (rise=5) => temperature=High (8.0/0.0)
(Month=2-Mar) and (mean-temperature>=37.5) and (rise=6) => temperature=Very high (5.0/0.0)
IF (month=2-Feb) and (mean-temperature=Low) and (rise=2) THEN payout=3.82 (2.0/0.0)
IF (month= 1-mar) and (mean-temperature=Low) and (rise=2) THEN payout=4.31(2.0/0.0)
IF (month=1-mar) and (rise<=0)=>payout=0(15.0/0.0)
(Temperature=V.High) and (payout>0) and (Risk =High) THEN insurance=yes(12.0)
IF (temperature=Normal) and (payout=00 and (Risk=Low) THEN insurance=no (18.0)

Table II: Sample Rule Generation

With the derived results of the payout value, the risk profile factors are also included according to the areas and then the insurance is recommended by the decision tree approach by YES or NO as shown in Figure 5.

District	blocks	month	max.t	min-	mear	Trigge	M. R	temperatu	payout	Risk	insuran
Nammal	Rasipuran	1-Feb	37	27	32	26	6	V.High	9.2	High	yes
erode	Ammapet	2-Feb	36.7	29	33.1	31	2.1	Low	3.82	Low	yes
karur	kadavur	1-Jan	32	24	28	26	2	Normal	0	Low	no
karur	Aravakuri	2-Jan	39	23	31	27	4	Moderate	3.99	Low	yes
coimbat	annur	2-Mar	34	24	29	27	2	very Low	4.31	Low	yes
erode	Ammapet	2-Mar	38	28	32.8	37.5	-4.7	Normal	0	Low	no
salem	omalur	1-Mar	36	22	29	26	3	Moderate	6.57	Medi	yes
coimbat	annur	1-Jan	33	20	26.5	26	0.5	Normal	0	Low	no
coimbat	annur	1-Feb	32	18	25	29.5	-4.5	Normal	0	Low	no

Figure 5 Sample predicted payout results

A. *Correlation between Weather Attributes and District Payouts*

Weather Attributes and Payout Area for insurance claims scheme forms the selection factor for the government to learn far by displaying bivariate data in a graphical form that maintains the pairing fixed payout for districts, such pairwise display of variables is called a scatter plot.

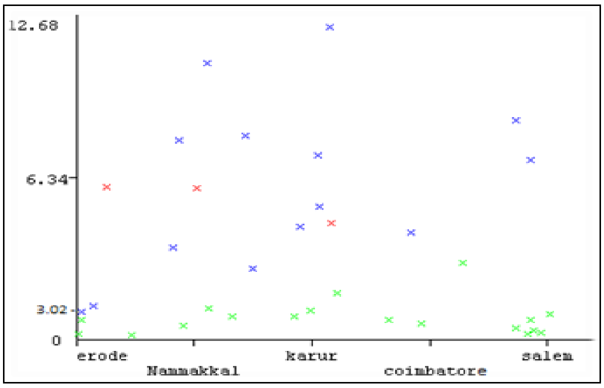


Figure 6 : District Vs. Payout shows the predicted payout values over five districts named Erode ,Coimbatore , Karur , Salem ,Namakkal.

In Figure 6, the estimated respective payout is seen for each selected district which shows the payout covered areas clearly.

VI. PERFORMANCE METRICS

The performance of the proposed CB-KNN-PAYRULE rule-based crop payout prediction model is analyzed and compared with the existing Rule–based classification model named JRipper, CART, C4.5 rules by using MATLAB 2018a.

Algorithm	Predictive Accuracy (%)	Accuracy Rank	Avg model size	Avg.No of Rules
JRip	88.32	8.3	2.78	1.90
C4.5Rules	82.14	8.1	4.50	2.54
CART	84.20	8.4	3.67	2.56
CBKNN-PAYRULE	89.75	8.7	2.67	1.67

Table III. Accuracy Measures of Algorithms.

These results are compared with the results reported in **Table III** indicates that the hybrid CBKNN–PAYRULE algorithm achieves better accuracy with a comparable number of rules generation techniques. The result shows that the accuracy measures with the model size and the average number of rules generated collectively for seasonal weather datasets for five districts.

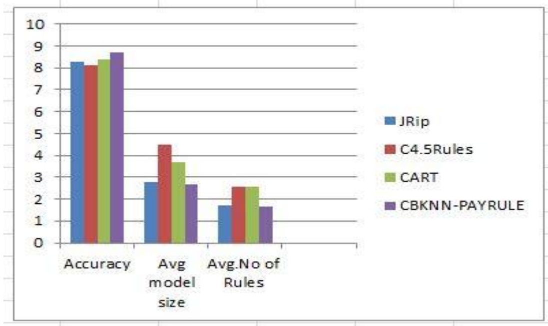


Figure 7: Chart illustrating the comparison of average accuracy rank , rules and model size

In Figure 7 , the comparison of accuracy rank, rules generated and model size of hybrid CBKNN-PAYRULE with the JRipper, CART, C4.5 algorithms is made. This work discovered that the samples with an average number of rules generated with the specified model size significantly differs with the model size when compared to JRip, CART, C4,5. This difference spots the characteristics of the developed hybrid algorithm. Other Rule Generation algorithm like JRip exposed the data samples with a smaller number of longer rules with increasing model size, but the Hybrid CBKNN-PAYRULE revealed data samples with a greater number of shorter rules.

Finally, the payout percentage covered is calculated simultaneously for each district from the predicted results and displayed diagrammatically in Figure 8. Namakkal covers 32% of maximum payout agricultural blocks covered under insurance, Salem of about 22% and so on.

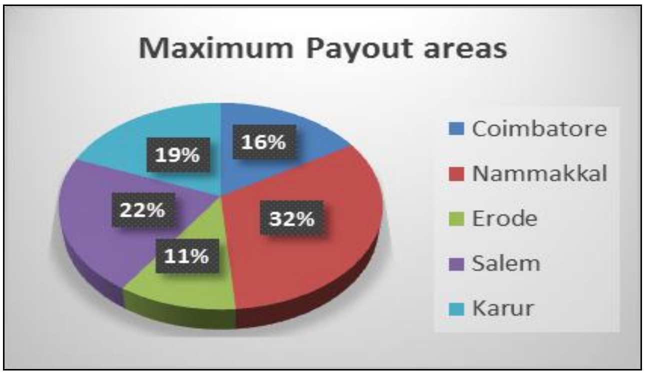


Figure 8: Chart describing the number of agricultural blocks covered under insurance payout of each district of Tamilnadu

Considering all the simultaneous predicted results for each agricultural blocks the work determines the percentage of payout covered areas for crop insurance in five districts Erode, Coimbatore, Karur, Salem, Namakkal. And from the chart it is known that Namakkal district is predicted to have more crop insurance payout claims.

VII. CONCLUSION

This paper introduces a new approach for building the class-based KNN with rule-based classifier to categorize the adverse temperature of agriculture blocks of five districts and estimate the area-wise insurance payout value based on payout term sheet by the classifier model. With the payout value, the decision tree approach helps to recommend the insurance eligible areas. The hybrid CBKNN-PAYRULE algorithm is well suited for this research to estimate the weather insurance payout by categorizing the temperature for five districts named Erode, Coimbatore, Karur, Salem, and Namakkal. And from the results, it is known that Namakkal district is predicted to have the maximum payout by the decision tree approach. The proposed technique supports the simultaneous prediction of the insurance payout to be paid in case of adverse weather factors of groundnut crop for five districts with high accuracy. Currently, this work focuses on the selected crop for the specified selected area. In future, this work can be extended to all districts of Tamilnadu and for all crops under insurance by including more meteorological attributes for decision making.

REFERENCES

1. A Brief survey of Data Mining Techniques Applied to Agricultural Data, Aman Vohra1, Nitin Pandey2, S.K. Khatri, International Journal of Computer Applications (0975 – 8887) Volume 95– No. 9, June 2014
2. K.P.Mangani, R.Kousalya, Weather Based Prediction Model for Recommending the Crop Insurance using CART Algorithm, International Journal of Computer Trends and Technology (IJCTT) - Volume 67 Issue 4 - April 2019.
3. Abdallah Alashqur, A novel methodology for constructing Rule-based naïve Bayesian classifiers, International Journal of Computer Science & Information Technology (IJCSIT) Vol 7, No 1, February 2015.
4. M Ramzan Talib, Tose ef Ullah, M UmerSarwar, M KashifHanif and NafeesAyub, S, Application of Data Mining Techniques in Weather Data Analysis, International Journal of Computer Science and Network Security, VOL.17 No.6, June 2017
5. Xavier Gine, Robert Townsend, and James Vickery, Statistical analysis of rainfall insurance Payouts in southern India, Amer. J. Agr. Econ. 89 (Number 5, 2007): 1248–1254,Copyright 2007
6. Baojing Sun, ChanghaoGuo and G. Cornelis van Kooten,,Hedging weather risk for corn production in Northeastern China The efficiency of weather-indexed insurance, Agricultural Finance Review, Vol. 74 No. 4, 2014,pp. 555-572, Emerald Group Publishing Limited,0002-1466
7. Rajesh Kumar, "Decision Tree for the Weather Forecasting",International Journal of Computer Applications (0975 – 8887) Volume 76– No.2, August 2013
8. M.Mayilvaganan, P.Vanitha, Correlation Analysis of Meteorological Data in Region of Tamil Nadu Districts Based On K- Means Clustering Algorithm, International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 3, May-June 2015
9. Li Fenglian, Zhang Xueying DUChunlei, Huang Lixia, A Hybrid NRS-CART Algorithm and Its Application on Coal Mine Floor Water-inrush Prediction, 2015,1EEE.
10. Johan Huysmans, Rudy Setiono, Bart Baesens, and Jan Vanthienen, Minerva: Sequential Covering for Rule Extraction,IEEE ,April 2008.
11. George D. Magoulas, Extensions of the k-Nearest Neighbour methods for Classification Problems, 2015,IEEE
12. H Malik, S Mishra,"Feature selection using Rapid Miner and classification through probalistic neural network for fault diagnostics of power transformer,2014,IEEE
13. Swe Aung1 , ItaruNagayama2 , and Shiro Tamaki3,"Dual-kNN for a Pattern Classification Approach", IEIE Transactions on Smart Processing and Computing, vol. 6, no. 5, October 2017
14. A. Lebloisa and Philippe Quirion b, Agricultural insurances based on meteorological indices: realizations, methods and research challenges, meteorological applications,,Meteorol. Appl. 20: 1–9 (2013), Royal Meteorological Society
15. Harshit Dubey,Vikram Pudi ,Class Based Weighted K-Nearest Neighbor over Imbalance Dataset, Springer-Verlag Berlin Heidelberg 2013
16. World Bank Group (2012), "Weather based crop insurance in India", Policy Research Working Paper No. 5985, World Bank Group, Washington, DC.
17. Department of Agriculture, Cooperation and Farmers Welfare, Ministry of Agriculture & Farmers Welfare ,"Operational guideline restructured weather based crop insurance scheme" March 2016,KrishiBhawan, New Delhi-110001
18. Omer Akgobek, A rule induction algorithm for knowledge discovery and classification, Turkish Journal of Electrical Engineering and Computer Sciences, January 2013.
19. Available:http://www.cs.bc.edu/~alvarez/ML/covering.html
20. Available:http://www.afcindia.org.in/PDF/research_reports/WBCIS-FINAL%20REPORT-060211-PDF%20TO%20AFC/WBCIS-FINAL%20REPORT-060211.pdf
21. http://tawn.tnau.ac.in/General/HomePublicUI.aspx
22. http://www.aicofindia.com/AICEng/General_Documents/Product_Profiles/WBCIS_FAQ.pdf

AUTHORS PROFILE

K.P.Mangani received her B.Sc. computer science, in 2000, from Madurai Kamaraj University, Madurai and MCA in 2003, from Madurai Kamaraj University, Madurai. She attained her M.Phil in the area of Data Mining from Periyar University, Salem in 2009. At present, she is pursuing her Ph.D in the area of Data Mining in Dr.NGP Arts and Science College under Bharathiyar University,Coimbatore. She is a life member of the professional body Indian Society of Systems for Science and Engineering (ISSE).

Dr._R. Kousalya received her B.Sc., from P.S.G.R. Krishnammal College for women, Coimbatore in 1997 and acknowledged M.C.A from Bharathiyar University, Coimbatore in 2000. She attained her M.Phil. in the area of Data Mining from Manonmaniam Sundaranar University, Tirunelveli in 2003. She achieved her Ph.D. degree in the area of data mining from Manonmaniam Sundaranar University, Tirunelveli in 2016. At present she is working as a Professor & Head at the Department of Computer Applications in Dr.N.G.P Arts and Science College, Coimbatore.

Her research interest lies in the area of Data Mining and Web Mining. She has published more than 20 research papers in various National and International journals in the area of Data Mining.. She is a member of the professional body Computer Society of India(CSI).