

The Association Rule of Corn Disease Symptoms by Using Frequent Pattern Growth and Random Forest

Achmad Yasid¹, Budi Dwi Satoto²

Faculty of Engineering
University of Trunojoyo Madura
Bangkalan, Indonesia

¹ayasid@trunojoyo.ac.id, ²budids@gmail.com

Abstract—Despite proper soil health management, pests and diseases control is also an important task in corn farming for improving both the quality and the quantity of the crops production. One way to address this challenge is to identify the emerging symptoms accurately to help define the appropriate solution. Downy, Leaf Blight, Rot midrib, Rot Stem, Leaf Rust, Burnt, Dwarf, Leaf spot are among the common symptoms on corn crops. Failure to identify these symptoms in the early stage could result adverse effects on the corn crops and reduce the production in the long run. This study therefore aims to classify corn disease symptoms to know the real disease. In this study, Decision trees and Frequent Pattern Growth (FP Growth) are employed. Analysis from the data indicated accuracy values of decision tree J4.5 (ID3) of 95%, J.48 of 96% and Random Forest of 95%. While for the relation pattern with 9 attributes and 2 classes, obtained the most often arise rule is the mycelium symptom or cotton with the leaves color turning red / brown / ash / chlorotic with accuracy of 92%.

Keywords—frequent pattern growth; decision tree; random forest; corn disease symptoms

I. INTRODUCTION

Corn is one of the leading agricultural products in Madura in addition to rice, tobacco and nuts. Local corn varieties are still the main choice for farmers where 90 percent of farmers in Madura still use local corn varieties. The problem is the cultivation still lack optimal. In conjunction with increasing food crop production, corn has important role. But in its management there are obstacles including pests and diseases. The diseases that often attack corn include Downy, Leaf Spot, Rot midrib, Rot Rod, Leaf Rust, Burnt, Dwarf, Leaf Spots [1]. If the diseases are not controlled, the corn production reduced. The steps to control it are by selecting seeds or varieties and pests right handling. One of the ways to find out the corn attack pattern is to look for patterns of disease relations by using association rules [2]. By knowing the interrelationship between diseases can be known the handling priority so the expected operational costs not too high. Association Rule is a procedure in Market Basket Analysis to find knowledge in the relationships form between items in a dataset and display them in the patterns form that explains the relationship between variables or attributes [3]. The Association Rule Method has several methods that can be used including the Apriori method [4], decision tree [5], Frequent Pattern Tree [6], Frequent

Pattern Growth [7], Weighted Tree and many more [8]. This study used Decision Tree J.48 which is the development of Decision Tree J.45 algorithm.

II. LITERATURES

A. Corn disease

Corn disease is diseases that attack the corn plants, can be explained, among others are:

1) Downy mildew (*Peronosclerospora maydis*)

The disease symptoms occur on the leaves surface of white to yellowish followed by chlorotic lines and another characteristic is that in the morning at the leaf bottom there is white velvet layer. Downy mildew on corn plants causes extended systemic symptoms throughout the plant and cause local symptoms. Plants infected with downy mildew at young age usually do not form fruit, but if the infection in older plants they still form fruit but generally with stunted growth [9].

Control can be done by using downy mildew resistant varieties, such as Srikandi, Lamuru, and Gumarang. In addition, simultaneous planting can be carried out and the corn free period at least two weeks to one month in each year. If infected it can be eradicated or totally destroyed. For prevention can use metallaixill fungicide on plant seeds with a dose of 0.7 grams of active ingredient in each kg of seed.

2) Leaf spots (*Bipolaris maydis Syn*)

This disease symptoms occur when yellowish or reddish-brown leaves appear. When the corn seed is exposed to this spot can wither or die within 3-4 weeks. If cobs infected will cause the seeds to break and rot, even the cob can fall. This disease infection can be carried by the wind or raindrops and can cause the first infection in corn plants. Control can use leaf resistant varieties, such as Bima 1, Sukmaraga, and Palakka. If you see plants that have been infected, they must be eradicated immediately. Fungicides can also be carried out with the active ingredients of mancozeb and carbendazim [10].

3) Leaf Blight (*Rhizoctonia solani*)

This disease symptoms occur when oval-shaped spots appear and the spots are elongated and develop into necrotic and called blight. Spots are grayish green or brown and appear

early on the lower leaves then develop towards the upper leaves. Severe infections can cause plants to die quickly or dry out and the fungus does not infect the cob or klobot.

Control can use leaf blight resistant varieties, such as Bisma, Pioner 2 and 14, and Semar 2 and 5. If visible plants that have been infected must be eradicated immediately. It can also be done using *Trichoderma viride* antagonist fungi and giving fungicides with active ingredients of mankozeb and dithiocarbamate [11].

4) Leaf rust (*Puccinia polysora*)

Symptoms of this disease occur when small patches appear round to oval on the surface of corn leaves at the top and bottom. These patches produce uredospores that are round or oval and play an important role as a source of inoculum in infecting other corn plants and their distribution through the wind. Rust disease can occur in the lowlands to high and the infection develops well in the rainy season or dry season.

Control can be done by using leaf rust resistant varieties such as Lamuru, Sukmaraga, and Semar 10. If visible plants that have been infected must be eradicated immediately. Can also be done by giving fungicides with benomil active ingredient [12].

5) Stem Rot (*Fusarium sp.*)

This disease symptoms generally occur after the flowering phase. The base of infected stem changes color from green to brown, the inside decomposes, so it is easy to fall, the outer skin is thin. At the base of the infected stem shows pink, brownish brown or brown color. This disease can be spread by wind, rain water, and insects.

Control can be done by using stem rot resistant varieties, such as BISI-1, Surya, CPI-2, and Pioneer-8. In addition it can be done crop rotation, balanced fertilization, avoid giving high N and low K, and good drainage to prevent attacks. Biological control can also be carried out with *Trichoderma sp* antagonists [13].

B. Data Mining

Data mining is process that uses statistical techniques, mathematics, artificial intelligence, and machine learning to extract and identify useful information and related knowledge from various databases. Data mining is process that employs one or more computer learning techniques to analyze and extract knowledge or knowledge automatically [3].

1) Decision tree

One method in Data Mining is Decision Tree, is one of the classification methods on Machine Learning. Decision Tree is a classification method that uses a tree structure, where each node represents attribute, and the branch represents value of attribute, while the leaf is used to represent the class. Classification is the process of finding a collection of patterns or functions that describe and separate data classes from one another, so that they can be used to predict data that does not have particular data class [14].

Association rule is pattern of attribute relationships that influence each other. The need to know this pattern is to

determine the attributes priority in decision tree. The attribute table for disease symptoms is shown in Table 1.

TABLE I. CORN DISEASE SYMPTOMS

TID	Attribute of corn diseases symptoms
Land 1	<i>a,c,e,f,m,p</i>
Land 2	<i>a,c,d,e,f,m,p</i>
Land 3	<i>a,b,c,e,f,g,m</i>
Land 4	<i>b,c,f,i</i>
Land 5	<i>b,c,e,n,p</i>
Land 6	<i>k,l</i>

Attribute control data can be described in binary format, where each row of land corresponds to symptom item and each column corresponds to attribute control. An item can be treated as binary variable worth 1 (y) if the item is in transaction and 0 if the opposite.

Itemset $I = \{i_1, i_2, \dots, i_d\}$ is the set of all items in market basket data and $T = \{t_1, t_2, \dots, t_N\}$ is the set of all transactions. Each t_i transaction contains a subitemset of I . In the association analysis the collection of one or more items is called itemset. If the itemset contains k items, it is called k -itemset. Null sets are itemset that do not contain items. Transaction width is defined as the number of items in the transaction. Example itemset: $\{i_1, i_2, i_3\}$ [15].

Attribute control data can be described in binary format, where each row of land corresponds to a symptom item and each column corresponds to an attribute control. An item can be treated as a binary variable worth 1 (y) if the item is in a transaction and 0 if the opposite.

Itemset $I = \{i_1, i_2, \dots, i_d\}$ is the set of all items in market basketball data and $T = \{t_1, t_2, \dots, t_N\}$ is the set of all transactions. Each t_i transaction contains a subitemset of I . In the association analysis the collection of one or more items is called itemset. If the itemset contains k items, it is called k -itemset. Null sets are itemset that do not contain items. Transaction width is defined as the number of items in the transaction. Example itemset: $\{i_1, i_2, i_3\}$ [15].

The support count (σ) is the frequency of the appearance of certain itemsets in the transaction. Mathematically, it can be stated by:

$$\sigma(X) = |\{t_i \mid X \subseteq t_i, t_i \in T\}| \quad (1)$$

with:

X = itemset

t_i = transaction i^{th}

T = transaction set

$|\cdot|$ = number of elements in a set

Example: $\sigma(\{i_1, i_2, i_3\}) = 2$

Association Rule is implication expression of $X \rightarrow Y$ form, where X and Y are disjoint itemsets, $X \cap Y = \emptyset$. The

association rule strength can be measured by the value of support and confidence. Support determines how often the rule appears in the data, while confidence determines how often items Y appear with X in the same transaction. Mathematically can be stated by:

$$\text{Support, } s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (2)$$

$$\text{Confidence, } c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (3)$$

with notation :

$X, Y = \text{itemset}$

$\sigma = \text{support count}$

$N = \text{total transaction amount}$

Example of *support* value:

$$s(i_1, i_3 \rightarrow i_4) = \frac{\sigma(i_1, i_3, i_4)}{N} = \frac{2}{5} = 0.4$$

Example of *confidence* value:

$$c(i_1, i_3 \rightarrow i_4) = \frac{\sigma(i_1, i_3, i_4)}{\sigma(i_1, i_3)} = \frac{2}{3} = 0.67$$

Association rules exploration can be expressed as if there is set of T transactions, how to find all the rules that have support \geq minsup and confidence \geq minconf, where minsup and minconf are related to the limits of support and confidence values [16].

2) Decision Tree J.48

The algorithm J.48 is the development of conventional decision tree induction algorithm which is very important, namely ID3. The algorithm which is the development of ID3 can classify data with decision tree methods that have advantage able to process numerical and discrete data, can handle missing attribute values, produce rules that are easily interpreted, and the fastest among algorithms that use main memory on the computer [17]. In the application of several cases of classification techniques, this algorithm able to produce better performance.

3) Frequent Pattern Growth

FP-growth algorithm is very efficient algorithm in finding frequent itemset in data set by generating prefix-tree structure or called FP-tree. Frequent Pattern Growth (FP-growth) is alternative algorithm that can be used to determine the most frequently occurring data sets (data itemset) in data set. FP-growth algorithm uses data structure called FP-tree (Frequent Pattern Tree) in conducting frequent itemset searches [18].

As the name implies, FP-tree is tree structure where each branch contains frequent itemset information and each node stores item information and frequency. The parent and child sequence of this tree is determined by the frequency of each item. Items with greater frequency will be the parent for items with smaller frequency. Likewise the opposite item with smaller frequency will be child for larger item. Thus, the leaf

from this tree is the item with the smallest frequency. This algorithm does not do candidate generation in the frequent itemset search process so that it can reduce database scans repeatedly in the mining process and can run faster [19].

FP-tree has header table that contains item information and frequency and has pointer to the node in the tree. The items list in this table is arranged based on the frequency, starting with items with the largest frequency then followed by items with smaller frequencies and so on. When compared with the Apriori algorithm and the variants or techniques of candidate generation, which requires several checks on the database, FP-growth only requires two checks on the database. The first check calculates the frequency of each item and then puts it in the table. Then proceed with second check that will enter the frequency information of each item into the FP-tree [20].

After the header table and FP-tree are created, frequent itemset search is only carried out through this structure, so that until this stage the examination of the database is no longer needed. The next level of frequent itemset is obtained by creating a FP-tree conditional with the FP-growth algorithm. This algorithm works recursively in making FP-tree conditionals. In the FP-growth algorithm, FP-tree tracing is mandatory step to obtain FP-tree conditionals. To make FP-tree conditionals are needed twice FP-tree search. As result, most of the time used by this algorithm is spent doing FP-tree search.

Therefore to reduce search time can be achieved with modification by using an additional data structure in the array form so the search process in FP-tree can be reduced [21]. FP-tree search is only done once because the information needed to make FP-tree conditionals is available in array. This FP-growth algorithm modification produces a new algorithm called FP-growth algorithm.

4) Random forest

Random forest (RF) is algorithm used in the classification of large amounts of data. Random forest classification is done through the trees merging by conducting training on the sample data owned. The trees use that increasingly affect the accuracy that will be obtained is better. Determination of classification by random forest is taken based on the results of the voting of tree formed. Winners of the tree formed are determined by the most votes. Tree construction on random forest until it reaches maximum size of the data tree. However, the random forest tree construction is not pruned which is method to reduce space complexity. Development is carried out by applying random feature selection method to minimize errors. Tree formation with sample data using variables taken randomly and run classification on all the trees that are formed [22].

C. Level of accuracy, sensitivity and specificity

The accuracy calculation as shown in the formula 4,5,6,7 [23].

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (4)$$

$$\text{Sensitivity / Recall} = TP / (TP + FN) \quad (5)$$

$$\text{Specificity} = TN / (TN + FP) \quad (6)$$

$$\text{Precision Rate} = TP / (TP + FP) \quad (7)$$

With description of TP (true positive) = record number of corn disease symptom classified as the corn diseases cause, TN (true negative) = records number of non bacteria corn disease classified as not corn disease, FN (false negative) = records number of corn disease classified as non-corn disease and FP (false positive) = records number of not corn disease classified as corn disease [24].

III. METODOLOGY

The methodology in this study can be described as shown in figure 1.

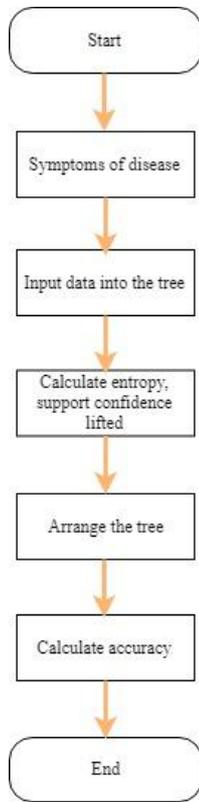


Fig. 1. Methodology

In this study, the experimental data processing is done by using 3 algorithms, namely Decision tree C4.5, Decision Tree J.48, Random Forest and FP Growth.

IV. DISCUSSION

This study uses 100 records consisting of 15 attributes and 6 classes that are used to form decision tree. The attributes used include symptoms of small / round spots on leaves of leaf color (red / brown / ash) chlorotic dry / wilted / dwarf, thin bark rotten midrib (red, ash) mycelium (cotton) klobot infection sticking, rotten brown cob of corn seeds, Sclerotium fungus (spore) with 5 classes is action that needs to be done in connection with the disease, namely fungicide-eradication-variety, variety-drainagee-fungicide-rotation, fungicide, fertilization-rotation, fungicide-eradication-rotation, and variety-annihilation -fungicide.

TABLE II. TABLE SYMPTOMS OF CORN PLANT DISEASE

Name of Corn Disease	Action taken			
	Selection of varieties	Eradication	Fungicide	-
Leaf blight	Selection of varieties	Eradication	Fungicide	-
Stem rot	Selection of varieties	Idrainase	Fungicide	Crop rotation
Downy	Selection of varieties	Eradication	Fungicide	Crop rotation
Rotten Cob_fusarium	Fungicide	-	-	-
Rotten cob_diplodia	Fungicide	-	-	-
Rotten cob_gibberella	Fertilization	Crop rotation	-	-
Stem rot	Fungicide	-	-	-
Leaf rust	Selection of varieties	Destruction	Fungicide	-
Mozaic virus	Selection of varieties	Eradication	Fungicide	Crop rotation

From the fifteen attributes using the decision tree algorithm J4.5 or ID3, Correctly Classified Instances 96 (96%), Incorrectly Classified Instances 4 (4%), Kappa statistics 0.9492, Mean absolute error 0.0247, Root mean squared error 0.1112, Relative absolute error 9.3885%, Root relative squared error 30.6874% of the Total Number of Instances 100. While the decision tree results are formed using Classifier J.48, the tree structure and accuracy values are obtained. The second result uses Decision Tree J48, the following tree structure is obtained

```

Round spot <= 0
| small round spot on leaf <= 0
| | mycelium (cotton) <= 6
| | | presence of pest<= 0
| | | | klobot infection(sticking) <= 0: fungicide--(22.0)
| | | | infeksi_klobot(sticking) > 0
| | | | | leaf color (red/brown/grey/chlorotik) <= 4: rotation fertilization--(8.0)
| | | | | leaf color (red/brown/grey/chlorotik) > 4: fungicide---(11.0/1.0)
| | | presence of pest> 0: variety-eradication -fungicide-rotation(23.0/2.0)
| | mycelium (cotton) > 6: variety-drainage-fungicide-rotation(15.0/1.0)
| small round sport on leaf > 0: variety-eradication-fungicide-(11.0)
Round spot> 0: variety-eradication-fungicide-(10.0)
    
```

with Confusion Matrix for each class that shows the best treatment to be done is shown as follows

```

a b c d e f <-- classified as
11 0 1 0 0 0 | a = variety-eradication-fungicide-
    
```

0 14 0 0 0 0 | b = variety-drainage-fungicide-rotation
 0 2 31 0 1 0 | c = fungicide---
 0 0 0 8 1 0 | d = fertilization-rotation--
 0 0 0 0 21 0 | e = variety-eradication-fungicide-rotation
 0 0 0 0 0 10 | f = variety-eradication-fungicide-

Accuracy is calculated by calculating accuracy, sensitivity and specificity

TABLE III. ACCURACY CALCULATION

TP rate	FP rate	Precision	Recall	F-Measure	ROC Area	Class
0.6368	-	1.0000	0.6368	0.6646	0.6868	varieties-eradication-fungicide
1.0000	0.012	0.6479	1.0000	0.6708	0.6903	varieties-drainage-fungicide-rotation
0.6535	0.015	0.0674	0.6535	0.6632	0.6813	fungicide
0.6174	-	1.0000	0.6174	0.6535	0.6896	fertilization-rotation
1.0000	0.025	0.6340	1.0000	0.6632	0.0688	varieties-eradication-fungicide-rotation
1.0000	-	1.0000	1.0000	1.0000	1.0000	varieties-removal-fungicide

Furthermore, by using the Decision Tree J.48, the result of Correctly Classified Instances 95 (95%) Incorrectly Classified Instances 5 (5%) Kappa statistics 0.9367, Mean absolute error 0.0296, Root mean squared error 0.1327, Relative absolute error 11.2045%, Root relative squared error 36.578%, Total Number of Instances 100.

And trees formed by using J.48 were then processed to obtain item relations from corn disease symptoms by using 9 attributes including 'small / round spots on the leaves,' leaves (red / brown / ash / chlorotic), 'dry / wilted / dwarf, skin thin stems, 'midribs' (red / gray), 'mycelium (cotton),' klobot infection 'sticks,' stinking infections 'brown corn seeds,' sclerotium fungi (spores) and 2 classes namely fungicide-drainage-rotation-eradication varieties - fertilization and fungicides. Using 50 training data obtained Correctly Classified Instances 46 (92%), Incorrectly Classified Instances 4 (8%), Kappa statistics 0.8983, Mean absolute error 0.1648, Root mean squared error 0.2314, Relative absolute error 62.4124%, Root relative squared error 63.8584 %, Total Number of Instances 100, the result is a rule as shown in the following explanation.

FP Growth found 13 rules (displaying top 10)

- [mycelium (cotton) = y]: 12 ==> [Leaf color (red / brown / ash / chlorotic) = y]: 12 <conf: (1)> elevator: (1.61) lev: (0.09) conv: (4.56)
- [mycelium (cotton) = y]: 12 ==> [mushroom sclerotium (spore) = y]: 12 <conf: (1)> lift: (2) lev: (0.12) conv: (6)
- [foul cobs infection = y, mycelium (cotton) = y]: 9 ==> [leaf color (red / brown / ash / chlorotic) = y]: 9 <conf: (1)> elevator: (1.61) lev : (0.07) conv: (3.42)
- [mycelium (cotton) = y]: 12 ==> [leaf color(red / brown / ash / chlorotic) = y, sclerotium (spore) = y]: 12 <conf: (1)> elevator: (2.5) lev: (0.14) conv: (7.2)
- [color_ leaves (red / brown / ash / chlorotic) = y, mycelium (cotton) = y]: 12 ==> [sclerotium (spore) = y]: 12 <conf: (1)> lift: (2) lev: (0.12) conv: (6)
- [fungus sclerotium (spore) = y, mycelium (cotton) = y]: 12 ==> [Warna_daun (red / brown / ash / chlorotic) = y]: 12 <conf: (1)> elevator: (1.61) lev: (0.09) conv: (4.56)
- [foul cobs infection = y, mycelium (cotton) = y]: 9 ==> [sclerotium fungus (spore) = y]: 9 <conf: (1)> elevator: (2) lev: (0.09) conv : (4.5)
- [rotten cobs infection = y, mycelium (cotton) = y]: 9 ==> [leaf color (red / brown / ash / chlorotic) = y, sclerotium (spore) = y]: 9 <conf: (1)> elevator: (2.5) lev: (0.11) conv: (5.4)
- [color_daun (red / brown / ash / chlorotic) = y, foul cobs infection = y, mycelium (cotton) = y]: 9 ==> [sclerotium fungus (spore) = y]: 9 <conf: (1)> elevator: (2) lev: (0.09) conv: (4.5)
- [foul cobs infection = y, sclerotium fungus (spore) = y, mycelium (cotton) = y]: 9 ==> [leaf color (red / brown / ash / chlorotic) = y]: 9 <conf: (1)> elevator: (1.61) lev: (0.07) conv: (3.42)

V. CONCLUSION

From the analysis and testing results, it can be concluded that by using algorithm of the decision tree J4.5 or ID3 obtained 96 Correctly Classified Instances, accuracy results of Decision Tree J.48 shows 95% Correctly Classified Instances, and the Random Forest Classified Instances 95% with the best rule result of FPGrowth on [mycelium (cotton) = y]: 12 ==> [Leaf color(red / brown / ash / chlorotic) = y]: 12 <conf: (1)> elevator: (1.61) lev: (0.09) conv: (4.56)

REFERENCES

- [1] A. Soliman, et al., "Developed and validated inoculation and disease assessment methods for Goss's bacterial wilt and leaf blight disease of corn", Crop Protection, vol. 112, p. 159-167, 2018.
- [2] R. Kost, B. Littenberg, and E. S. Chen, "Exploring generalized association rule mining for disease co-occurrences", in AMIA Annual Symposium Proceedings, American Medical Informatics Association, 2012.
- [3] S. Ramasamy and K. Nirmala, "Disease prediction in data mining using association rule mining and keyword based clustering algorithms", International Journal of Computers and Applications, p. 1-8, 2007.
- [4] C. Jia, R. Lu, and L. Chen, "A Frequent Pattern Mining Method for Finding Planted Motifs of Unknown Length in DNA Sequences", International Journal of Computational Intelligence Systems, vol. 4, no. 5, pp. 1032-1041, 2011.
- [5] S. Sivakumari, R. P. Priyadarsini, and P. Amudha, "Accuracy evaluation of C4. 5 and Naive Bayes classifiers using attribute ranking method",

- International journal of computational intelligence systems, vol. 2, no. 1, pp. 60-68, 2009
- [6] P. Gupta and B. B. Sagar, "Determining Weighted, Utility-Based Time Variant Association Rules Using Frequent Pattern Tree", *Journal of Engineering and Education*, vol. 14, 2018
- [7] J. Heng, et al., "Research and application of a combined model based on frequent pattern growth algorithm and multi-objective optimization for solar radiation forecasting", *Applied Energy*, vol. 208, pp. 845-866, 2017.
- [8] F. Séverac, et al., "Non-redundant association rules between diseases and medications: an automated method for knowledge base construction", *BMC medical informatics and decision making*, vol. 15, no. 1, p. 29, 2015.
- [9] S. Tanaka, et al., "A secreted *Ustilago maydis* effector promotes virulence by targeting anthocyanin biosynthesis in maize", *Elife*, vol. 3, p. e01355, 2014.
- [10] L. Gao, et al., "Use of a novel two-stage cultivation method to determine the effects of environmental factors on the growth and sporulation of several biocontrol fungi", *Mycoscience*, vol. 50, no. 4, pp. 317-321, 2007.
- [11] H. Jacobs, et al., "Translocation of carbon by *Rhizoctonia solani* in nutritionally-heterogeneous microcosms", *Mycological Research*, vol. 108, no. 4, pp. 453-462, 2004.
- [12] N.Y.Z. Ramirez-Cabral, L. Kumar, and F. Shabani, "Global risk levels for corn rusts (*Puccinia sorghi* and *Puccinia polysora*) under climate change projections", *Journal of Phytopathology*, vol. 165, no. 9, pp. 563-574, 2017.
- [13] I. M. Singha, et al., "Identification and characterization of *Fusarium* sp. using ITS and RAPD causing Fusarium wilt of tomato isolated from Assam, North East India", *Journal of Genetic Engineering and Biotechnology*, vol. 14, no. 1, pp. 99-105, 2016.
- [14] W.N. Khotimah, et al., "Tuna fish classification using decision tree algorithm and image processing method", in *Computer, Control, Informatics and its Applications (IC3INA), International Conference on 2015, IEEE*, 2015.
- [15] B. Vo, F. Coenen, and B. Le, "A new method for mining Frequent Weighted Itemsets based on WIT-trees", *Expert Systems with Applications*, vol. 40, no. 4, pp. 1256-1264, 2013.
- [16] D. J. Prajapati, S. Garg, and N. Chauhan, "Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment", *Future Computing and Informatics Journal*, vol. 2, no. 1, pp. 19-30, 2017.
- [17] R. Panigrahi and S. Borah, "Rank Allocation to J48 Group of Decision Tree Classifiers using Binary and Multiclass Intrusion Detection Datasets", *Procedia Computer Science*, vol. 132, pp. 323-332, 2018.
- [18] S. Thakare, S. Rathi, and R. Sedamkar, "An Improved PrePost Algorithm for Frequent Pattern Mining with Hadoop on Cloud", *Procedia computer science*, vol. 79, pp. 207-214, 2016.
- [19] B. Zeng, et al., "The improvement of weighted association rules arithmetic based on FP-tree", in *Advanced Computer Theory and Engineering (ICACTE), 3rd International Conference on. 2010. IEEE*, 2010.
- [20] C. Lucchese, S. Orlando, and R. Perego, "Fast and memory efficient mining of frequent closed itemsets", *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 21-36, 2006.
- [21] K. W. Chon, S. H. Hwang, and M. S. Kim, G. Miner, "A fast GPU-based frequent itemset mining method for large-scale data", *Information Sciences*, vol. 439, pp. 19-38, 2018.
- [22] M. B. Kursa, "Robustness of Random Forest-based gene selection methods", *BMC bioinformatics*, vol. 15, no. 1, p. 8, 2014.
- [23] J. N. Mandrekar, "Simple statistical measures for diagnostic accuracy assessment", *Journal of Thoracic Oncology*, vol. 5, no. 6, pp. 763-764, 2010.
- [24] K. J. Van Stralen, et al., "Diagnostic methods I: sensitivity, specificity, and other measures of accuracy", *Kidney international*, vol. 75, no. 12, pp. 1257-1263, 2009.