Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2024. Vol. 4. No. 1; e-ISSN: 2808-4519

Classification of Tile Productivity Data Based on Tile Type Using Random Forest Algorithm in Langkat Regency

Muhammad Arif Ridho[1]*, Relita Buaton [2], Muammar Khadapi[3]

[1, 2 ,3] *STMIK Kaputama*
muhammadarifridho2@gmail.com[1] , bbcbuaton@gmail.com[2], khdafi5@gmail.com[3]*

Abstract

This study aims to classify data on the productivity of census in Langkat Regency based on the type of census by using the Random Forest algorithm. Ubinan is a method used to measure the productivity of food crops, and in this study, the data was processed with various variables such as planting area, type of fertilizer, type of pesticide, and production volume. The Random Forest algorithm was used to build a classification model that could predict the productivity of the tares with very high accuracy, reaching 99.58% in the training stage. The model categorizes the productivity of the samples into several levels, namely Very Low, Low, Medium, High, and Very High. The implementation of this system is also equipped with a MATLAB GUI interface, which makes it easier for users to train and test data efficiently. With this system, users can see the prediction results through intuitive visualization. This research is expected to help farmers and policy makers in improving agricultural productivity through data-based analysis.

Keywords: *Random Forest, Classification, Cropping, Matlab*

1. Introduction

Food crops play an important role in the agricultural sector in Langkat Regency, North Sumatra. Crops such as rice and palawija are the backbone for thousands of farmers in the area. However, the agricultural sector in Langkat faces a major challenge, namely how to increase productivity to meet the growing market demand. Therefore, a deep understanding of the factors that affect productivity is crucial for farmers. Along with the advancement of information technology and data analysis, there is a great opportunity to adopt a data-driven approach in agricultural practices [1]. This study aims to investigate the relationship between tuber species and their productivity in Langkat Regency. By utilizing data analysis techniques and Random Forest algorithms, this study develops a prediction model that is able to identify productivity levels based on plant characteristics. The findings of this study are expected to provide new insights that are useful for farmers in increasing the productivity and efficiency of their agricultural production. The Central Statistics Agency (BPS) periodically conducts an Agricultural Survey every 10 years, with the last survey being carried out in 2023. One of the focuses of the survey is food crops, which aims to collect information related to 1 2 food crop productivity in various sub-districts. However, the data generated from the survey is often complex due to the variety of food crop types and the number of farmers differing. To deal with this complexity, the application of Data Mining technology is needed. The classification of tuber productivity based on plant type using the Random Forest algorithm is expected to help in analyzing farmers' planting patterns and provide an estimate of the productivity of food crops in each different subround. Data mining techniques allow the discovery of hidden patterns in large and complex data. By applying the Random Forest algorithm, this study aims to predict the productivity of agricultural land in Langkat Regency based on plant characteristics. It is hoped that the results of this study can provide useful information for farmers in making more data-based decisions, so that they can increase overall agricultural productivity.

2. Theoretical Foundations

2.1. Understanding Data Mining

Data Mining is an analytical process that aims to find knowledge from large and complex data sets. This technique is at the intersection of statistics and computer science. In more detail, data mining is the result of a combination of statistics, computer science, artificial intelligence, and machine learning [2]. Data mining is the process of analyzing large and complex data sets to find patterns and knowledge. It is an interdisciplinary field that combines techniques from statistics, computer science, artificial intelligence, and machine learning. Data mining involves methods such as classification, regression, grouping, and learning association rules to uncover useful information from data [3]. Data mining is an activity that involves the use and collection of data to find relationships or patterns in large data sets. This process includes in-depth data analysis using various techniques from the fields of statistics, computer science, artificial intelligence, and machine learning. Data mining aims to identify hidden trends, correlations, and anomalies that can provide valuable insights for decision-making. Methods often used in data mining 12 include classification, regression, clustering, and association analysis, all of which help in uncovering useful information from complex data [4].

2.2. Definition of Data

According to Susanto in the book Education management information system: Concepts, principles, and applications. Data is facts or information that can be used as input to generate new information. This data can be material for discussion, decision-making, calculations, or measurements. Data is no longer limited to text, but also includes sounds, still or moving images, in two or three dimensions. In addition, virtual data generated through computer engineering is now also increasingly used. Data in various forms is very important for analysis and utilization in various fields [4]. Data is a source of information that is still in its raw form. This means that the data has not been processed or analyzed so that it does not immediately provide insight or understanding. The data can be numbers, text, sounds, images, or videos, of which 16 are collected from a variety of sources. In its raw form, data needs to go through a process of cleaning, transformation, and analysis to become useful information. As a valuable resource, raw data is the basis for evidence-based decision-making in a variety of fields, including business, science, and technology [5].

According to Susanto in the book Education management information system: Concepts, principles, and applications. Data is facts or information that can be used as input to generate new information. This data can be material for discussion, decision-making, calculations, or measurements. Data is no longer limited to text, but also includes sounds, still or moving images, in two or three dimensions. In addition, virtual data generated through computer engineering is now also increasingly used. Data in various forms is very important for analysis and utilization in various fields [4]. Data is a source of information that is still in its raw form. This means that the data has not been processed or analyzed so that it does not immediately provide insight or understanding. The data can be numbers, text, sounds, images, or videos, of which 16 are collected from a variety of sources. In its raw form, data needs to go through a process of cleaning, transformation, and analysis to become useful information. As a valuable resource, raw data is the basis for evidence-based decision-making in a variety of fields, including business, science, and technology [6].

2.3. Definition of Tiles

According to [7] Ubinan is used as a survey approach that aims to evaluate plant productivity per hectare. Currently, the use of tubers to measure crop productivity is still limited to commodities such as rice, corn, soybeans, cassava, and sweet potatoes. In addition to recording productivity figures, ubinan also includes additional data such as land type, planting techniques, intensification rate, seed variety, number of seeds planted, and the use of fertilizers and pesticides, accompanied by qualitative information relevant to the productivity. In Indonesia, the estimated rice production is based on the harvested area reported by farmers in each sub-district, then multiplied by the productivity figures from the tuber survey conducted by the Central Statistics Agency (BPS) and the Agriculture Office at the district level. Not only covering productivity figures, the tuber survey also involves other supporting data such as land type, planting techniques, harvesting techniques, crop management organization, seed variety, number of seeds used, use of fertilizers and pesticides, and other qualitative information relevant to plant productivity.

2.4. Definition of Food Crops

In the journal [8] mentioned, Food crops are a diverse group of plants, which play a role in providing carbohydrates and proteins as the main source of energy for humans. They play an important role as a key aspect in food that provides vital energy 18 for the functioning of the human body. Generally, this food crop has a life cycle that lasts in one growing season. Food crops are a type of plant that is used as a source of foodstuffs, both directly consumed and after processing. They contain a variety of essential nutrients that are necessary for maintaining the health and survival of humans and other organisms. The diversity of these types of food crops enriches the variety of food available to the human population, as well as supports food security in various regions [8].

2.5. Definition of Classification

Classification in data mining is one of the techniques used to predict the class of a new object based on previously classified data. This technique aims to build a model that can map objects with certain attributes into predefined classes. This model can then be used to classify new objects with attributes that are not yet known to the class. Algorithms commonly used in classification include Decision Tree, Naive Bayes, Support Vector Machines, and Random Forest [9].

2.6. Random Forest Classification Algorithm

Random Forest is a machine learning algorithm used for classification and regression. This algorithm consists of a large number of individual decision trees that work as ensembles. Each tree in the Random Forest provides a class prediction, and the class with the most votes is the model's prediction. The basic idea of Random Forest is to 23 improve the performance of the model by combining the results of several different decision trees [9]. Random Forest works by creating a number of decision trees from a subset of training data. Each tree is trained using a random sample of data, and on each node within the tree, only a random subset of features is considered for separation. This helps to make the trees uncorrelated and improve the overall accuracy of the model [10], [11].

3. Analysis and Design

3.1. Definition of Classification

Classification in data mining is one of the techniques used to predict the class of a new object based on previously classified data. This technique aims to build a model that can map objects with certain attributes into predefined classes. This model can then be used to classify new objects with attributes that are not yet known to the class. Algorithms commonly used in classification include Decision Tree, Naive Bayes, Support Vector Machines, and Random Forest [12], [13].

3.2. Random Forest Classification Algorithm

Random Forest is a machine learning algorithm used for classification and regression. This algorithm consists of a large number of individual decision trees that work as ensembles. Each tree in the Random Forest provides a class prediction, and the class with the most votes is the model's prediction. The basic idea of Random Forest is to 23 improve the performance of the model by combining the results of several different decision trees [14]. Random Forest works by creating a number of decision trees from a subset of training data. Each tree is trained using a random sample of data, and on each node within the tree, only a random subset of features is considered for separation. This helps to make the trees uncorrelated and improve the overall accuracy of the model [15].

Table 1: Data Transformation Description

Column Name	Code	Information
Ubinan Name	1	Paddy Rice
	2	Paddy Farms
	3	Corn
	4	Soybean
	5	Cassava
	6	Sweet potato
Subround Tanam	1	Subround 1 (Januari - April)
	2	Subround 2 (Mei - Agustus)
	3	Subround 3 (September - December)
Planting Area	-	Planting area
Subround Panen	1	Subround 1 (January - April)
	2	Subround 2 (May - August)
	3	Subround 3 (September - December)
Planting Area	-	Harvest area
Production Volume	-	Production volume of ubinan
Types of fertilizers used	1	Urea Fertilizer
	2	NPK Fertilizer
	3	The Practical Blow
	4	Fertilizer from Animal Waste
	5	Other Non-organic Fertilizers
	6	Not Using Fertilizer
Types of Pesticides Used	1	Fungisida
	2	Herbisida
	3	Insecticide
	4	Rodentisida
	5	Other Types of Pesticides
	6	No Use of Pesticides
Categories Productivity	1	Very Low
	2	Low
	3	Keep
	4	Tall
	5	Very High

This data sharing process is important to ensure that the model can accurately predict outcomes when faced with new data that has never been seen before. Usually, the data is divided into two sets proportionally, for example 80% as training data and 20% as test data. This division can be done randomly to ensure that the data representation on both sets is even and unbiased. 56 By determining the right training and test data, as well as performing the necessary data transformations, the developed classification model can be optimized to provide more accurate and reliable prediction results. This division also helps in assessing how well the model can generalize from existing data to new data, which is an important indicator of the model's performance in real applications.

Table 2: Sample Training Data and Test Data

Ubinan Name	Subround	Planting Area	Subround	Harvest Area	Production Volume	Types of fertilizers used	Amount of Urea and NPK Fertilizers	Types of Pesticides Used	Categories Productivity
1	1	2000	1	1500	1800	1	5	3	5
1	1	1800	2	1300	1600	2	10	4	4
2	2	2100	2	1600	1900	1	5	2	5
3	3	2200	1	1700	2000	3	0	1	4
2	2	2500	3	1800	2100	4	15	5	3
1	2	2400	1	1900	2200	5	0	2	4
4	1	2000	2	1800	2000	6	0	3	2
5	3	3000	1	2500	2700	1	20	5	4
6	1	2700	2	2300	2500	2	10	4	5
3	2	2600	3	2000	2300	3	0	2	3

3.3. Bootstrapping

Bootstrapping is a resampling technique used to estimate statistics from a data by taking random samples from the original dataset with substitution. In the productivity classification of tubers using the Random Forest algorithm, bootstrapping is very important in the formation of individual decision trees that are part of the overall random forest model. In this step, a large number of samples of training data are randomly taken from the original dataset with replacement, so some data points may appear more than once in 3 samples, while others may not appear at all. Each of the 57 generated samples was used to train an individual decision tree in the Random Forest. This process helps to increase variability and reduce overfitting, as each tree in the forest is built on a different subset of data. By implementing bootstrapping, Random Forest can create several different models, which are then combined to produce a more stable and accurate final prediction. This technique allows Random Forest to handle large and complex datasets more effectively, providing a significant advantage in terms of accuracy and generalization capabilities. Through the use of bootstrapping, the ubinan productivity classification model can capture various patterns and relationships that may exist in the data, which ultimately improves the model's ability to predict ubinan productivity more accurately. From the original dataset with 10 rows, we'll create some bootstrap samples. We create 3 bootstrap samples each with 10 rows of data.

Table 3: Sample 1

Ubinan Name	Subround	Planting Area	Subround	Harvest Area	Production Volume	Types of fertilizers used	Amount of Urea and NPK Fertilizers	Types of Pesticides Used	Categories Productivity
1	1	2000	1	1500	1800	1	5	3	5
1	1	1800	2	1300	1600	2	10	4	4
2	2	2100	2	1600	1900	1	5	2	5
3	3	2200	1	1700	2000	3	0	1	4
2	2	2500	3	1800	2100	4	15	5	3
1	2	2400	1	1900	2200	5	0	2	4
4	1	2000	2	1800	2000	6	0	3	2
5	3	3000	1	2500	2700	1	20	5	4
6	1	2700	2	2300	2500	2	10	4	5
3	2	2600	3	2000	2300	3	0	2	2

3.4. Selection of Random Feature Subsets

In the Random Forest algorithm, the random selection of a subset of features is an important step to improve accuracy and reduce correlations between built decision trees. In this study, the researcher will use the "Type of Fertilizer Used" feature 59 as one of the main features in building a decision tree based on the bootstrap sample that has been created. Here are the steps to be taken:

Dataset and Feature Description "Types of Fertilizers Used"
1 = Urea Rifle
2 = NPK Rifle
3 = Organic Fertilizer
4 = Animal Waste Fertilizer
5 = Other Non-Organic Fertilizers
6 = No Fertilizer

Manual calculation using the "type of fertilizer used" feature
1.Calculating Initial Entropy
From sample 1, the distribution of "Productivity Category" is as follows:
Category 5: 3 data
Category 4: 4 data

Category 3: 2 data
Category 2: 1 data Total data = 10 Initial Entropy:
Entropy(S)=−(0.3×−1.737)−(0.4×−1.322)−(0.2×−2.322)−(0.1×−3.322)
Entropy(S)≈0.5211+0.5288+0.4644+0.3322≈1.8465

2. Calculate Entropy After Split Based on "Type of Fertilizer Used"
Based on the value of "Type of fertilizer used":
Fertilizer 1 (Urea): 2 data (Category 5, 4)
Fertilizer 2 (NPK): 2 data (Category 5, 4)
Fertilizer 3 (Organic): 2 data (Category 4, 3)
Fertilizer 4 (Animal Waste): 1 data (Category 3)
Fertilizer 5 (Other Non-Organic): 1 data (Category 4)
Fertilizer 6 (No Fertilizer): 1 data (Category 2)
Calculate the entropy of each subset and combine them to get the total entropy after the split.

3.Fertilizer 3 (Organic):
Total After Split
Entropy Setelah Split = 0.2+0.2+0.2+0+0+0 = 0.6
Gain(S,"Type of Fertilizer Used")=Entropy(S)−Entropy After Split Gain(S,"Type of Fertilizer Used") = 1.8465−0.6 = 1.2465
Building a decision tree based on "type of fertilizer used"
Root Node: "Type of Fertilizer Used"

Sample 1
Shot 1 (Urea):
Data: 2, Productivity Category: [5, 4], Prediction: 5+4 = 4.5 (rounded to 5)
Shot 2 (NPK):
Data: 2, Productivity Category: [5, 4], Prediction: 5+4 = 4.5 (rounded to 5)

Fertilizer 3 (Organic):
Data: 2, Productivity Category: [4, 3], Prediction: 4+3 = 3.5 (rounded to 4)

Fertilizer 4 (Animal Waste):
Data: 1, Productivity Category: [3], Prediction: 3 Fertilizer 5 (Other Non-Organic):
Data: 1, Productivity Category: [4], Prediction: 4 Fertilizer 6 (No Fertilizer):
Data: 1, Productivity Category: [2], Prediction: 2

Sample 2
Divide the data by "type of fertilizer used": Fertilizer 1 (urea): Productivity category: [5, 5], Prediction: 5
Fertilizer 2 (NPK): Productivity Category: [4, 5], Prediction: 5

Fertilizer 3 (Organic): Productivity Category: [4, 3], Prediction: 4
Fertilizer 4 (Animal Waste): Productivity Category: [3], Prediction: 3
Fertilizer 5 (Other Non-Organic): Productivity Category: [3], Prediction: 3 Fertilizer 6 (No Fertilizer): Productivity Category: [2], Prediction: 2

Sample 3
Divide the data by "type of fertilizer used": Fertilizer 1 (urea): Productivity category: [5, 5, 4], Prediction: 5
Fertilizer 2 (NPK): Productivity Category: [4, 5], Prediction: 5
Fertilizer 3 (Organic): Productivity Category: [3, 4, 4], Prediction: 4
Fertilizer 4 (Animal Waste): Productivity Category: [3], Prediction: 3 Fertilizer 5 (Other Non-Organic): - (no fertilizer data 5)
Fertilizer 6 (No Fertilizer): Productivity Category: [2], Prediction: 2

Testing with Test Data:

Table 4: Majority Vote for Final Prediction

No	Tree Prediction 1	Tree Prediction 2	Tree Prediction 3	Final Prediction (Majority Vote)	Actual Productivity Categories
1	5	5	5	5	5
2	4	5	5	5	4
3	5	5	5	5	5
4	4	4	4	4	4
5	3	3	3	3	3
6	4	4	4	4	4
7	2	2	2	2	2
8	4	4	4	4	4
9	5	5	5	5	5
10	3	3	3	3	3

Calculating Accuracy From 10 predictions:
Correct Predictions: 9 (All except row 2)
Incorrect Predictions: 1 (Row 2)
The Accuracy Result of the above calculation model is: 90%

4. Interface Discussion

In this section, the results of programming design made using the GUI (Graphical User Interface) interface of MATLAB will be explained. The GUI is designed to make it easier for users to operate the ubinan productivity prediction system without having to understand the technical details of programming. The interface consists of several main menus which include the main page, the training data menu, the test data menu, and the single data menu. The display of this menu will be explained further and presented in the following image.

4.1. Main Page Menu

The main page is the starting point when the user opens the app. On this page, users can select the available menu which is to start training data. The display of this menu is presented in the image below.

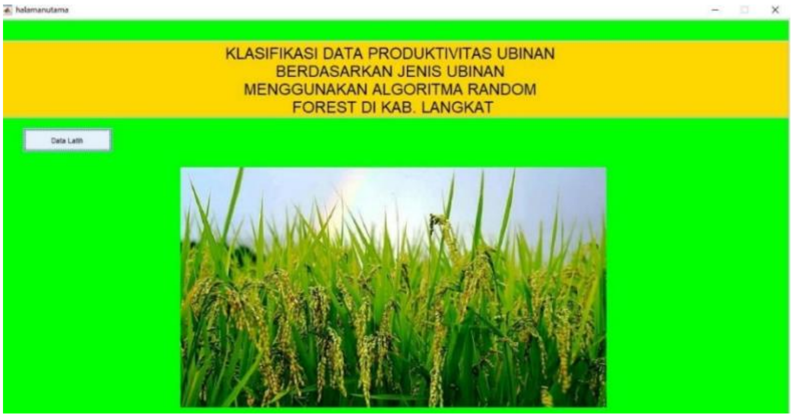

Fig. 1: Home Menu

4.2. Menu Training Data Menu

Training Data is designed as part of the system used to process training training training productivity prediction models. In this menu, the training data that has been collected and prepared will be input into the system to train the Random Forest model. By utilizing this training data, the system will build a pattern that is used to predict ubinan productivity. The display of this prediction menu can be seen in the image below.

Fig. 2: Training Data Menu

4.3. Test Data Menu

The Test Data menu serves as part of the system used to test the performance of a model that has been trained using training data. Through this menu, test data containing new ubinomia information will be input into the system to evaluate how well the model can predict productivity based on data that has never been seen before. The display of this prediction menu can be seen in the image below.

Fig. 3: Test Data Menu

4.4. Single Data Test Menu

The Single Data Test menu is used to perform model testing by manually entering a set of data. Users can enter data related to tuber productivity, such as the type of tubers, planting area, type of fertilizer, and production volume one by one. This menu is useful for testing how the model predicts the productivity of ubinan based on one specific data input. The display of this help menu can be seen in the image below.

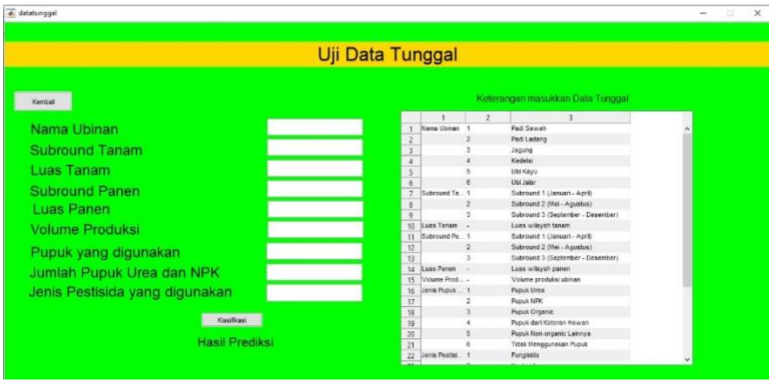

Fig. 4: Single Data Test Menu

4.5. Implementation

In this section, we will explain the steps taken in the trial of the Random Forest program using MATLAB to predict the productivity of ubinan. This test aims to determine the extent of the system's performance in processing data so that it can produce accurate predictions related to plant productivity. The data used in the training consisted of training data and training target data. This data will be trained using the Random Forest algorithm to generate a productivity prediction model. The Random Forest algorithm works by forming multiple decision trees from a subset of the training data and then making predictions through a majority voting mechanism of the entire decision tree. The training process is carried out until the model reaches an optimal level of accuracy. Once the model is trained using the training data, the process continues by entering the test data to start the testing phase. Test data is used to evaluate the model's performance in predicting productivity based on new data that has never been seen by 78 models. After the test is completed, the results of the tuber productivity prediction can be known. The implementation of the program to predict tuber productivity using MATLAB is as follows: To classify the productivity of tuber plants, users can input training data. Afterward, the train data button can be pressed. The stages of this process are presented in the following figure.

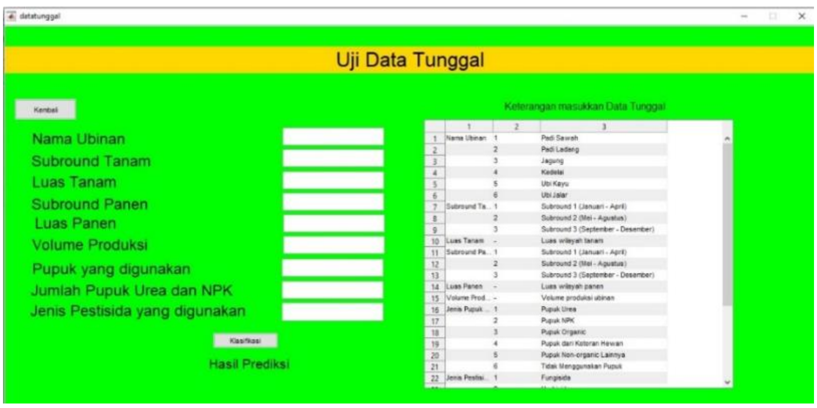

Fig. 5: Process

The data training trained is ubinan productivity. Once all the data has been inputted, the training process can begin. After the data training button is pressed, a message will appear if the data has been successfully trained and along with the accuracy of the data training as shown in the image below.

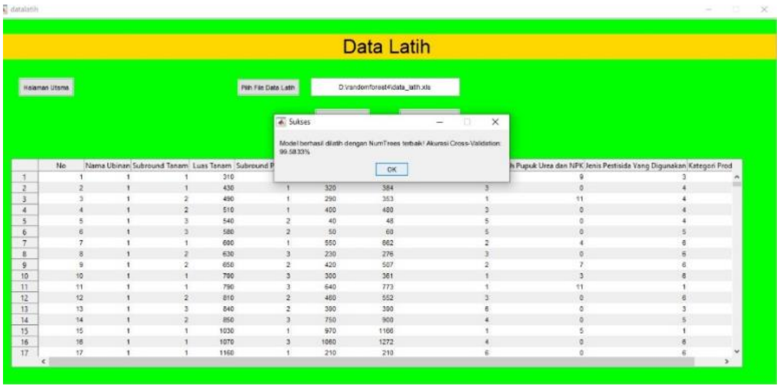

Fig. 6: Data Training Success Messages

From the training process above, the Random Forest model was successfully trained using the training data that had been uploaded. Based on the notification window that appears, the model has achieved the level of accuracy measured through the Cross Validation method, with an accuracy of 99.5833%. This shows that the model has been trained quite well, and is able to classify ubinan productivity data with a relatively low error rate. This notification provides information that the training process completed successfully, and the model is ready to be used to make predictions on test data or new data. The accuracy obtained can also be used as a reference for users in assessing the performance of the model before proceeding to the testing stage with other data.

Fig. 7: Rule Decison Tree

Tile Productivity Classification After training the data, the Rule Decision Tree Tile Productivity Classification can be evaluated through the decision tree structure shown in the image above. This tree shows various rules that are generated based on variables such as land area (Luas_Panen), production volume (Volume_Produksi), type of fertilizer used (Jenis_Pupuk_Yang_digunakan), and the name of the tuber (Nama_Ubinan). At each node, there is a separation based on a certain threshold value that leads to the next branches until it reaches the final decision node (yes/no). The classification process starts from the root node, where Luas_Panen becomes the main dividing variable with a threshold value of 925. If it is smaller than this value, the tree will continue to parse into other variables, such as Nama_Ubinaan or Volume_Produksi, until it reaches a leaf node that determines the productivity of the tubers based on the rules applied. This decision tree allows users to see which factors most affect the classification results and how combinations of those various variables can be used to predict outcomes effectively.

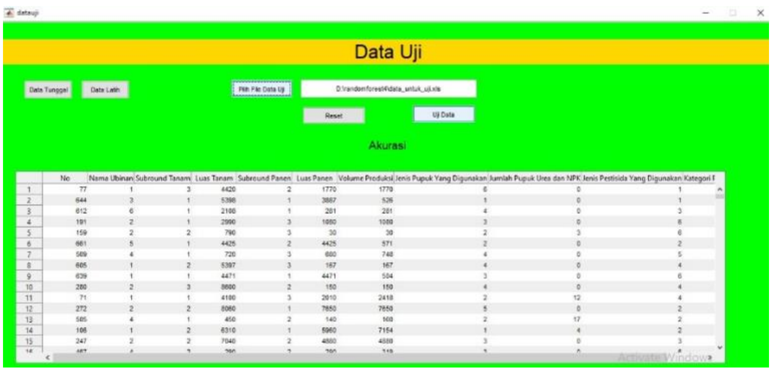

Fig. 8: Testing Process

Furthermore, to test data on other data, users can use the Test Data Menu as shown in the image above. In this menu, test data that is different from the training data is uploaded to test the performance of the prediction model that has been trained before. Users can select the test data file by using the Test Data File button, as seen with the file named "D:\rawdatatest\data_uji.xls" that has been selected. After the data is uploaded, the user can press the Test Data button to start the testing process, The uploaded test data includes the same variables as the training data. The Accuracy indicator at the bottom of the screen can be used to evaluate how accurate the model is in predicting results based on this test data. This testing process helps to evaluate whether the trained model is able to provide accurate predictions when faced with new data that is different from the training data. After that, users can head to the Single Data Menu to manually classify data by entering data one by one. This menu allows users to enter variables related to tuber productivity, such as the type of tubers, the area of 82 plants, the production volume, the type of fertilizer, and the type of pesticide are displayed manually, as seen in the image below.

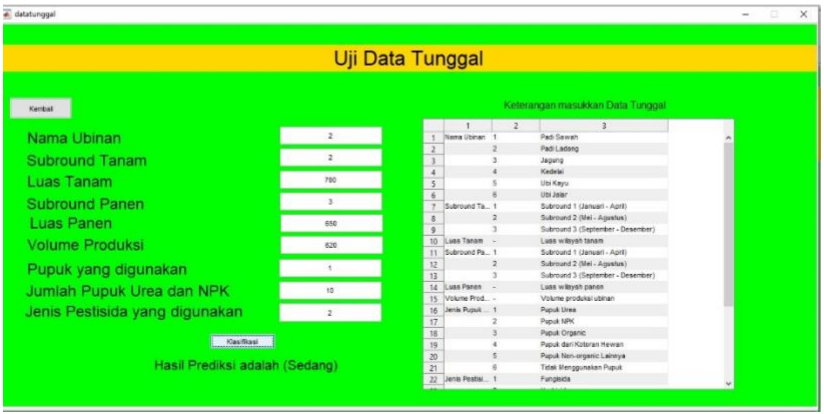

Fig. 9: Prediction Results

After the data is entered and the classification process is carried out by pressing the Classification button, the system generates a prediction for the productivity of the tubers based on the data that has been entered. The prediction results show that the productivity of tubers is in the Medium category. The right side of the screen also displays a Single Data Input Description, which guides the user in understanding the input values and the appropriate categories for each variable. It provides a direct reference for users to ensure that the data input is correct before running the prediction. Based on the analysis that has been carried out above, the ubinan productivity prediction model has been successfully tested using various approaches, including single data tests and data tests with larger datasets. From the results of a single data test, the system is able to generate predictions with the Medium productivity category, which is based on input variables such as planting area, production volume, type of fertilizer, and pesticide used.

These results show that the model built using the Random Forest algorithm has the ability to classify the productivity of ubinan quite accurately. The manual input process through Single Data Test allows users to test individual variables one at a time, thus providing flexibility in evaluating the prediction results for specific scenarios. In addition, through tests conducted using a larger dataset and evaluation using the Confusion Matrix, this model has been proven to be able to handle data variations well, with a high level of accuracy in classifying productivity categories from Very Low to Very High. Thus, this system can be an effective tool for farmers or related parties in predicting the productivity of tubers, as well as assisting in decision-making related to fertilizer use and land management.

5. Conclusion

Based on the research that has been conducted regarding the classification of tuber productivity in Langkat Regency using the Random Forest algorithm, several conclusions can be drawn as follows:

Random Forest Algorithm Model: The Random Forest algorithm has been proven to be effective in predicting the productivity of the tubers with a fairly high level of accuracy, reaching 99.58% during the training process. This model is able to classify productivity based on input variables such as planting area, type of fertilizer, pesticide, and production volume. Ubinan Productivity Classification: From the results of system implementation and testing, the model is able to divide ubinan productivity into several categories, namely Very Low, Low, Medium, High, and Very High. The prediction process based on test data and single data shows that the model can predict with high accuracy, which is proven through evaluation using the Confusion Matrix.MATLAB GUI-Based System: The use of the MATLAB GUI interface makes it easier for users to input data, conduct training, and test the model efficiently. The interface also provides intuitive visualization of results and makes it easier for users to understand the predictions given by the system. 4. Data Transformation Process: The data used in this study is transformed and processed in such a way that it fits the format that can be processed by the Random Forest model. This process helps in the identification of significant patterns in tuber productivity.

References

[1] Abidin, Z. (2022). Paradigma Usaha Tani Tanaman Ternak Terintegrasi Optimal: Perspektif Sumber Daya Terbatas dan Musim Berbeda. Penerbit Nem.

[2] Aini, Z. Implementasi Random Forest dan Gradient Boosting Pada Klasifikasi Indeks Pembangunan Manusia (Ipm) (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).

[3] Amalia, R. (2020). Penerapan data mining untuk memprediksi hasil kelulusan siswa menggunakan metode naïve bayes. Jurnal Informatika Dan Sistem Informasi, 6(1), 33 42.

[4] Apriliah, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. Sistemasi, 10 (1), 163.
[5] Dahlan, I. A. (2022). Klasifikasi Cuaca Provinsi Dki Jakarta Menggunakan Algoritma Random Forest Dengan Teknik Oversampling. Jurnal Teknoinfo, 16(1), 87-92. DR HA Rusdiana, M. M. (2021). Sistem informasi manajemen pendidikan: Konsep, prinsip, dan aplikasi.
[6] Fitrah Ilhami. Hartatik, H., Kwintiana, B., Nengsih, T. A., Baradja, A., Harto, B., Sudipa, I. G. I., ... & Gugat, R. M. D. (2023). DATA SCIENCE FOR BUSINESS: Pengantar & Penerapan Berbagai Sektor. PT. Sonpedia Publishing Indonesia.
[7] Helsalia, A., Pratama, H., Kristiani, M., & Marpaung, Y. B. (2021). Perancangan Aplikasi Pemesanan Obat di Apotek Dengan Analisis Design UML yang Menerapkan GIS dan LBS. J. Tek. Inform. Penerapan GIS dan LBS pada Anal. Des. UML, 1(1).
[8] Masdian, A. R., Bashit, N., & Hadi, F. (2023). ANALISIS PRODUKTIVITAS PADI MENGGUNAKAN ALGORITMA MACHINE LEARNING RANDOM FOREST DI KABUPATEN BATANG TAHUN 2018-2022. Elipsoida: Jurnal Geodesi dan Geomatika, 6(1), 43-51.
[9] Putra, R. F., Zebua, R. S. Y., Budiman, B., Rahayu, P. W., Bangsa, M. T. A., Zulfadhilah, M., ... & Andiyan, A. (2023). Data Mining: Algoritma dan Penerapannya. PT. Sonpedia Publishing Indonesia.
[10] Rezki, M., Kholifah, D. N., Faisal, M., Priyono, P., & Suryadithia, R. (2020). Analisis Review Pengguna Google Meet dan Zoom Cloud Meeting Menggunakan Algoritma Naïve Bayes. Jurnal Infortech, 2(2), 264-270.
[11] Ruslan, K. (2019). Memperbaiki Data Pangan Indonesia Lewat Metode Kerangka Sampel Area.
[12] Saleh, J., Adrian, A., & Sanger, J. (2022). SISTEM KLASIFIKASI KELULUSAN MAHASISWA DENGAN ALGORITMA RANDOM FOREST. Jurnal Ilmiah Realtech, 18(1), 10-14. Sari, I. R. F., & Utami, A. (2021). Rekayasa Perangkat Lunak Berorientasi Objek Menggunakan PHP. Penerbit Andi.
[13] Setiawan, Z., Fajar, M., Priyatno, A. M., Putri, A. Y. P., Aryuni, M., Yuliyanti, S., ... & Wijaya, A. (2023). Buku Ajar Data Mining. PT. Sonpedia Publishing Indonesia.
[14] Timisela, N. R., Turukay, M., Parera, W. B., & LAWALATA, M. (2021). Efisiensi relatif agroindustri pala banda dengan pendekatan Data Envelopment Analysis (DEA). SEPA: Jurnal Sosial Ekonomi Pertanian dan Agribisnis, 9(1).
[15] Triscowati, D. W., Sartono, B., Kurnia, A., Dirgahayu, D., & Wijayanto, A. W. (2020). Classification of rice-plant growth phase using supervised random forest method based on landsat-8 multitemporal data. International Journal of Remote Sensing and Earth Sciences (IJReSES), 16(2), 187-196.