The 2nd International Conference on Multidisciplinary Engineering
(2nd ICOMDEN 2024)

The 2nd International Conference on Multidisciplinary Engineering: 00049 (2024) https://doi.org/10.29103/icomden.v2.xxxx
Volume 2, 2024
eISSN: xxxx-xxxx
Research Original Article

Implementation of the Support Vector Machine Method in Determining the Best Quality of Sap

Azhari Putra Sayani[1] **Safwandi**[2] **Fajriana**[3]
[1]Department of Informatics Enginering, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, azhari.200170259@mhs.unimal.ac.id
[2]Department of Informatics Enginering, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, safwandi@unimal.ac.id
[2]Department of Informatics Enginering, Universitas Malikussaleh, Bukit Indah, Lhokseumawe, 24353, Indonesia, fajriana@unimal.ac.id
Corresponding Author: azhari.200170259@mhs.unimal.ac.id **| Phone: +6282288828716**

Abstract

Rubber trees (Hevea brasiliensis) are the main source of natural rubber and play an important role in Indonesia's industry. Determining the quality of rubber sap is a challenge for companies, as traditional manual processes are time-consuming and prone to human error. PT Poly Kencana Raya, a company in Besitang, North Sumatra, currently still uses conventional methods in determining the quality of rubber latex it produces. This research aims to design a web-based system with the application of the Support Vector Machine (SVM) method to facilitate the determination of rubber latex quality. SVM was chosen as a classification method because of its ability to determine the optimal hyperplane that can separate data from two different classes, namely feasible and unfit. The built system utilizes the main criteria such as tree age, tapping time, moisture content, color, and texture in determining the quality of the sap. Implementation. This study used 120 samples of test data, with accurate prediction results on 111 data, resulting in an accuracy rate of 92.5%. This decision support system is expected to increase efficiency and accuracy in rubber sap selection and support the development of rubber production quality in Indonesia. This research also opens up opportunities for further development by adding other classification methods for accuracy comparison or adding training data to optimize prediction results.
Keywords: Rubber Trees, Support Vector Machine, Data Mining

Introduction

The selection of high-quality rubber latex is essential and poses a unique challenge for companies engaged in industrial sectors. Currently, assessments for sorting quality latex still rely on traditional, manual methods, which require more time and effort and can impact product quality [1]. In processes like these, there is also a high risk of errors due to human error. Rubber trees contribute to the nation's economy, especially in the industrial sector, as rubber is widely used across various industries, including automotive and textile. Thus, maintaining rubber quality is crucial for competitiveness between companies.

PT Poly Kencana Raya, located in Besitang, North Sumatra, is one of the companies processing rubber latex in Indonesia and plays a role in Indonesia's economic landscape. However, the company currently uses conventional methods to determine the quality of rubber latex, without computerized technology to facilitate quality assessment, resulting in a relatively lengthy evaluation process. The desired technology would be a decision support system to streamline and improve efficiency in quality selection. Data mining have been widely adopted for their efficiency in problem-solving. To increase accuracy and efficiency in latex selection, machine learning could be a promising option.

In previous research with the title Decision Support System for Determining the Quality of Rubber Sap Using the COPRAS Method Decision Support System Determining the Quality of Rubber Sap Using the COPRAS Method. The calculation results show that PB 260 sap is the best option. In the implementation of the login form, main menu, alternative data, criteria data, COPRAS process, and reports. The conclusion shows that the results of system calculations with manual techniques are comparable. The purpose of this research is to improve the quality of employee performance in rubber latex processing companies in North Sumatra Province by using the COPRAS method, which was chosen because it has the ability to make optimal decisions by considering all positive and negative aspects of the method.

One effective machine learning method for such applications is the Support Vector Machine (SVM). In previous research with the title Classification of Dental and Oral Diseases Using the Support Vector Machine Method. The results showed that the Support Vector Machine (SVM) method provided the best accuracy of 94.442% in the classification of dental and oral diseases. Compared to other kernels, the RBF kernel gives the best results. Convergence of classification results is influenced by the number of iterations. It is hoped that this system will help in the early diagnosis of oral and dental diseases [2].

Previous research related to the SVM method entitled Twitter Sentiment Analysis Regarding the Implementation of Permendikbudristek Number 30 of 2021 Using TextBlob and Support Vector Machine. The research was carried out by dividing training data and test data using a ratio of 70: 30 randomly by the system. Where the training data is 111 data, the test data is 48 data. Based on the results of testing with textblob assisted by applying the Support Vector Machine algorithm using 48 test data with 34 positive data, and 14 negative data, this study obtained an accuracy of 70.8% [3].

Based on these problems, it is necessary to conduct research to classify worthy and unworthy rubber sap based on quality, in order to make it easier and become an option for PT Poly Kencana Raya in choosing the quality of rubber sap.

Literature Review

1. Rubber Trees

Rubber sap is an important plantation crop that produces rubber sap (Hevea brasiliensis). Rubber sap is a sap obtained from the trunk of a rubber tree. Rubber sap is one of the leading commodities in Indonesia. Usually the quality of rubber sap is determined by high dry rubber content (KKK) with low water content. Dry rubber content can also be used as a parameter to assess the productivity of rubber trees. A good tree will produce high levels of latex [4].

Rubber plants become raw materials that are widely used as the main ingredient in the manufacture of various productions, so the quality of the rubber latex must be considered properly. In connection with this quality, there are several things that must be considered including tree age, tapping time, moisture content, color and texture [5].

2. Data Mining

According to [6], data mining is a tool that allows users to quickly access large amounts of data. A more specific understanding of data mining, namely a tool and application using statistical analysis on data. Data mining is a process of extracting or extracting large, previously unknown, but understandable and useful data and information from large databases and is used to make very important business decisions Data mining describes a collection of techniques with the aim of finding unknown patterns in the data that has been collected. Data mining allows users to discover knowledge in databases that would not otherwise be known to exist.

3. Support Vector Machine (SVM)

SVM is a machine learning method that aims to find the best hyperplane that separates two classes in the input space. The SVM classification algorithm uses training data to form a classification model, the model formed is used as a prediction of new data classes that have never existed before called testing data [7].

The Support Vector Machine method itself is divided into two, namely Linear Support Vector Machine and Non-Linear Support Vector Machine. The basic principle of a linear classifier, namely the case of linearly separated data classification by separating the two classes on a hyperplane with soft margins. However, SVM has been developed to work on non-linear problems by applying the function of the kernel trick to high-dimensional space [8].

$$f(x) = w_1 * x_1 + \cdots + wn * xn + b \qquad (1)$$

Description:
$f(x)$ = Predicted value
w = Weight value
x = Input value
b = Bias value

Here is the formula for Support Vector Machine using the Gradient Descent algorithm:

$$w_j = w_j + \eta \times \left(y_i - f(x_i)\right) \times x_{ij}$$

$$b = b + \eta \times \left(y_i - f(x_i)\right)$$

Description:
η = Learning rate (e.g., $\eta = 0{,}01$)
y_i = Actual target class value
$f(x_i)$ = Predicted result
x_{ij} = j-th feature of the i-th data point

4. Confusion Matrix

The confusion matrix is a table that shows the classification of the number of correct and incorrect test data. confusion matrix is used in the evaluation to get classification results that are divided based on actual data and prediction data [9].

Table 1. Confusion Matrix

	Positive Prediction	Negative Prediction
Actual Positive	True Positive (TP)	False Negative (FN)

Actual Negative	False Positive (FP)	True Negative (TN)

Materials & Methods

1. Data Collection

The data was obtained from PT Poly Kencana Raya, with a total of 600 data with 480 training data and 120 test data. This calculation will produce output that shows the results of feasible and infeasible, along with the accuracy level of the svm method used.

Table 2 Data for Classification Process

No.	Tree Age	Tapping Time	Moisture Content	Color	Texture	Acceptability
1	16 years	05.00	55	White - Yellow	Very Soft	Acceptable
2	17 years	06.00	45	White – Yellow	Soft	Acceptable
3	16 years	05.00	45	White – Yellow	Soft	Acceptable
4	20 years	06.00	50	White – Yellow	Soft	Acceptable
5	16 years	05.00	45	White - Yellow	Soft	Acceptable
…	…	…	…	…	…	…
600	16 years	06.00	55	White - Yellow	Soft	Acceptable

2. Method

The system scheme that will be created is as follows:

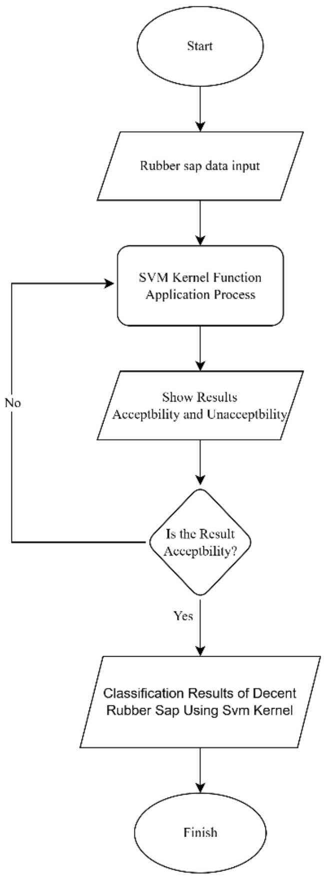

Figure 1 System Scheme

The system starts by running the initial process of the designed system. After that, rubber sap data is inputted into the system to start the classification process. Next, the SVM kernel function is applied to the input data, where this method is used to classify the rubber sap data. The result of the classification is then displayed, with information on whether the rubber sap data meets the eligibility requirements or not. If the results show that the data is eligible, the system will display the results of the classification of eligible rubber sap. However, if the result is not eligible, the system will return to the process of applying the SVM kernel function to start the classification again. Finally, the system displays the result of the proper rubber sap classification and the process is completed.

Results and Discussion

From the data below, the criteria of tapping time, color, and texture will be weighted first.

Table 3 Conversion Data Label

No.	Tree Age	Tapping Time	Moisture Content	Color	Texture	Acceptability
1	5	4	5	4	4	1
2	4	3	3	4	3	1
3	5	4	3	4	3	1
4	1	3	4	4	3	1
5	5	4	3	4	3	1
…	…	…	…	…	…	…
600	5	3	5	4	3	1

Table 4 Data Normalization

No.	Tree Age	Tapping Time	Moisture Content	Color	Texture
1	1	1	1	1	1
2	0,75	0,666666667	0,5	1	0,666666667
3	1	1	0,5	1	0,666666667
4	0	0,666666667	0,75	1	0,666666667
5	1	1	0,5	1	0,666666667
…	…	…	…	…	…
600	1	0,666666667	1	1	0,666666667

After having normalized value data, then build the SVM model to predict the best rubber sap quality data. The formula used is equation (1). In the construction of the model has a requirement for the formation of weights in seeing the results of iterations performed. To determine whether data is in class 1 or -1, SVM calculates the prediction results in a way:

If $(x) \geq 0$, then predict class 1 (Acceptbility).
If $(x) < 0$, then predict class -1 (Unacceptbility).

In the first iteration, the value of y in the first data is Acceptbility or 1. For $(x0)$, it is 0, which indicates the prediction of class 1. The prediction results are in accordance with the first iteration, so the weights and biases do not need to be updated. Conversely, if the results of the prediction and class values are different, it is necessary to update the weights and biases. SVM uses an optimization algorithm such as *Gradient Descent* to update the weight w and bias b based on the prediction error. The weight update formula for each iteration is as follows:

$$w j = w j + \eta \times (y i - f (x i)) \times x i j$$

$$b = b + \eta \times (y i - f (x i)$$

Table 5 Result Of SVM Calculation

F(x).	w1	w2	w3	w4	w5	b
0	0	0	0	0	0	0
0,01	0	0	0	0	0	0
0,0199	0	0	0	0	0	0
0,0297	0	0	0	0	0	0

0,0394	0	0	0	0	0	0
…	…	…	…	…	…	…
0,57215	-0,1991	0,4093	-0,1781	-0,2812	-0,1763	1,62508

Table 6 Results of Test Data

No.	Tree Age	Tapping Time	Moisture Content	Color	Texture	Acceptability
1	1	1	0,5	0,666666667	0,666666667	1
2	0,25	0,333333333	0	0,333333333	0,333333333	-1
3	0,75	1	0,5	1,666666667	0,666666667	1
4	1	0,666666667	1	1	1	1
5	1	1	0,5	0,666666667	0,666666667	1
…	…	…	…	…	…	…
120	1	0,666666667	1	1	0,666666667	1

Table 7 Confusion Matrix Results

	Positive Prediction	Negative Prediction
Actual Positive	True Positive = 111	False Negative = 0
Actual Negative	False Positive = 9	True Negative = 0

Based on table 7, the results obtained from the test data amounted to 111 prediction data that matches the actual data, and 9 prediction data that does not match the actual data.

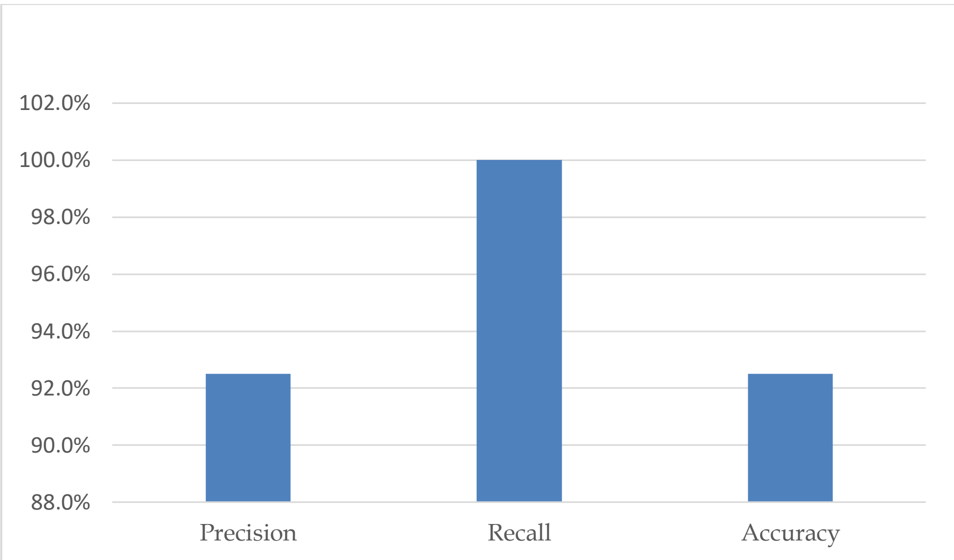

Figure 2 Evaluation SVM of Confusion Matrix

In Figure 2, it can be seen that the accuracy results of research using Support Vector Machine (SVM) obtained a precision value of 92,5%, recall 100%, and accuracy of 92,5%.

Conclusions

The conclusion that can be drawn from the results of this study is that the best rubber sap selection classification system is calculated using the st Vector Machine (SVM) algorithm and with criteria: tree age, tapping time, moisture content, color, texture. In addition, this study uses 120 rubber sap test data, where the prediction results that match the actual consist of 111 data and 9 test data that do not match the actual. The implementation of the Support Vector Machine (SVM) algorithm allows the classification of unfit rubber sap with a high level of accuracy. SVM is a machine learning method that is very effective in classification, and in this context, SVM can predict whether the incoming rubber sap at PT Poly Kencana Raya falls into the worthy or unworthy category. The accuracy results obtained in using the Support Vector Machine (SVM) method resulted in an accuracy of 92.5%.

Acknowledgments

The author wishes to extend his deepest gratitude to Mr. Safwandi, S.T., M.Kom, as the primary supervisor, for his invaluable guidance, constructive feedback, and unwavering support throughout this research journey. His expertise and mentorship have been instrumental in shaping this work. The author also expresses heartfelt thanks to Dr. Fajriana, S.Si., M.Si., as the secondary supervisor, for her insightful advice, encouragement, and patience, which greatly contributed to the completion of this study. Additionally, the author would like to acknowledge the support of friends and peers who provided motivation, collaboration, and understanding during the research process. Their encouragement and camaraderie made a meaningful impact, and for that, the author is deeply grateful.

References

[1] A. Anriani, I. F. Astuti, and F. Agus, "Model Penentuan Kualitas Getah Karet Menggunakan Metode Multi Attributive Border Approximation Area Comparison (Mabac)," *IKRA-ITH Inform. J. Komput. dan Inform.*, vol. 8, no. 2, pp. 34–43, 2024, doi: 10.37817/ikraith-informatika.v8i2.2954.

[2] N. W. S. Agustini, D. Priadi, and R. V. Atika, "Profil Kimia dan Aktivitas Antibakteri Fraksi Aktif Nannochloropsis sp. sebagai Senyawa Penghambat Bakteri Penyebab Gangguan Kesehatan Mulut," *J. Pascapanen dan Bioteknol. Kelaut. dan Perikan.*, vol. 17, no. 1, p. 19, 2022, doi: 10.15578/jpbkp.v17i1.781.

[3] F. O. Dayera, Musa Bundaris Palungan, "G-Tech : Jurnal Teknologi Terapan," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 186–195, 2024, [Online]. Available: https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/122

[4] N. I. Ningrum, A. Azanuddin, and D. Suherdi, "Sistem Pendukung Keputusan Menentukan Kualitas Getah Karet Menggunakan Metode COPRAS," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 4, p. 374, 2022, doi: 10.53513/jursi.v1i4.5346.

[5] D. Aldo, "Rekomendasi Kualitas Getah Karet Terbaik Berbasiskan Sistem Pendukung Keputusan dengan Metode MAUT," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1592–1602, 2022, doi: 10.47065/bits.v4i3.2631.

[6] I. Ahmad, S. Samsugi, and Y. Irawan, "Implementasi Data Mining Sebagai Pengolahan Data," *J. Teknoinfo*, vol. 16, no. 1, p. 46, 2022, [Online]. Available: http://portaldata.org/index.php/portaldata/article/view/107

[7] E. Suryati, Styawati, and A. A. Aldino, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," *J. Teknol. Dan Sist. Inf.*, vol. 4, no. 1, pp. 96–106, 2023, [Online]. Available: https://doi.org/10.33365/jtsi.v4i1.2445

[8] W. R. U. Fadilah, D. Agfiannisa, and Y. Azhar, "Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine," *Fountain Informatics J.*, vol. 5, no. 2, p. 45, 2020, doi: 10.21111/fij.v5i2.4449.

[9] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021.