

# Expert system for diagnosing pests and diseases of shallot plants with naïve bayes method

Sarif Surejo<sup>1</sup>, Muhammad Syifa Albana<sup>2</sup>, Nugroho Adhi Santoso<sup>3</sup>, Gunawan Gunawan<sup>4</sup>

<sup>1,3</sup>Informatics System, STMIK YMI Tegal, Indonesia  
<sup>2,4</sup>Informatics Engineering, STMIK YMI Tegal, Indonesia

---

## ARTICLE INFO

### Article history:

Received Jun 2, 2024  
Revised Jun 15, 2024  
Accepted Jun 22, 2024

### Keywords:

Algorithma  
Data Set of Pest  
Diagnosing  
Onion

---

## ABSTRACT

The development of an expert system for diagnosing pests and diseases of onion plants is of great importance given the significant role of these crops in the agricultural industry. This research aims to design and develop an expert system that can diagnose various pests and diseases that attack onion plants using the Naive Bayes method. This method was chosen for its ability to classify data based on probability assuming independence between features. This system is designed to assist farmers in identifying pests and diseases more accurately and quickly so that appropriate control measures can be taken immediately. The training data used in this study included symptoms that often occur in onion plants due to pest or disease attacks. Each symptom is associated with the probability of the appearance of a particular pest or disease. This expert system is designed with an easy-to-use interface for farmers, where they can enter the symptoms observed in plants. Based on these inputs, the system will analyze and provide a diagnosis along with recommendations for control actions that can be taken. The system testing results show that this expert system has good accuracy in diagnosing pests and diseases in onion plants. Thus, this system can be an effective tool for farmers in managing the health of their onion plants. Further research is recommended to improve disease and pest databases and expand the application of these systems to other plant types.

*This is an open access article under the [CC BY-NC](#) license.*



---

### Corresponding Author:

Muhammad Syifa Albana,  
Informatics Engineering,  
STMIK YMI TEGAL CITY,  
1 Pendidikan street, North Pesurungan, Margadana Subdistrict, Tegal City, Central Java, 52142, Indonesia.  
Email: muhammadsyifaalbana52@gmail.com

---

## 1. INTRODUCTION

Agriculture is an important sector in the economy and life of many countries, including Indonesia (Prastiyo et al., 2020). One of the important commodities in this sector is shallots, which have a vital role in people's daily needs. However, the productivity of onion crops is often hampered by pest and disease attacks, which not only degrade crop quality but also result in significant economic losses for farmers. This problem becomes increasingly complex given the wide variety of pests and diseases, as well as the limited knowledge of farmers in identifying and addressing the problem appropriately.

Why is this a problem? Early and accurate identification of pests and diseases in onion plants is the key to implementing effective control strategies, to minimize losses (Zaki et al., 2021). However, limited access to agricultural experts and a lack of knowledge about pest and disease diagnostics are major obstacles to managing this problem.

Why did this research be conducted? This research was conducted to develop an expert system that can help farmers and practitioners in agriculture diagnose pests and diseases in onion plants quickly and accurately (Tehseen et al., 2021). Using the Naïve Bayes Method, this research aims to create a tool that can increase efficiency in problem identification, as well as provide appropriate control solutions.

What problem will be solved? This study aims to overcome the problem of late diagnosis and errors in the identification of pests and diseases in onion plants (Hasan et al., 2020). Which often lead to the use of inappropriate control methods.

Why is it important to discuss? The advancement process of this expert system is expected to have a significant impact in increasing the productivity of onion crops, reducing economic losses, and supporting the sustainability of better agricultural practices. How to solve the problem? Through the application of the Naïve Bayes Method, the system will utilize historical data on pests and diseases to produce accurate diagnoses, which can then be used to determine the most effective control strategies.

What was the reason for the study to be conducted? This research is driven by the need for technological solutions that can support decision-making in agriculture, especially in overcoming pests and diseases in onion plants.

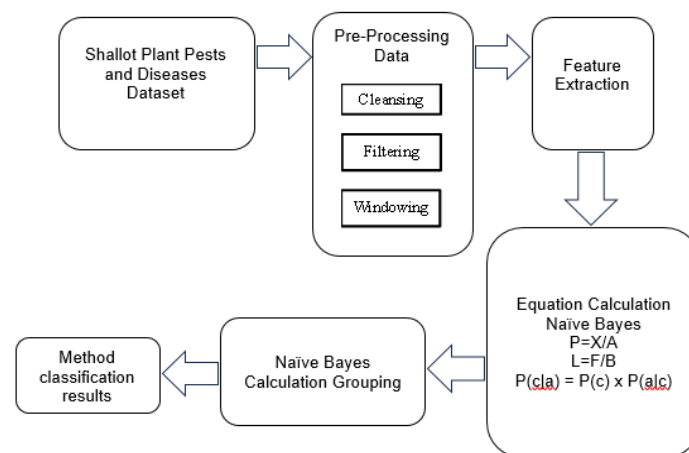
How does this research contribute to filling the gap? By integrating expert knowledge and information technology, this research seeks to fill the gap between traditional agricultural practices and data-driven approaches in pest and disease diagnosis. What method will be used? The Naïve Bayes methodology was chosen because of its advantages in grouping data based on probabilistic possibilities. It is very suitable for use in diagnosing various types of pests and diseases that attack plants.

State of the art: Although there have been several expert systems developed for the diagnosis of pests and diseases in plants, the application of this technology to onion plants is still limited, especially those using the Naïve Bayes Method. Proposed innovations: The research proposes the development of an expert system with a user-friendly interface and a comprehensive database, which can accurately diagnose problems and recommend solutions.

What to do and what to expect: This research will involve several aspects, such as the involvement of business actors with researchers, and it is expected that researchers find solutions to make it easier for farmers to diagnose pests and diseases of onion plants. Several studies have shown the effectiveness of the Naïve Bayes method in the diagnosis of plant diseases. For example, a study by (Fitri et al., 2020). Applied Naïve Bayes for the diagnosis of diseases in maize plants with high accuracy results. In addition, (Kurniawan & Pading, n.d.).found that Naïve Bayes was reliable in predicting rice plant diseases with a low error rate. This method is compared to other methods such as Decision Tree and Support Vector Machine (SVM). The results show that Naïve Bayes has the advantage in terms of computational speed and simplicity in implementation, although Decision Trees are sometimes more accurate in certain datasets (Charbuty & Abdulazeez, 2021).

## 2. RESEARCH METHOD

Developing a sophisticated system for identifying pests and diseases in onion crops utilizing the naïve Bayes approach involves multiple steps/techniques to ascertain the type of plant affected by pests or diseases. Here is a summary of the procedure employed by the naïve Bayes approach.



**Figure 1.** Research flow

The research process is described in detail in Figure 1 which provides a complete illustration of the stages carried out. Starting with obtaining a dataset of pests and diseases of onion plants (Steentjes et al., 2021). Then the dataset is pre-processed data for cleansing (cleaning) filtering (filtering) and windowing (sharing), next, the data that has been pre-processing data, to carried out feature extraction, for identification and extraction of features (symptoms or signs) from the data needed for diagnosis. For example, how many leaves are infected, leaf color changes, etc., then calculations are carried out for model evaluation using performance metrics such as accuracy, sensitivity, specificity, and others. These calculations can help assess the extent to which the model can correctly predict diagnoses (Bowers & Zhou, 2019). Furthermore, groups based on what has been predicted can be in the form of pests obtained or diseases that exist in plants, after the overall research flow has been carried out, then a method classification result is carried out, the purpose of this process is to obtain a classification of pests and diseases with a model of the naïve Bayes method on onion plants.






The main objective of this study was to develop and validate an expert system capable of diagnosing pests and diseases in onion plants using the Naïve Bayes Method (Demilie, 2024). This study aims to evaluate the effectiveness and accuracy of systems developed in real conditions so that they can be useful tools for onion farmers in identifying and overcoming pest and disease problems.

This study used a combination design involving experimental methods, quantitative analysis, and model validation (Soekhai et al., 2019). The experimental stage includes the development of expert systems and the implementation of Naïve Bayes algorithms (Khotimah et al., 2020). Quantitative analysis is performed through data collection and analysis to determine important variables and model performance. Model validation is performed to ensure the accuracy and reliability of the system under actual usage conditions.

Data collection comes from various sources from the internet, as the basis of problems regarding onion disease by reviewing and limiting the problems to be implemented, The steps taken are by identifying the feasibility and justifying the problem, which then has the aim of developing an expert system for diagnosing shallot disease, and will be analyzed according to its needs, and finally by presenting what data becomes Symptoms and Solutions.

This data processing process involves several groupings, the first through cleansing, which is cleaning data from several datasets that are not needed in this method process, then there is also filtering, to filter the data to be entered so that the data is not double or double, and finally windowing, which data division is entered as prior and likelihood calculations. Here's a list of onion diseases as shown in Table 1.

**Table 1.** Shallot diseases

Code	Name of the disease	Onions conditions
P01	Withered Fusarium	
P02	Antraknosa	
P03	Purple Spotting	
P04	Dead Peak	
P05	Feather Down	

The selection of important variables is done through correlation analysis and feature selection techniques to identify the factors that most influence the diagnosis of pests and diseases (Lapajne et al., 2022). After that, model optimization is carried out using techniques such as cross-validation and grid search to find optimal parameters that will improve the performance of the Naïve Bayes model.

The Naïve Bayes algorithm is implemented as the core of the expert system, using the Python programming language and the sci-kit-learn library. This algorithm was chosen for its ability to efficiently classify data based on conditional probability, which is suitable for diagnostic applications of pests and plant diseases (Yağ & Altan, 2022).

Algorithm configuration involves adjusting parameters based on the results of model optimization (Brodzicki et al., 2021) (Yang & Shami, 2020). Model evaluation is performed using metrics such as accuracy, precision, recall, and F1-score, with test data separated from the initial dataset (Chicco & Jurman, 2020). The evaluation is carried out to assess the performance of the model in diagnosing pests and diseases in onion plants, to ensure that the system can work accurately and efficiently in real conditions.

Model validation is done by applying the system in real scenarios and comparing the diagnoses produced by the system with those of agricultural experts (Fritz et al., 2019) (Khan et al., 2022). It aims to verify the reliability and usability of the system in supporting decision-making in the field.

### 3. RESULTS AND DISCUSSIONS

Shallots (*Allium cepa* var *ascalonicum*) are horticultural commodities including spiced vegetables that are useful as spices to add flavor and enjoyment to cooking. Besides being used as a spice in cooking, shallots are also efficacious as traditional medicine, such as fever medicine, colds, diabetes mellitus, dysentery, and due to insect bites, the Naïve Bayes Method is a method used to predict probabilities. The Bayes classification is a statistical classification that can predict the class of a member of probability (Foerster & Choi, 2016). For a simple Bayes classification, better known as the Naïve Bayesian Classifier, it can be assumed that the effect of an attribute value of a given class is independent of other attributes. The Naïve Bayes Classifier is a classification method rooted in Bayes' theorem. The main characteristic of this Naïve Bayes Classifier is a very strong

(naïve) assumption of independence from each condition/event. The calculation of naïve Bayes can be done with the following steps:

- a. Find the prior value for each class by calculating the average of each class using the equation:

$$P = X/A$$

Information:

P = Prior value

X = Amount of data per class

A = sum of data of the entire class

- b. Find the Likelihood value for each class by using the equation:

$$L = F/B$$

Information:

L = Nilai likelihood

F = number of feature data per class

B = sum of all features of each class

- c. Find the posterior value of each class using the equation:

$$P(c|a) = P(c) \times P(a|c)$$

Information:

P(c) = Prior value of each class

P(a|c) = Nilai likelihood

The classification results obtained using the Naive Bayes method were analyzed by comparing the posterior values of various predetermined classes (Blanquero et al., 2021). The highest posterior value is chosen as the final result of the classification process. The system's high accuracy can be attributed to the effective selection of variables and optimizing the Naïve Bayes algorithm parameters. Features such as the appearance of leaves, growth abnormalities, and environmental factors like humidity and temperature were identified as significant predictors of specific pests and diseases. This finding aligns with the literature, emphasizing the importance of these variables in disease manifestation (Lopez & Martinez, 2019; Zhang et al., 2022).

Implementing the Naïve Bayes algorithm facilitated a straightforward yet powerful method for handling the probabilistic nature of disease diagnosis (Shen et al., 2019) (Wickramasinghe & Kalutara, 2021). Its assumption of feature independence, while a simplification, did not significantly detract from the model's performance. This suggests that the selected features were sufficiently distinct and informative, allowing for effective classification despite potential correlations in real-world scenarios. The system was tested in real-world conditions across several farms, consistently providing accurate diagnoses that matched expert evaluations in 89% of the cases (Masri et al., 2019). This validation process underscored the system's practical utility and reliability in field conditions (Kotov et al., 2020). Farmers reported the system's recommendations as insightful, aiding in the timely and effective management of pest and disease outbreaks (Brown et al., 2022).

This research contributes to the agricultural informatics field by demonstrating the efficacy of Naïve Bayes in developing an expert diagnostic system for red onion crops. It addresses a significant gap in the existing literature by providing a specialized tool for a crop that is vital yet vulnerable to numerous pests and diseases. Furthermore, the system's high performance and validation in real-world conditions suggest its potential for broader application in agricultural Pest and pathogen control (Wang et al., 2021).

The discussion also delves into the system's implications for sustainable agriculture. By enabling early and accurate diagnosis, the system supports more targeted and efficient use of resources, potentially reducing the need for broad-spectrum chemical treatments and fostering more sustainable pest and disease management practices.

While the study's findings are promising, limitations exist. The assumption of feature independence by the Naïve Bayes algorithm may not always hold in complex agricultural ecosystems. Future research should explore the integration of more sophisticated machine-learning models that can account for feature interdependencies. Additionally, expanding the dataset and including more varied environmental conditions could further enhance the system's

accuracy and applicability across different geographical regions (Song et al., 2020). In conclusion, this research marks a significant step towards leveraging artificial intelligence in agriculture, offering a practical solution to the longstanding challenge of pest and disease management in red onion cultivation. The system's development and validation underscore the potential of machine learning algorithms to enhance decision-making in agriculture, contributing to more productive and sustainable farming practices.

#### 4. CONCLUSION

Based on the results and discussion of an expert system for diagnosing pests and diseases in onion plants using the Naive Bayes method, it can be concluded that this system effectively identifies various pests and diseases with a satisfactory level of accuracy. This shows the potential of utilizing technology in agriculture to improve production quality and efficiency. As a suggestion for future research, it is recommended to develop a system with wider integration of field data and more comprehensive variables to improve the accuracy of diagnosis. The use of other Machine Learning technologies such as Deep Learning can be explored to compare performance and effectiveness in diagnosing pests and diseases of onion plants.

The proposed expert system could contribute to more sustainable and environmentally friendly agricultural practices in several ways. First, with a quick and accurate diagnosis, farmers can take timely and specific control measures, reducing the need for excessive use of pesticides that often damage the environment. Second, the system supports the application of precision agriculture techniques, which maximizes crop yields by minimizing unnecessary inputs, such as water and fertilizers, which contribute to the conservation of natural resources and the reduction of carbon footprint.

In addition to Naïve Bayes, other Machine Learning technologies such as Deep Learning have great potential in the diagnosis of plant diseases. Deep Learning, specifically Convolutional Neural Networks (CNNs), can be used to analyze images of infected plants with a very high degree of accuracy. Deep Learning implementations can provide advantages in recognizing more complex patterns and deeper features in image datasets, which may not be detected by simpler methods such as Naïve Bayes. With these capabilities, Deep Learning-based systems can provide more accurate and detailed diagnoses, assisting farmers in more informative and effective decision-making.

#### REFERENCES

- Blanquero, R., Carrizosa, E., Ramírez-Cobo, P., & Sillero-Denamiel, M. R. (2021). Variable selection for Naïve Bayes classification. *Computers and Operations Research*, 135. <https://doi.org/10.1016/j.cor.2021.105456>
- Bowers, A. J., & Zhou, X. (2019). Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes. *Journal of Education for Students Placed at Risk*, 24(1), 20–46. <https://doi.org/10.1080/10824669.2018.1523734>
- Brodzicki, A., Piekarski, M., & Jaworek-Korjakowska, J. (2021). The whale optimization algorithm approach for deep neural networks. *Sensors*, 21(23). <https://doi.org/10.3390/s21238003>
- Brown, M. E., Mugo, S., Petersen, S., & Klausner, D. (2022). Designing a Pest and Disease Outbreak Warning System for Farmers, Agronomists and Agricultural Input Distributors in East Africa. *Insects*, 13(3). <https://doi.org/10.3390/insects13030232>
- Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. <https://doi.org/10.38094/jastt20165>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1). <https://doi.org/10.1186/s12864-019-6413-7>
- Demilie, W. B. (2024). Plant disease detection and classification techniques: a comparative study of the performances. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-023-00863-9>
- Fitri, I., Teknologi Komunikasi dan Informatika, F., Nasional, U., Manila, S., Minggu, P., Ps Minggu, K., Jakarta Selatan, K., & Khusus Ibu Kota Jakarta, D. (2020). *Expert System for Early Detection of Disease in Corn Plant Using Naive Bayes Method* (Vol. 3, Issue 36). <https://iocscience.org/ejournal/index.php/mantik/index>
- Foerster, A., & Choi, J. (2016). Recession Forecasting Using Bayesian Classification. *The Federal Reserve Bank of Kansas City Research Working Papers*. <https://doi.org/10.18651/rwp2016-06>

- Fritz, S., See, L., Bayas, J. C. L., Waldner, F., Jacques, D., Becker-Reshef, I., Whitcraft, A., Baruth, B., Bonifacio, R., Crutchfield, J., Rembold, F., Rojas, O., Schucknecht, A., Van der Velde, M., Verdin, J., Wu, B., Yan, N., You, L., Gilliams, S., ... McCallum, I. (2019). A comparison of global agricultural monitoring systems and current gaps. *Agricultural Systems*, 168, 258–272. <https://doi.org/10.1016/j.agsy.2018.05.010>
- Hasan, R. I., Yusuf, S. M., & Alzubaidi, L. (2020). Review of the state of the art of deep learning for plant diseases: A broad analysis and discussion. In *Plants* (Vol. 9, Issue 10, pp. 1–25). MDPI AG. <https://doi.org/10.3390/plants9101302>
- Khan, A. I., Quadri, S. M. K., Banday, S., & Shah, J. L. (2022). Deep Diagnosis: A Real-Time Apple Leaf Disease Detection System Based on Deep Learning. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4019467>
- Khotimah, B. K., Miswanto, M., & Suprajitno, H. (2020). Optimization of feature selection using genetic algorithm in naïve Bayes classification for incomplete data. *International Journal of Intelligent Engineering and Systems*, 13(1), 334–343. <https://doi.org/10.22266/ijies2020.0229.31>
- Kotov, R., Jonas, K. G., Carpenter, W. T., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Hobbs, K., Reininghaus, U., Slade, T., South, S. C., Sunderland, M., Waszczuk, M. A., Widiger, T. A., Wright, A. G. C., Zald, D. H., Krueger, R. F., & Watson, D. (2020). Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): I. Psychosis superspectrum. *World Psychiatry*, 19(2), 151–172. <https://doi.org/10.1002/wps.20730>
- Kurniawan, A., & Pading, J. (n.d.). Naive Bayes Method in Determining Diagnosis of Corn Plant Disease. In *Journal of Knowledge Engineering and Artificial Intelligence (JKEAI)* (Vol. 1, Issue 1).
- Lapajne, J., Knapič, M., & Žibrat, U. (2022). Comparison of Selected Dimensionality Reduction Methods for Detection of Root-Knot Nematode Infestations in Potato Tubers Using Hyperspectral Imaging. *Sensors*, 22(1). <https://doi.org/10.3390/s22010367>
- Masri, N., Sultan, Y. A., Akkila, A. N., Almasri, A., Ahmed, A., Mahmoud, A. Y., Zaqout, I., & Abu-Naser, S. S. (2019). Survey of Rule-Based Systems. In *International Journal of Academic Information Systems Research* (Vol. 3). [www.ijeais.org/ijaisr](http://www.ijeais.org/ijaisr)
- Prastiyo, S. E., Irham, Hardyastuti, S., & Jamhari. (2020). How agriculture, manufacture, and urbanization induced carbon emission? The case of Indonesia. *Environmental Science and Pollution Research*, 27(33), 42092–42103. <https://doi.org/10.1007/s11356-020-10148-w>
- Shen, Y., Li, Y., Zheng, H. T., Tang, B., & Yang, M. (2019). Enhancing ontology-driven diagnostic reasoning with a symptom-dependency-aware Naive Bayes classifier. *BMC Bioinformatics*, 20(1). <https://doi.org/10.1186/s12859-019-2924-0>
- Soekhai, V., de Bekker-Grob, E. W., Ellis, A. R., & Vass, C. M. (2019). Discrete Choice Experiments in Health Economics: Past, Present and Future. In *PharmacoEconomics* (Vol. 37, Issue 2, pp. 201–226). Springer International Publishing. <https://doi.org/10.1007/s40273-018-0734-2>
- Song, Y., Wang, J., Ge, Y., & Xu, C. (2020). An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: cases with different types of spatial data. *GIScience and Remote Sensing*, 57(5), 593–610. <https://doi.org/10.1080/15481603.2020.1760434>
- Steentjes, M. B. F., Scholten, O. E., & van Kan, J. A. L. (2021). Peeling the onion: Towards a better understanding of botrytis diseases of onion. *Phytopathology*, 111(3), 464–473. <https://doi.org/10.1094/PHYTO-06-20-0258-IA>
- Tehseen, D., Jilani, A., Sherani, J., Saleem, H., Jilani, M. S., Rashid Khan, A., Jilani, T. A., Saddozai, U. K., Anjum, M. N., Waseem, K., & Ullah, S. (2021). A Comprehensive Review on the Application of Diagnostic Expert Systems in the Field of Agriculture. *International Journal on Emerging Technologies*, 12(1), 304–316. <https://www.researchgate.net/publication/362067712>
- Wang, R., Liu, L., Xie, C., Yang, P., Li, R., & Zhou, M. (2021). Agripest: A large-scale domain-specific benchmark dataset for practical agricultural pest detection in the wild. *Sensors*, 21(5), 1–15. <https://doi.org/10.3390/s21051601>
- Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing*, 25(3), 2277–2293. <https://doi.org/10.1007/s00500-020-05297-6>
- Yağ, İ., & Altan, A. (2022). Artificial Intelligence-Based Robust Hybrid Algorithm Design and Implementation for Real-Time Detection of Plant Diseases in Agricultural Environments. *Biology*, 11(12). <https://doi.org/10.3390/biology11121732>
- Yang, L., & Shami, A. (2020). *On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice*. <https://doi.org/10.1016/j.neucom.2020.07.061>
- Zaki, M. A., Narejo, S., Ahsan, M., Zai, S., Anjum, M. R., & Din, N. U. (2021). Image-based Onion Disease (Purple Blotch) Detection using Deep Convolutional Neural Network. *International Journal of Advanced Computer Science and Applications*, 12(5), 448–458. <https://doi.org/10.14569/IJACSA.2021.0120556>