Jurnal Studi Multidisiplin Ilmu
*ISSN: 3026-7412, Vol 3, No 1, 2025, 45-57* https://doi.org/10.35912/jasmi.v3i1.6937

# K-Nearest Neighbors Based Matic Motorcycle Damage Prediction System Web Application Preventive Maintenance Bengkel Sahabat Motor

**Anggi Wijaya[1*], Sulyono Sulyono[2]**
Institut Informatika dan Bisnis Darmajaya, Bandar Lampung, Indonesia[1,2]
anggiwijaya500@gmail.com[1*], sulyono@darmajaya.ac.id[2]

**Article History**
Received on 1 November 2024
1st Revision on 16 November 2024
2nd Revision on 1 December 2024
3rd Revision on 20 December 2024
Accepted on 9 January 2025

**Abstract**

**Purpose**: This study develops a web-based matic motorcycle damage prediction system using the K-Nearest Neighbors (KNN) algorithm at Bengkel Sahabat Motor to support early damage detection, preventive maintenance, and cost reduction.

**Methodology**: A quantitative approach with waterfall System Development Life Cycle (SDLC) was used. Data were collected through observation, interviews, and workshop records. The system was built using Personal Home Page (PHP), html, Cascading Style Sheets (CSS), JavaScript, and MySQL. KNN with Euclidean distance and K=3 was applied, using a three-level symptom scale. System design used Unified Modeling Language (UML) and validation was conducted through black box testing.

**Results**: The system accurately classifies motorcycle damage, with test outputs correctly identifying "Engine Overheating" based on nearest neighbor distances. Black box testing achieved 100% acceptance across 143 test items, categorized as “Very Good.” Diagnosis time decreased from 30 to 10 minutes per case.

**Conclusions**: The KNN-based system effectively automates motorcycle damage classification and improves diagnostic efficiency.

**Limitations**: The study is limited to a single workshop, small dataset, no IoT integration, and lacks formal accuracy metrics.

**Contributions**: This study provides a practical machine learning-based predictive maintenance system for motorcycle workshops, offering a replicable framework for digital diagnostics in the automotive service sector.

**Keywords:** *Euclidean Distance, K-Nearest Neighbors Algorithm, Motorcycle Damage Prediction System, Preventive Maintenance, Web-Based Decision Support System*

**How to Cite:** Wijaya, A., & Sulyono, S. (2025). K-Nearest Neighbors Based Matic Motorcycle Damage Prediction System Web Application Preventive Maintenance Bengkel Sahabat Motor. *Jurnal Studi Multidisiplin Ilmu*, 3(1), 45-57.

## 1. Introduction

The Matic (automatic transmission) motorcycle has become one of Indonesia most ubiquitous transportation modes, favored for its ease-of-use relative to manual transmission alternatives. Unlike manual motorcycles, matic transmissions use a Continuously Variable Transmission (CVT) system a belt-and-pulley mechanism that automatically adjusts gear ratios without rider intervention making the vehicle more accessible to a broader demographic but also introducing a set of specific mechanical vulnerabilities that differ substantially from manual variants (Kristianto & Supatman, 2024). The CVT system's complexity, combined with the generally limited technical knowledge of matic motorcycle users, creates a persistent problem: damage that could be detected and addressed early through preventive maintenance instead progresses to serious mechanical failure requiring expensive repair, because users neither recognize nor act on early warning symptoms (Epriliani, Wijaya, & Kusuma, 2022).

Automotive repair workshops (bengkel) are the primary institutional interface between motorcycle users and technical expertise. However, the diagnostic capability of workshop mechanics varies considerably with only a subset of technicians possessing the deep mechanical knowledge needed to accurately diagnose complex matic damage patterns from symptom presentations (Alfrido & Gautama, 2022). This diagnostic inconsistency, combined with the high volume of vehicles processed in busy workshops, creates conditions in which damage may be misidentified, under identified, or identified too late for cost-effective preventive repair. A systematic, algorithmic approach to symptom-based damage classification would address both the consistency and the speed dimensions of the diagnostic challenge.

K-Nearest Neighbors (KNN) a supervised machine learning algorithm that classifies new instances by identifying their closest matches in a historical training dataset using distance metrics is particularly well-suited to the motorcycle damage prediction problem. KNN instance-based, non-parametric learning approach does not assume a specific underlying data distribution, making it adaptable to the heterogeneous symptom patterns characteristic of matic motorcycle damage (Priyandoko, 2022; Riyadi, Amin, & Hakim, 2021). Critically, KNN classification logic is directly interpretable: the system identifies which historical cases most closely resemble the current symptom presentation, making the prediction rationale transparent to both mechanics and customers. This interpretability distinguishes KNN from black-box alternatives and supports user trust in the system's outputs.

Recent advances in machine learning–based predictive maintenance and fault diagnosis have significantly improved the reliability of mechanical systems, including automotive transmission systems. Studies have demonstrated that hybrid and data-driven models such as K-Nearest Neighbors, Support Vector Machines, and deep learning approaches can effectively classify mechanical faults using vibration and sensor data, particularly in continuously variable transmission systems (Babu, Senthilnathan, Pancholi, Kumar, Prabha, Mohammed, & Wahab, 2021; Vrba, Cejnek, Steinbach, & Krbcova, 2021). Furthermore, recent research highlights that combining feature extraction techniques with supervised learning algorithms improves diagnostic accuracy and robustness under varying operational conditions (Almounajjed, & Sahoo, 2022; Zhao, Yao, Deng, Ding, Ding, Jia, & Liu, 2023). In automotive applications, intelligent diagnostic systems have been widely adopted to enhance early fault detection, reduce downtime, and support preventive maintenance strategies, especially in systems with complex mechanical structures such as CVT-based transmissions (Narendiranath, 2021; Wang, Yu, & Wu, 2022).

In addition, web-based decision support systems integrated with machine learning algorithms have become an emerging trend in intelligent maintenance systems, enabling real-time and user-friendly diagnostic platforms for end-users and technicians (Rusnanda & Haliza, 2021). Recent studies also emphasize the importance of interpretable and lightweight models for industrial deployment, where K-Nearest Neighbors remains relevant due to its simplicity and transparency in classification logic (Almounajjed & Sahoo, 2022). Moreover, predictive maintenance frameworks increasingly incorporate multi-source data fusion and adaptive learning mechanisms to handle dynamic operational environments and improve classification reliability (Chaleshtori, & Aghaie, 2022; Tan, Ning, Peng, Xia, & Wu, 2022). These developments indicate a strong shift toward intelligent, accessible, and data-driven diagnostic systems in automotive maintenance contexts, supporting the relevance of applying KNN-based web systems for matic motorcycle damage prediction in real workshop environments.

## 2. Literature Review

### *2.1 Matic Motorcycle Damage Diagnosis: Current Approaches*

Several machine learning and expert system approaches have been applied to motorcycle damage diagnosis in prior research. Kristianto and Supatman (2024) applied the Naive Bayes algorithm for matic motorcycle damage diagnosis, identifying that the primary challenge is users' lack of technical knowledge in identifying symptom-cause relationships. Epriliani et al. (2022) implemented Naive Bayes for motorcycle damage prediction at Bengkel Citra Djaya Motor, exploiting the algorithm's probabilistic classification strengths. Nuryahya and Muflihah (2023) applied the Certainty Factor method for carburetor motorcycle diagnosis via web mobile, providing confidence-weighted diagnosis outputs. Mulyani and Natsir (2023) applied forward chaining logic for motorcycle damage diagnosis at Bengkel Rahmat Cort, using rule-based reasoning to trace from symptoms to damage conclusions. While all four approaches have demonstrated effectiveness in their respective deployment contexts, each involves limitations that KNN's instance-based approach is designed to address specifically: Naive Bayes assumes conditional independence among symptoms, which may not hold in practice; Certainty Factor requires expert-formulated rules; and forward chaining is inflexible when new damage patterns emerge outside the rule base. KNN avoids all three limitations by learning directly from historical case data without independence assumptions or rule formulation (Ardhani, 2020).

### *2.2 K-Nearest Neighbors Algorithm*

KNN is a supervised learning algorithm that classifies a new instance by identifying its K nearest neighbors in the training dataset using a distance metric, then assigning the majority class among those neighbors as the prediction (Riyadi, Amin, & Hakim, 2021). The algorithm is non-parametric making no assumptions about the underlying data distribution which gives it flexibility for datasets with complex, non-linear class boundaries characteristic of symptom-to-damage mappings. KNN's primary hyperparameter is K (the number of nearest neighbors): small K values increase sensitivity to noise (risk of overfitting), while large K values reduce sensitivity to local patterns (risk of blurring decision boundaries) (Ayu, Sari, & Rahman, 2023). Priyandoko (2022) confirms that KNN produces flexible, non-linear decision boundaries that are particularly well-suited to heterogeneous classification problems such as multi-symptom vehicle damage identification. The standard Euclidean distance metric for two data points A (training) and B (testing) across n feature dimensions is:

$$d(A, B) = \sqrt{[\Sigma_{i=1}{}^{n}(A_i - B_i)^2]} \quad (1)$$

Formula 1 where $A_i$ is the value of the i-th attribute in training data point A, and $B_i$ is the corresponding attribute in testing data point B. The K data points with smallest d (A,B) values are selected as nearest neighbours, and the majority class label among these K neighbours is assigned to B (Riyadi et al., 2021; Prayoga, 2023).

### *2.3 System Development: Waterfall and Black Box Testing*

The Waterfall SDLC methodology provides a sequential, phase-gated development framework encompassing requirements gathering, system analysis, system design, implementation, testing, and maintenance each phase completing before the next begins (Maulana et al., 2021; Tan & Hajjah, 2020). Waterfall is appropriate for systems with well-defined, stable requirements, as in the current study where functional specifications (prediction, CRUD management, reporting) are fully determinable from the research design stage. Black box testing validates system functionality by verifying input-output behaviour against specification without requiring knowledge of internal implementation the appropriate validation method for user-facing web systems where end-user experience of correct functional behaviour is the primary quality criterion (Mulki, Andrianto, & Permana, 2023). Evaluation uses a Likert scale with five qualification tiers: 90–100% = Very Good (no revision); 75–89% = Good (minor revision); 65–74% = Adequate (moderate revision); 55–64% = Insufficient (major revision); 0–54% = Very Poor (rebuild).

*2.4 Prior Empirical Studies*

Table 1. Summary of prior studies on KNN and machine learning for vehicle damage prediction

| Author(s) & Year | Application | Method | Key Finding Relevant to KNN-Based Vehicle Damage Prediction |
|---|---|---|---|
| Kristianto & Supatman (2024) | Matic motorcycle damage diagnosis | Naive Bayes | Naive Bayes effectively diagnoses matic motorcycle damage; main limitation is the system's probabilistic dependency on training data distribution; KNN instance-based approach may complement this for small datasets |
| Epriliani et al. (2022) | Motorcycle damage prediction, Citra Djaya Motor | Naive Bayes | Naive Bayes motorcycle damage prediction achieves satisfactory accuracy; simplicity and probabilistic classification are strengths; does not capture local feature similarity patterns that KNN provides |
| Nuryahya & Muflihah (2023) | Motorcycle carburettor damage diagnosis, web mobile | Certainty Factor | Certainty factor provides explainable confidence levels in diagnosis; limitation is rule-base dependency; KNN data-driven approach avoids the need for expert rule formulation |
| Mulyani & Natsir (2023) | Motorcycle damage diagnosis, Bengkel Rahmat Cort | Forward Chaining | Forward chaining expert system provides structured reasoning path from symptom to diagnosis; effective but inflexible when new damage patterns emerge outside rule base |
| Maulana et al. (2021) | KNN in machine learning | KNN | KNN demonstrates strong classification performance for pattern recognition tasks; its non-parametric nature makes it adaptable to diverse data distributions including automotive symptom datasets |
| Riyadi et al. (2021) | KNN for text and data classification | KNN, Euclidean distance | KNN with Euclidean distance metric achieves reliable classification across diverse data types; key hyperparameter is K selection — small K risks overfitting, large K blurs decision boundaries |
| Prayoga (2023) | KNN for historical data prediction | KNN | KNN applied to historical pattern-matching prediction yields high accuracy when training data is representative; performance depends critically on dataset quality and Euclidean distance metric scaling |
| Present Study (2024) | Matic motorcycle damage prediction, Bengkel Sahabat Motor (K=3, n=15 damage categories) | KNN, Euclidean distance, PHP/Laravel/MySQL, Waterfall, Black Box | KNN with Euclidean distance and 3-level symptom weighting (1=mild, 2=moderate, 3=severe) predicts damage category with illustrative example: nearest neighbour = "Engine Overheating" (d=2.8284); diagnosis time reduced from 30 to 10 minutes; 100% black box test pass rate (4 respondents) |

Table 1 summarizes prior studies on machine learning approaches for vehicle and motorcycle damage prediction, particularly focusing on the use of K-Nearest Neighbors and other classification methods. The studies compare different algorithms such as Naive Bayes, certainty factor, forward chaining, and KNN, highlighting their strengths and limitations in diagnostic accuracy and interpretability. Overall, the table shows that KNN is widely recognized for its simplicity and effectiveness in pattern-based classification, especially when combined with Euclidean distance for similarity measurement. It also indicates that hybrid and rule-based systems are useful but often lack flexibility compared to data-driven approaches like KNN. The present study builds on these findings by applying KNN in a web-based system for matic motorcycle damage prediction with improved practical implementation in a workshop setting.

## 3. Research Methodology

### *3.1 Research Design and Setting*

A quantitative experimental systems development approach employing the Waterfall SDLC was adopted. The research was conducted at Bengkel Motor Sahabat Motor, a motorcycle repair workshop in Bandar Lampung, Lampung Province, Indonesia. Data collection spanned the problem identification through requirements analysis phases, with system development and testing conducted in the subsequent phases.

In addition, the system development process incorporated a structured iterative validation mechanism to ensure that each stage of the Waterfall SDLC was properly verified before proceeding to the next phase. Requirement analysis was conducted through direct observation of motorcycle diagnostic workflows and structured interviews with workshop mechanics to identify key symptoms and classification patterns of matic motorcycle damage. The design phase translated these requirements into system architecture using UML diagrams, while the implementation phase operationalized the K-Nearest Neighbors algorithm within a web-based environment using PHP, MySQL, and related front-end technologies. Testing was performed using black box methods involving mechanics and administrative users to evaluate functional correctness, usability, and system reliability. This structured development approach ensured that the resulting system was not only technically valid but also aligned with real-world operational needs of Bengkel Motor Sahabat Motor, particularly in improving diagnostic consistency and reducing service time.

### *3.2 Data Collection Methods*

Four complementary data collection methods were employed. Direct observation at Bengkel Sahabat Motor documented current damage diagnosis workflows, average diagnosis times, and common matic damage symptom patterns presented by visiting vehicles. Structured interviews with workshop mechanics provided expert knowledge on symptom-to-damage mappings, the most frequently encountered matic damage categories, and the features of an effective prediction interface from the mechanic's operational perspective. Historical damage records (*dokumentasi*) provided the training dataset of previously diagnosed damage cases with associated symptom profiles. Literature review of academic journals on KNN implementation, motorcycle damage diagnosis systems, and web application development informed the system design choices and algorithm parameterization (Fauzy, Hidayat, & Pratama, 2020; Nurhidayati, Prasetyo, & Utomo, 2024).

### *3.3 Training Data: Symptom Weighting*

This three-level weighting converts qualitative symptom severity observations into numerical feature vectors compatible with the Euclidean distance metric. The weighting enables the KNN algorithm to distinguish between different degrees of the same symptom for example, mild engine vibration versus severe engine vibration rather than treating all occurrences of a symptom as equivalent.

Table 2. Symptom severity weighting scale for KNN training data

| Weight Value | Severity Level | Operational Definition |
|---|---|---|
| 1 | Ringan (Mild) | Minor damage not affecting core motor performance; rider can continue use but maintenance is recommended |

| Weight Value | Severity Level | Operational Definition |
|---|---|---|
| 2 | Sedang (Moderate) | Damage beginning to impair motor performance; noticeable symptoms affecting ride quality; prompt attention required |
| 3 | Berat (Severe) | Damage causing the motor to malfunction significantly; motor cannot be operated safely; immediate repair required |

Table 2 presents the symptom severity weighting scale used in the K-Nearest Neighbors training process. It defines three levels of severity mild, moderate, and severe each assigned a numerical weight from 1 to 3. The scale operationalizes motorcycle damage symptoms by indicating their impact on engine performance, ranging from minor issues that allow continued use to severe conditions requiring immediate repair. This weighting system enables structured input representation for more accurate classification in the prediction model.

Fifteen damage categories were identified from historical records and mechanic interviews. Table 3 presents the ten primary categories with their associated symptoms and repair recommendations.

Table 3. Primary Matic motorcycle damage categories, symptoms, and repair recommendations

| No. | Damage Category | Primary Symptoms | Recommended Action |
|---|---|---|---|
| 1 | Engine Overheating (Mesin Cepat Panas) | Engine temperature abnormally high; reduced performance in prolonged use | Change oil; inspect radiator fan; add coolant |
| 2 | ECU Error / Total Failure | Motor will not start; warning indicators illuminate abnormally | Diagnose ECU; replace if faulty |
| 3 | Fuel Overconsumption | Fuel consumption exceeds normal range for operating conditions | Service CVT; inspect injector; check air-fuel ratio |
| 4 | Engine Stalling | Engine stops unexpectedly during operation | Check fuel tank; inspect ignition system and spark plug |
| 5 | Engine Misfiring (Mesin Brebet) | Irregular engine firing; hesitation during acceleration | Clean injector; adjust air-fuel mixture |
| 6 | Idle Vibration | Abnormal vibration at idle speed | Replace CVT chain; check engine mounts |
| 7 | CVT Noise | Grinding or rattling sound from CVT housing | Replace worn rollers or clutch pads |
| 8 | Power Loss on Incline | Motor unable to maintain speed when climbing | Replace V-belt and CVT rollers |
| 9 | Heavy Engine Pull | Engine feels sluggish; acceleration response delayed | Clean air filter; service CVT |
| 10 | Black Exhaust Smoke | Excessive dark smoke from exhaust pipe | Replace spark plug; adjust fuel mixture |

Table 3 outlines the primary matic motorcycle damage categories, their associated symptoms, and recommended repair actions. It includes ten common fault types such as engine overheating, ECU failure, fuel overconsumption, and CVT noise. Each category is linked with observable symptoms and corresponding maintenance recommendations. This structured mapping supports diagnostic consistency and serves as the knowledge base for the K-Nearest Neighbors classification system, enabling accurate and interpretable damage prediction.

### *3.4 KNN Algorithm Implementation: Illustrative Calculation*

Table 4. KNN distance calculation: Top-3 nearest neighbours for example test case (K = 3)

| **Rank** | **Training Data (Damage Category)** | **Euclidean Distance d** | **$d^2 = \Sigma(x_i - y_i)^2$** | **Classification Contribution** |
|---|---|---|---|---|
| 1st | Engine Overheating (Mesin Cepat Panas) | d = 2.8284 | $\sqrt{8}$ = 2.8284 | Selected (K=3, rank 1) |
| 2nd | ECU Error / Total Failure | d = 3.0000 | $\sqrt{9}$ = 3.0000 | Selected (K=3, rank 2) |
| 3rd | Fuel Overconsumption | d = 3.1622 | $\sqrt{10}$ = 3.1622 | Selected (K=3, rank 3) |
| Prediction | Majority class among K=3 nearest neighbours: Engine Overheating (Mesin Cepat Panas) → System output: "*Ganti oli, periksa kipas radiator, tambahkan pendingin*" | | | |

Table 4 show the KNN prediction process for a test input vector is demonstrated through the following illustrative example. For a vehicle presenting symptoms coded as a weighted vector $y = [y_1, y_2, ..., y_n]$, the Euclidean distance to each training case $x_i$ is computed, and the K=3 training cases with smallest distances are selected as nearest neighbours.

### *3.5 Technology Stack*

The system is developed with the following technology components: PHP (backend server-side processing and KNN calculation engine); HTML, CSS, and JavaScript (frontend user interface and interactive symptom input forms); MySQL (relational database for training data, customer records, service history, and spare parts catalogue); Apache/XAMPP (local development web server environment); and deployed on cloud hosting at https://bengkelmotor.my.id/. System design uses UML including Use Case Diagram (defining interactions among Admin, Mechanic, and Manager roles) and Activity Diagrams (documenting the login workflow and the KNN prediction process workflow from symptom input to diagnosis output).

## 4. Results and Discussions

### *4.1 System Implementation*

The KNN-based matic motorcycle damage prediction system was successfully implemented and deployed at https://bengkelmotor.my.id/. The system integrates the KNN prediction engine with a comprehensive workshop management platform encompassing customer management, vehicle records, mechanic administration, spare parts catalogue, service recording, and management reporting all within a single web-based interface accessible to authorized users through role-based authentication.

Recent studies have shown that the integration of K-Nearest Neighbors in predictive maintenance and fault diagnosis systems significantly improves classification accuracy and operational efficiency in transportation and mechanical systems. The simplicity and instance-based learning nature of the algorithm make it suitable for real-time diagnostic environments, particularly in motorcycle and automotive applications where symptom variability is high. Prior research also indicates that KNN performs effectively when combined with structured feature weighting and distance-based similarity measures, enabling robust classification even in small to medium-sized datasets (Irsyad, 2021; Sahoo et al., 2021). Furthermore, machine learning-based diagnostic systems deployed in web environments have been shown to enhance accessibility for end users, allowing non-experts to obtain reliable maintenance recommendations without requiring deep mechanical expertise (Rusnanda & Haliza, 2021; Vrba et al., 2021).

From a system engineering perspective, the adoption of web-based decision support systems for vehicle maintenance aligns with the broader trend of Industry 4.0 digitalization in workshop management. Studies highlight that integrating machine learning models into web applications improves workflow efficiency, reduces diagnostic time, and enhances service consistency across technicians (Adhiatma et al., 2025; Suhadi et al., 2022). In addition, predictive maintenance frameworks that combine historical data analysis with real-time decision support mechanisms are increasingly recognized as essential tools for reducing operational costs and preventing catastrophic failures in mechanical systems (Taoufyq & Mansouri, 2024; Wang et al., 2022). These findings reinforce the relevance of implementing a K-Nearest Neighbors-based web system in Bengkel Sahabat Motor as a practical solution for standardized and efficient motorcycle damage diagnosis.

*4.1.1 Pre-Login Interface and Authentication*

The system's public homepage presents workshop information including service description, workshop mission, and system capabilities — to potential users before login. This information-first design aims to establish user confidence in the system's value proposition before requiring authentication. The login module implements credential-based authentication with role differentiation, redirecting authenticated users to role-appropriate dashboards (Admin, Mechanic, or Manager) with corresponding function access.

The system's user interface is designed with a clear separation between public and authenticated functionalities to enhance usability and user engagement. The public homepage provides essential information about the workshop, including service descriptions, organizational mission, and system capabilities, allowing potential users to understand the value of the platform before engaging further. This information-first approach is intended to build trust and increase user acceptance by presenting transparent and accessible service details. Once users proceed to authentication, the login module enforces credential-based access control and implements role differentiation to ensure secure and structured system usage. Depending on their assigned roles, Admin, Mechanic, or Manager users are automatically redirected to their respective dashboards, each offering tailored functionalities aligned with their operational responsibilities within the workshop management system (Mugiprakoso, Santoso, & Kurniawan, 2020).

*4.1.2 KNN Damage Prediction Module*

The prediction module is the system's functional centerpiece. The prediction workflow comprises three sequential screens. In the first screen, the mechanic enters vehicle information (brand, model, year, mileage) to contextualize the prediction. In the second screen, the mechanic selects observed symptoms from a comprehensive checklist and assigns severity ratings (Mild=1, Moderate=2, Severe=3), creating the input vector y for KNN computation. The system then calculates Euclidean distances between the input vector and all training data vectors, ranks the distances, and selects K=3 nearest neighbors. The third screen displays the prediction output: the majority-voted damage category, an indication of prediction confidence based on neighbor distance distribution, and the associated repair recommendation specific to the predicted damage category (Rahman, Setiawan, & Mahendra, 2024).

The prediction module represents the core functional component of the system, designed to operationalize the K-Nearest Neighbors algorithm into a structured, user-friendly diagnostic workflow. The module is implemented as a multi-stage process consisting of three sequential interfaces that guide mechanics through data input, symptom encoding, and diagnostic output generation. In the first stage, users input vehicle attributes such as brand, model, manufacturing year, and mileage. These attributes provide contextual grounding that may influence symptom interpretation and ensure that the prediction process is aligned with vehicle-specific conditions commonly encountered in real workshop environments (Rosa & Shalahuddin, 2020).

In the second stage, the system captures observed symptoms through a structured checklist interface, where mechanics select relevant indicators of motorcycle damage and assign severity levels using a standardized scale (Mild = 1, Moderate = 2, Severe = 3). This process transforms qualitative mechanical observations into a quantitative input vector y, enabling computational processing within the K-Nearest Neighbors framework. The system then computes Euclidean distances between the input vector and all stored training instances, measuring similarity in a multidimensional symptom space. Subsequently, all distances are ranked in ascending order, and the K = 3 nearest neighbors are selected as the reference set for classification. This stage is critical because it determines the quality of neighborhood representation, which directly affects prediction accuracy and system robustness, particularly in cases where symptom overlap exists across multiple damage categories (Septian, 2020).

In the third stage, the system generates the diagnostic output based on majority voting among the selected neighbors, producing the final predicted damage category. To enhance interpretability, the system also computes a confidence indicator derived from the distribution and proximity of the nearest neighbors, allowing users to assess the reliability of the prediction. Additionally, the module provides a structured repair recommendation aligned with the predicted damage type, ensuring that the output is not only descriptive but also actionable for workshop decision-making. This integrated design strengthens the system's role as a decision support tool by combining algorithmic classification with practical maintenance guidance, thereby reducing diagnostic ambiguity and improving consistency in repair recommendations across different mechanics.

#### *4.1.3 Service Recording and Data Management*

The service recording module enables mechanics to document completed repair transactions including spare parts used, repair costs, and assigned mechanic and generates printable service receipts for customers. This documentation serves two functions: it provides the customer with transparent service cost documentation, and it enriches the system's training dataset with new resolved damage cases, progressively improving the KNN model's coverage as operational data accumulates. The spare parts catalogue management module provides structured inventory tracking with categorization by component type (engine, CVT, electrical, brake, body), enabling both service transaction support and demand forecasting based on prediction trend analysis.

### *4.2 Black Box Testing Results*

All 143 test items across all four respondents were accepted without rejection, achieving a 100% acceptance rate and placing the system in the "Sangat Baik / Very Good" Likert qualification category (90–100%), which specifies "No Revision Required." The consistent 100% acceptance across the three mechanic respondents who are the system's primary operational users is particularly significant, confirming that the prediction workflow, symptom input interface, and output display are all intuitively usable by non-technical mechanical staff without extensive training.

### *4.3 Operational Impact and Discussion*

The most directly measurable operational impact of the system is the reduction in average diagnosis time from 30 minutes to 10 minutes per vehicle a 67% reduction that translates directly into increased workshop throughput, reduced per-vehicle labor cost, and faster customer service delivery. This time reduction arises from two mechanisms: first, the KNN prediction eliminates the iterative, memory-based diagnostic process that experienced mechanics undertake manually; second, the structured symptom input checklist provides a guided diagnostic protocol that ensures all relevant symptoms are systematically assessed rather than selectively noted.

The KNN algorithm's suitability for this application is confirmed by its operational performance in the system. The illustrative calculation demonstrates that the Euclidean distance metric effectively discriminates between damage categories with the nearest neighbors ($d = 2.8284$) clearly separated from the second nearest ($d = 3.0000$) indicating that the symptom weighting scheme creates a sufficiently differentiated feature space for reliable classification. However, it is important to note that the current study does not report formal accuracy metrics (precision, recall, F1-score) from a held-out test set evaluation, which would be necessary to formally quantify the system's classification accuracy across all 15 damage categories. Future research should address this limitation through systematic cross-validation evaluation.

Compared with prior approaches in the literature by Kristianto and Supatman (2024) and Epriliani, Wijaya, and Kusuma (2022), certainty factor by Nuryahya and Muflihah (2023), and forward chaining by Mulyani and Natsir (2023), the KNN approach implemented here offers two practical advantages for the workshop context. First, knowledge acquisition is automatic: the system learns from historical damage records without requiring expert formulation of classification rules or conditional probability distributions, enabling any workshop to populate and update the system without expert consultation.

Second, new damage patterns are naturally incorporated by adding new training cases to the database, whereas rule-based and probabilistic systems require explicit rule updates or distribution recalibration when new damage types emerge.

## 5. Conclusions

### *5.1 Conclusion*

This study successfully designed, developed, and implemented a web-based KNN damage prediction system for matic motorcycles at Bengkel Sahabat Motor, deployed at https://bengkelmotor.my.id/. The system applies the K-Nearest Neighbors algorithm with Euclidean distance metric and three-level severity weighting (K=3) to classify incoming symptom vectors against a historical training dataset of 15 matic motorcycle damage categories, producing damage type predictions with associated repair recommendations. Black box testing achieved 100% acceptance across all 143 test items from four respondents (one admin, three mechanics), qualifying the system as "Sangat Baik / Very Good" with no revision required. The system reduced average diagnosis time from 30 minutes to 10 minutes per vehicle (67% reduction), improving workshop throughput, customer service speed, and diagnostic consistency. The web-based architecture supports role-differentiated access for Admin, Mechanic, and Manager users, with integrated customer management, vehicle records, spare parts catalogue, service recording, and reporting functions that collectively digitalise the workshop's complete operational workflow.

### *5.2 Research Limitations*

Four limitations apply to this study. First, the research and training dataset are limited to a single workshop (Bengkel Sahabat Motor); the KNN model's performance on damage patterns specific to other workshop contexts or geographic regions would require validation with locally collected training data. Second, the study does not report formal accuracy metrics (precision, recall, F1-score, confusion matrix) from a held-out validation test set the 100% black box test success rate validates functional correctness of the software but does not establish the statistical accuracy of the KNN predictions across damage categories. Third, the historical training dataset is described as "relatively small" with a small training set, the KNN model may produce inaccurate predictions for damage categories underrepresented in the historical records, a limitation that will diminish as operational data accumulates. Fourth, the study does not include IoT sensor integration for real-time automated symptom detection; current symptom input is manually entered by mechanics, introducing subjectivity in severity rating assignment.

### *5.3 Suggestions and Directions for Future Research*

Four research directions are recommended. First, a formal accuracy evaluation using k-fold cross-validation should be conducted on the full training dataset to compute per-class precision, recall, and F1-score metrics, establishing a quantitative accuracy baseline and identifying damage categories where KNN classification is weakest. This would enable direct comparison with Naive Bayes, Certainty Factor, and Forward Chaining approaches applied to the same dataset. Second, comparison with alternative K values (K=1, K=5, K=7) should be empirically evaluated to determine whether K=3 is the optimal hyperparameter for this specific training dataset, as K optimization is a critical determinant of KNN performance. Third, IoT sensor integration incorporating real-time RPM monitoring, vibration sensors, and temperature sensors would enable automated symptom vector generation without mechanic manual input, reducing subjectivity and potentially increasing both prediction speed and accuracy. Fourth, multi-workshop deployment and dataset aggregation would substantially expand the training dataset, improve KNN coverage of rare damage types and enabling comparative accuracy analysis across different motorcycle brands, models, and usage patterns characteristic of the broader Bandar Lampung automotive market.


## Acknowledgement

The authors express sincere gratitude to Bengkel Motor Sahabat Motor particularly for their generous provision of operational data, their participation as system testing respondents, and their practical insights into matic motorcycle damage diagnosis that shaped the system's design. The authors also thank the Institut Informatika dan Bisnis Darmajaya, Bandar Lampung, for institutional support and research facilities.

## Author Contributions

AW contributed to the conceptualization of the study, system design, data collection, implementation of the K-Nearest Neighbors algorithm, and preparation of the original manuscript draft. SS contributed to methodological validation, supervision of the research process, critical review and editing of the manuscript, and provided overall academic guidance to ensure the quality and coherence of the study. Both authors have read and approved the final manuscript and agree to be accountable for all aspects of the work.

## References

Adhiatma, N., Ikhsan, M., Purnama, F., & Magfira, F. (2025). Implementation of web-based motorcycle workshop management application to improve efficiency and service quality. *SIMANTEC Journal (verify journal DOI if needed)*. https://doi.org/10.XXX/xxxxx

Alfrido, & Gautama. (2022). Analisis kualitas teknisi bengkel dalam diagnosis kerusakan kendaraan bermotor (Analysis of workshop technician quality in vehicle damage diagnosis). *Jurnal Teknik Mesin*, *15*(2), 45-52..

Almounajjed, A., & Sahoo, A. K. (2022). Wavelet-based multi-class support vector machine for stator fault diagnosis in induction motor. *Measurement and Control*, *45*(2), 261-273. https://doi.org/10.1177/014233122221109725

Ardhani, M. (2020). *Pengantar teknologi database dan sistem informasi (Introduction to database technology and information systems)*.

Ayu, D., Sari, P., & Rahman, A. (2023). Peran teknologi informasi dalam mendukung deteksi kerusakan kendaraan bermotor (The role of information technology in supporting vehicle damage detection). *Jurnal Sistem Informasi*, *12*(3), 112-125..

Babu, T. N., Senthilnathan, N., Pancholi, S., Kumar, S. P. N., Prabha, D. R., Mohammed, N., & Wahab, R. S. (2021). Fault analysis on continuous variable transmission using DB-06 wavelet decomposition and ANN classification. *Journal of Intelligent & Fuzzy Systems*, *41*(1), 1297-1307. https://doi.org/10.3233/JIFS-210199

Chaleshtori, A. E., & Aghaie, A. (2022). Data fusion techniques for fault diagnosis of industrial machines: A survey. *arXiv preprint*. Retrieved from https://arxiv.org/abs/2211.09551

Epriliani, R., Wijaya, S., & Kusuma, T. (2022). Implementasi algoritma Naive Bayes untuk memprediksi kerusakan sepeda motor pada Bengkel Citra Djaya Motor (Implementation of Naive Bayes algorithm for motorcycle damage prediction at Citra Djaya Motor Workshop). *Jurnal Teknik Informatika*, *8*(4), 234-245..

Fauzy, M., Hidayat, R., & Pratama, D. (2020). Sistem informasi manajemen bengkel kendaraan bermotor (Vehicle workshop management information system). *Jurnal Rekayasa Sistem*, *7*(2), 78-89..

Irsyad, H. H. (2021). Comparative performance of K-Nearest Neighbors and SVM algorithms for classification. *Systec Journal (journal DOI not always assigned in repository source)*. https://doi.org/10.XXX/xxxxx

Kristianto, B., & Supatman, A. (2024). Sistem pakar diagnosa kerusakan kendaraan bermotor jenis matic menggunakan metode Naive Bayes (Expert system for matic vehicle damage diagnosis using Naive Bayes). *Jurnal Artificial Intelligence*, *11*(1), 23-35..

Maulana, I., Saputra, H., & Wibowo, A. (2021). Implementasi algoritma K-Nearest Neighbor dalam machine learning (Implementation of K-Nearest Neighbor algorithm in machine learning). *Jurnal Data Mining*, *9*(3), 156-167..

Mugiprakoso, T., Santoso, B., & Kurniawan, D. (2020). Perkembangan sepeda motor sebagai alat transportasi masyarakat Indonesia (Development of motorcycles as Indonesian community transportation). *Jurnal Transportasi*, *13*(2), 89-102..

Mulki, F., Andrianto, S., & Permana, I. (2023). Pengujian sistem menggunakan metode Black Box Testing (System testing using Black Box Testing method). *Jurnal Software Engineering*, *10*(4), 201-212..

Mulyani, S., & Natsir, M. (2023). Sistem pakar diagnosis kerusakan sepeda motor di Bengkel Rahmat Cort menggunakan metode forward chaining (Expert system for motorcycle damage diagnosis at Bengkel Rahmat Cort using forward chaining). *Jurnal Expert System*, *14*(2), 67-78..

Narendiranath Babu, T., et al. (2021). Fault diagnosis in CVT systems using wavelet-based ANN methods. *Journal of Intelligent & Fuzzy Systems*. https://doi.org/10.3233/JIFS-210199

Nurhidayati, L., Prasetyo, A., & Utomo, P. (2024). Definisi dan klasifikasi kendaraan bermotor roda dua (Definition and classification of two-wheeled motor vehicles). *Jurnal Otomotif*, *16*(1), 45-56..

Nuryahya, D., & Muflihah, N. (2023). Sistem pakar diagnosa kerusakan sepeda motor bebek karburator dengan metode certainty factor berbasis web mobile (Expert system for carburettor motorcycle damage diagnosis using certainty factor, web mobile based). *Jurnal Mobile Computing*, *12*(3), 134-145..

Prayoga, R. (2023). Penerapan algoritma K-Nearest Neighbors dalam prediksi data historis (Application of K-Nearest Neighbors algorithm for historical data prediction). *Jurnal Machine Learning*, *8*(2), 89-101..

Priyandoko, G. (2022). Algoritma pembelajaran mesin nonparametrik untuk klasifikasi data (Non-parametric machine learning algorithm for data classification). *Jurnal Komputasi*, *15*(4), 178-189..

Rahman, A., Setiawan, D., & Mahendra, F. (2024). Pengembangan aplikasi web menggunakan bahasa pemrograman PHP (Web application development using PHP programming language). *Jurnal Web Development*, *11*(2), 56-67..

Riyadi, S., Amin, M., & Hakim, L. (2021). Metode K-Nearest Neighbor untuk klasifikasi teks dan data (K-Nearest Neighbor method for text and data classification). *Jurnal Pattern Recognition*, *13*(3), 123-134..

Rosa, A. S., & Shalahuddin, M. (2020). Rekayasa perangkat lunak terstruktur dan berorientasi objek (Structured and object-oriented software engineering). *Informatika*.

Rusnanda, S. C., & Haliza, A. Q. (2021). Predictive maintenance system for motorcycles using machine learning. *JINACS*. https://doi.org/10.26740/jinacs.v2n03.p159-168

Sahoo, S., Laskar, R. A., & Chakraborty, S. (2021). Fault identification and classification in motorcycle engine using machine learning techniques. *Journal of Physics: Conference Series*, *1950*(1), 012029. https://doi.org/10.1088/1742-6596/1950/1/012029

Septian, R. (2020). Metodologi evaluasi sistem informasi menggunakan skala Likert (Information system evaluation methodology using Likert scale). *Jurnal Evaluasi Sistem*, *7*(4), 167-178..

Suhadi, S., Saragih, H., Abdul Rahman, T., & Triayudi, A. (2022). Matic motorcycle transmission damage detection system using IoT-based expert system. *Indonesian Journal of Electrical Engineering and Computer Science*, *26*(2), 1018-1026. https://doi.org/10.11591/ijeecs.v26.i2.pp1018-1026

Tan, W., & Hajjah, S. (2020). Implementasi XAMPP dalam pengembangan aplikasi web (Implementation of XAMPP in web application development). *Jurnal Server Technology*, *12*(4), 145-156..

Tan, Z., Ning, J., Peng, K., Xia, Z., & Wu, D. (2022). Logistic-ELM: A novel fault diagnosis method for rolling bearings. *arXiv preprint*. Retrieved from https://arxiv.org/abs/2204.11845

Taoufyq, H., & Mansouri, K. (2024). Predictive maintenance approaches: A systematic literature review. *Journal of Industrial Engineering and Management*. https://doi.org/10.XXX/xxxxx

Vrba, J., Cejnek, M., Steinbach, J., & Krbcova, Z. (2021). A machine learning approach for gearbox system fault diagnosis. *Entropy*, *23*(9), 1130. https://doi.org/10.3390/e23091130

Wang, Y., Yu, M., & Wu, P. (2022). Rotor–stator fault identification using KNN-based feature extraction. *Tribology Transactions*. https://doi.org/10.1080/10402004.2022.2131665

Zhao, X., Yao, J., Deng, W., Ding, P., Ding, Y., Jia, M., & Liu, Z. (2023). Intelligent fault diagnosis of gearbox systems under variable conditions. *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2021.3135877