Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Design of an Automatic Indonesian Grammar Error Detection Application Using Machine Learning Algorithms

Robet[1*], Johanes Terang Kita Perangin Angin[2], Jerry Gavin[3*]

[1,3]*Informatics Engineering, STMIK Time, Medan, Indonesia*
[2]*Information Systems, STMIK Time, Medan, Indonesia*
robet@stmik-time.ac.id[1*], timejohanes@gmail.com[2], jerrygavin3@gmail.com[3*]

Abstract

Indonesian, as the country's official language, is crucial in both academic and professional settings. Therefore, writing well and adhering to grammatical standards is crucial. However, many grammatical errors persist in various types of writing. The objective of this research is to design and develop a web-based application that can automatically identify grammatical issues in Indonesian using machine learning techniques, specifically the Support Vector Machine (SVM). The SVM algorithm was chosen for its high accuracy in text classification. An Indonesian dictionary was used as the source dataset. This program can be used as a learning tool in addition to helping users identify and correct grammatical errors in real-time. With 100% accuracy, precision, and recall values, and 0% classification error, the test results demonstrate the application's excellent detection performance. These results demonstrate how well the SVM system is able to detect grammatical issues in Indonesian text.

Keywords: *Indonesian Grammar, Error Detection, Machine Learning, Support Vector Machine, Web Application, Text Classification*

1. Introduction

As the official language of the nation, Indonesian plays an important role in various aspects of daily life. Writing properly, including the use of correct grammar, is highly important, particularly in professional and academic contexts [1]. However, grammatical errors are still frequently found in many forms of writing, such as scientific papers, reports, and even informal writing [2]. The aim of this research is to develop an application capable of automatically identifying grammatical issues in Indonesian by employing machine learning algorithms, specifically the Support Vector Machine (SVM).

The issue of grammatical errors in Indonesian writing has become a serious concern. Alongside the rapid development of information technology, the production of written content continues to increase. However, not all writers possess the same grammatical proficiency. As a result, a large amount of writing still contains grammatical mistakes that may reduce the quality and credibility of the text. Moreover, manual grammar correction can be tedious and prone to human error [3].

To address this problem, this research proposes the development of an automatic Indonesian grammar error detection application using the SVM algorithm. SVM was chosen due to its capability to classify with high accuracy, especially for high-dimensional data such as text [4]. Although this field has been widely studied, achieving high accuracy in detection remains a challenge, particularly when dealing with complex grammatical errors. This study is expected to improve the precision of Indonesian grammar error detection, thereby helping users write more effectively [5].

With the presence of this application, it is expected that users, especially writers and students, will be able to improve the quality of their writing. The application provides real-time grammar correction suggestions so that users can immediately revise identified errors. Furthermore, this program can also serve as a learning tool for the Indonesian language. Based on these considerations, the authors were motivated to conduct this research with the hope of offering an effective solution to the problem of grammatical errors in Indonesian writing.

Based on the description above, the authors were encouraged to carry out a study entitled “Design of an Automatic Indonesian Grammar Error Detection Application Using Machine Learning Algorithms.”

2. Literature Review

2.1. Spelling Correction

Spelling Correction is a system that can correct spelling errors. Such errors may occur due to missing or extra characters, or because of the presence of incorrect characters. Spelling Correction is an essential component of Natural Language Processing (NLP) that functions to identify and correct spelling mistakes in text. It can be useful in various applications, such as information retrieval, auto-completion,

chatbots, and machine translation, where spelling errors may lead to misinterpretation or reduce the accuracy of results. This system typically works by identifying misspelled words and offering alternative words that are correct or more appropriate based on certain analyses.

2.2. Machine Learning

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that focuses on the development of algorithms and models that enable computers to learn from data and make predictions or decisions without being explicitly programmed. ML works by analyzing patterns in existing data and then applying them to new data. In the development of ML-based systems, the performance of algorithms improves as more data is processed, since the resulting models become better at recognizing patterns and making predictions.

2.3. Text Mining

Text mining is generally defined as a knowledge-based process that involves user interaction with a collection of documents using various analytical tools. The primary goal of text mining is to extract useful information from data through the identification and exploration of interesting patterns. Unlike conventional data mining, text mining focuses on unstructured textual data within document collections, rather than structured records in databases. Nevertheless, text mining shares many similarities with data mining systems in terms of high-level architecture. Both systems rely on preprocessing steps, algorithms for pattern discovery, and visualization elements that facilitate the analysis and exploration of results.

2.4. Support Vector Machine

Support Vector Machine (SVM) was first introduced in 1992 by Vapnik, together with Boser and Guyon. The basic concept of SVM is to use a linear classifier, which was later extended to address non-linear problems by applying the kernel trick in a high-dimensional space. SVM is capable of classifying both linear and non-linear data. In the process, predictor variables act as input data, while the corresponding target variables serve as the output. The main objective of SVM is to find the optimal classification function to distinguish between members of two classes in the training data. The concept of the "optimal" classification function can be illustrated geometrically. For linearly separable datasets, the linear classification function is associated with a separating hyperplane f(x), which lies between the two classes and separates them [6]. The SVM algorithm model is one of the classification methods, which works by finding a line (hyperplane) that separates two groups of data.

3. Method

The method in this research was aimed at thoroughly identifying the problem in order to fully understand its root causes. The problem analysis sought to explore the main issues faced, outline the limitations of the existing system, and determine the requirements that must be fulfilled by the proposed solution. By comprehensively understanding the problem, it is expected that the proposed solution will have a significant and relevant impact on user needs. The problem analysis in this study was divided into three parts: analysis of the current process in the field, analysis of the proposed algorithm, and analysis of the proposed system.

3.1. Proposed Algorithm Analysis

The proposed algorithm is the Support Vector Machine (SVM). This subsection explains, in simple terms, how the algorithm works in detecting Indonesian grammar errors. Figure 1 illustrates the flowchart of the implementation stages of the proposed algorithm in this research.

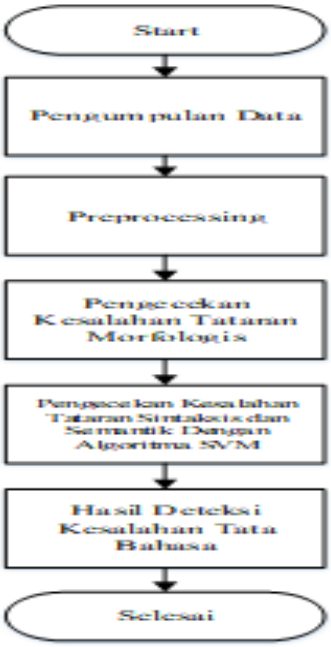

Fig. 1: Flowchart of SVM Implementation Stages

3.2. Proposed System Analysis

In this subsection, the proposed system analysis is conducted to support the development of a more effective solution related to the automatic detection of Indonesian grammar errors using machine learning algorithms. The analysis is carried out using a Use Case Diagram, as shown in Figure 2. below.

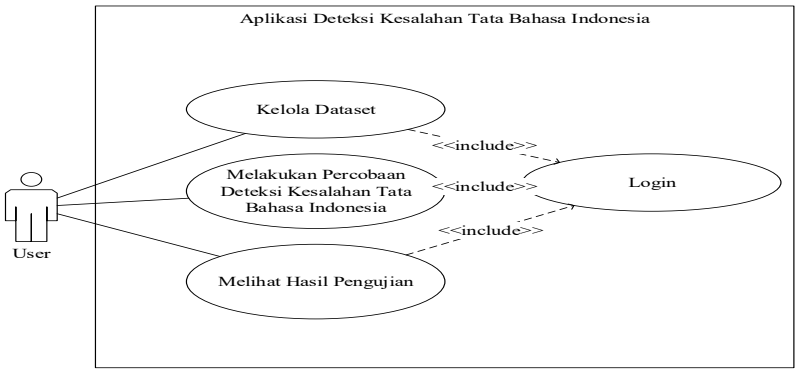

Fig. 2: Use Case Diagram of Indonesian Grammar Error Detection Application

3.3. System Design

System design covers the planning and design process that forms the foundation for developing the application or research project. It is a crucial stage to ensure that the system meets user needs and established objectives. The design process in this research is divided into two parts: interface design and database design.

3.3.1. Interface Design

Interface design is an important stage in application development, where visual and functional elements are designed to create an intuitive, attractive, and user-friendly interface. In this study, the interface design was created using Balsamiq Mockup 3. The interface of the Indonesian grammar error detection application includes the following:

A. Login Page Design.

This page design integrates components such as a login button and form, which must be clicked to access the system dashboard (Figure 3).

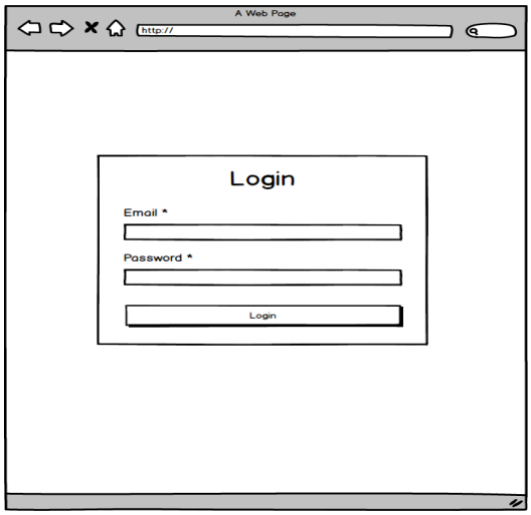

Fig. 1: Login Page Design

B. Dataset Management Page Design.
This page layout displays the list of accessible datasets. Users can also add, update, and delete datasets using the available buttons (Figure 4).

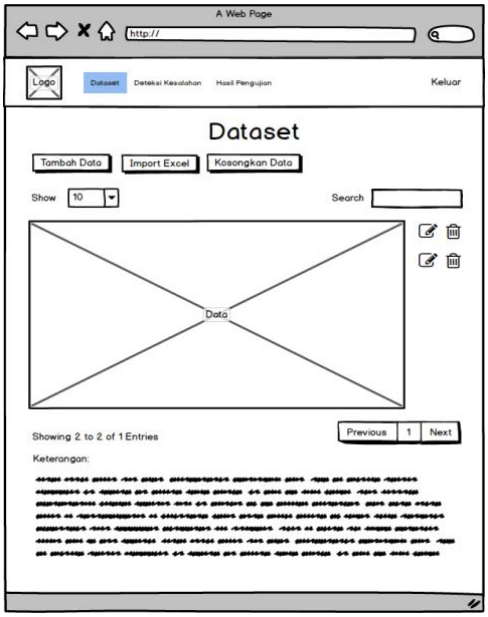

Fig. 2: Dataset Management Page Design

C. Add Dataset Form Page Design.
This form page provides input fields for entering training data. Users only need to fill in the form and then click the submit button to add a dataset (Figure 5).

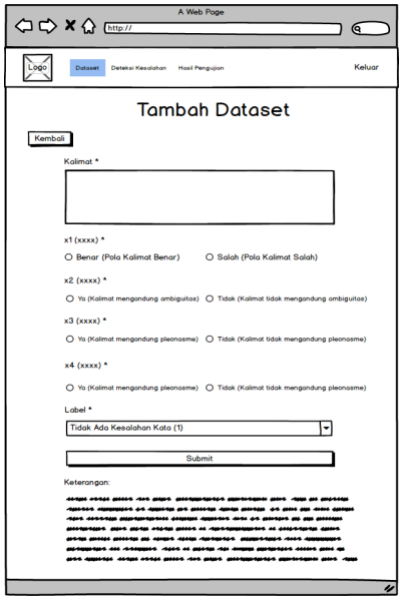

Fig. 3: Add Dataset Form Page Design

D. Edit Dataset Form Page Design.

This page allows users to modify existing dataset information by editing the form fields and clicking the submit button to update the dataset (Figure 6).

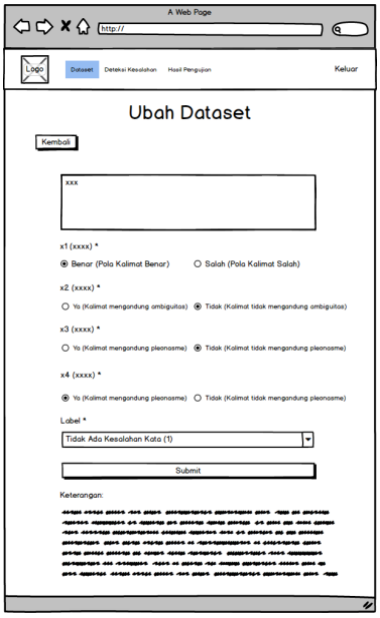

Fig. 4 Edit Dataset Form Page Design

E. Grammar Error Detection Page Design.

This page provides a form for checking Indonesian grammar errors using the Support Vector Machine algorithm (Figure 7).

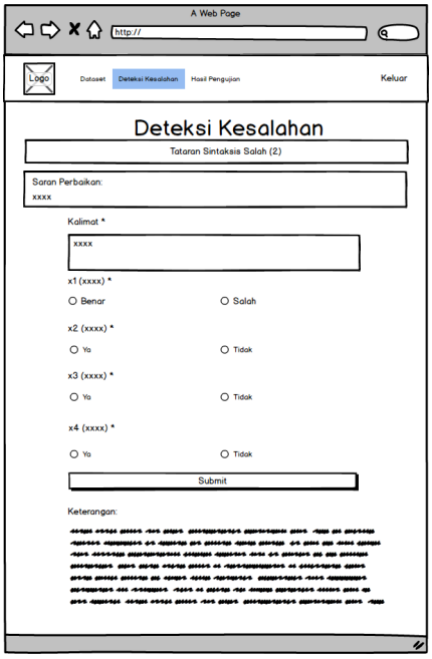

Fig. 5 Grammar Error Detection Page Design

F. Testing Page Design.
This page displays the testing results of the Confusion Matrix, showing accuracy, precision, recall, and misclassification error (Figure 8)

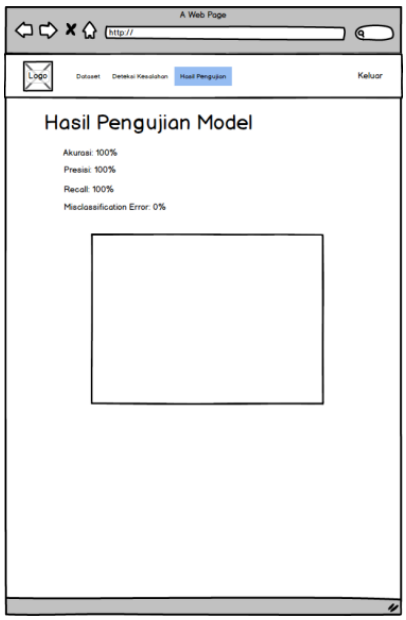

Fig. 6 Testing Page Design

3.3.2. Database Design

Database design is a crucial aspect of developing efficient and reliable software. In this stage, the system's database was designed using an Entity Relationship Diagram (ERD). Figure 9 shows the ERD of the Indonesian grammar error detection application.

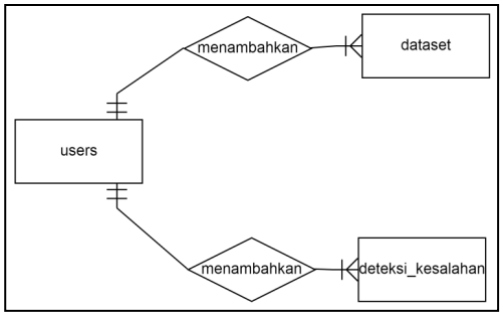

Fig. 7: ERD of the Indonesian grammar error detection application

Furthermore, the following tables present the structure of each entity in the ERD (Tables 1 to 3):

Table 1: Users Table Structure

Kolom	Jenis	Kosong	Bawaan
id *(Primary Key)*	int(11)	Tidak	
nama_lengkap	varchar(50)	Tidak	
email	varchar(50)	Tidak	
password	varchar(255)	Tidak	

Table 2: Dataset Table Structure

Kolom	Jenis	Kosong	Bawaan
id_dataset *(Primary Key)*	int(11)	Tidak	
id *(Foreign Key)*	int(11)	Tidak	
kalimat	text	Tidak	
x1	int(11)	Tidak	
x2	int(11)	Tidak	
x3	int(11)	Tidak	
x4	int(11)	Tidak	
label	text	Tidak	

Table 3: Grammar Error Detection Table Structure

Kolom	Jenis	Kosong	Bawaan
id_deteksi *(Primary Key)*	int(11)	Tidak	
id *(Foreign Key)*	int(11)	Tidak	
kalimat	text	Tidak	
kesimpulan	text	Tidak	
perbaikan	text	Tidak	

4. Result

The following will explain the appearance of Grammar Error Detection Application , which can be seen as follows.
1. Login Page

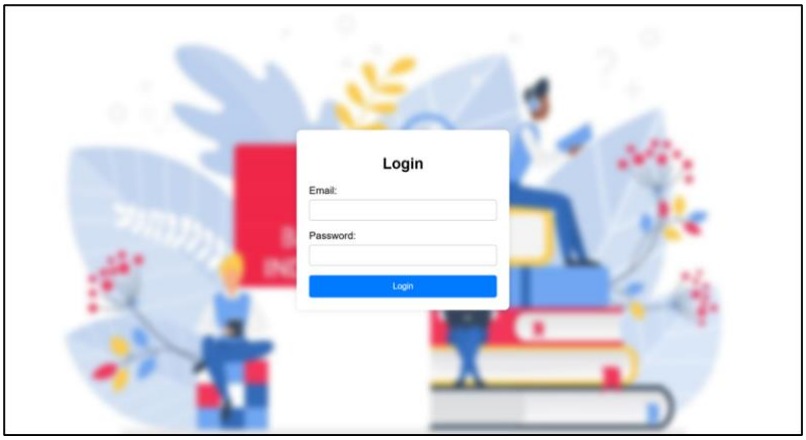

Fig. 10: Login Page

The login page consists of components such as a login button and form, which users must complete to access the system dashboard. The login/homepage is shown in Figure 10.

2. Dataset Management Page

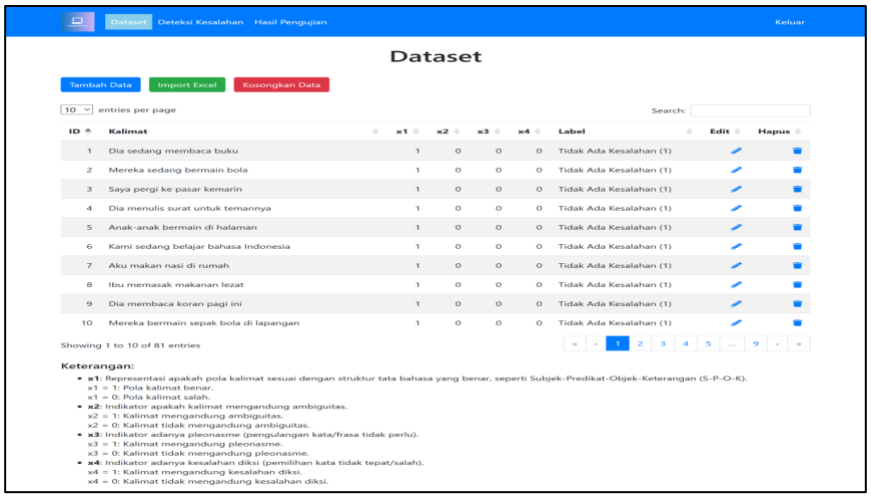

Fig. 11: Dataset Management Page

The dataset management page displays the available datasets. Users can add, update, and delete datasets using the provided buttons. The dataset management page is shown in Figure 11.

3. Add Dataset Form Page

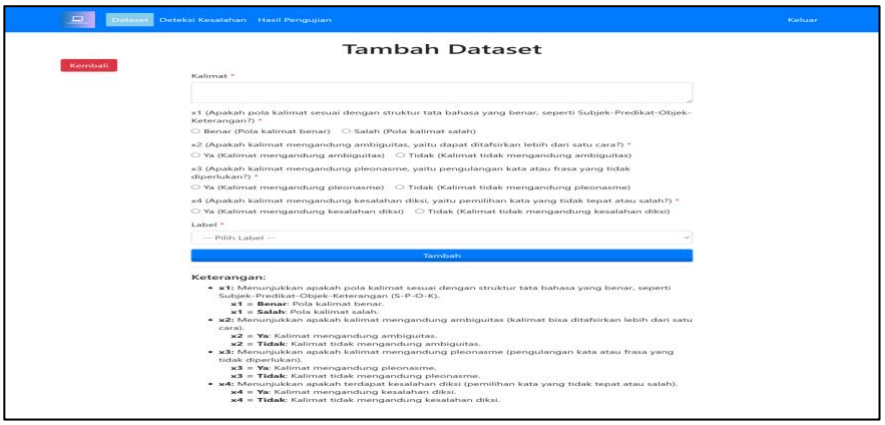

Fig. 12: Add Dataset Form Page

The add dataset form page provides input fields for entering training data. Users simply need to fill in the form and then click the submit button to add a dataset. Figure 12 shows the add dataset form design.

4. Edit Dataset Form Page

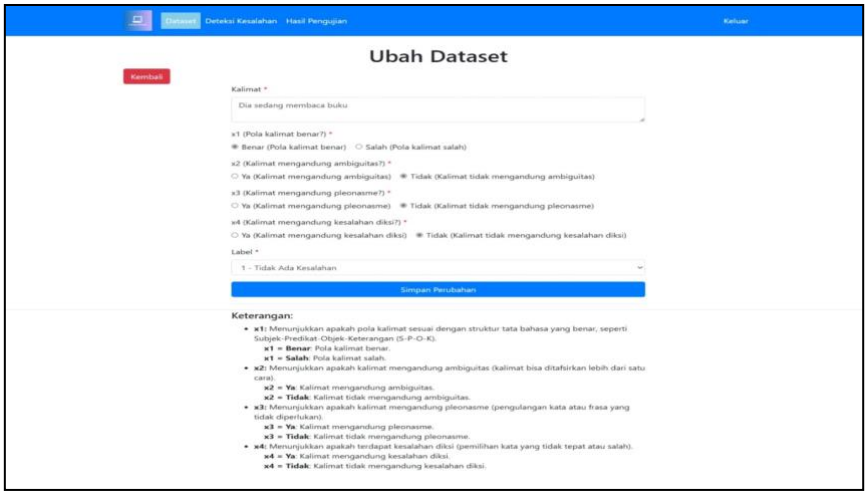

Fig. 13: Edit Dataset Form Page

The edit dataset form page provides input fields for modifying dataset information. Users can edit the pre-filled form and then click the submit button to update the dataset. Figure 13 shows the edit dataset form design.

5. Grammar Error Detection Page

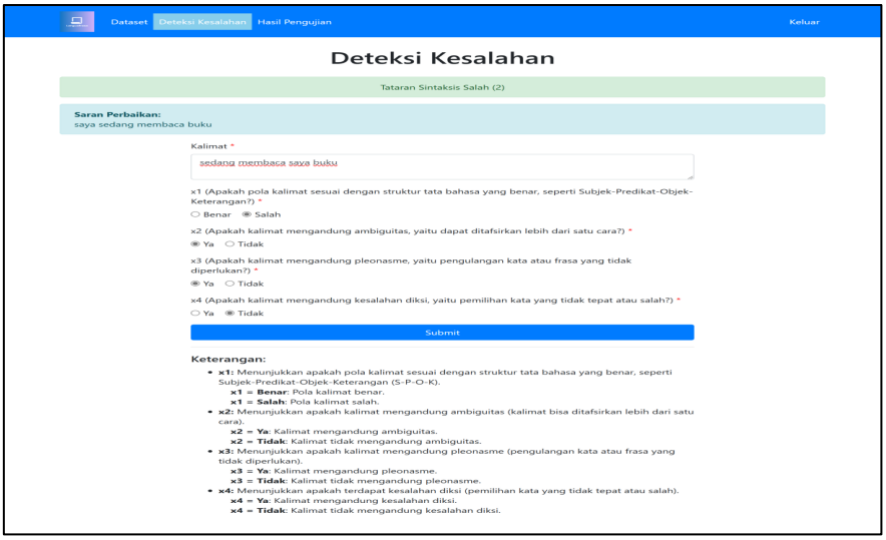

Fig. 14: Grammar Error Detection Page

This page contains a form for checking Indonesian grammar errors using the Support Vector Machine algorithm. Figure 14 shows the design of the grammar error detection page design.

6. Testing Result Page

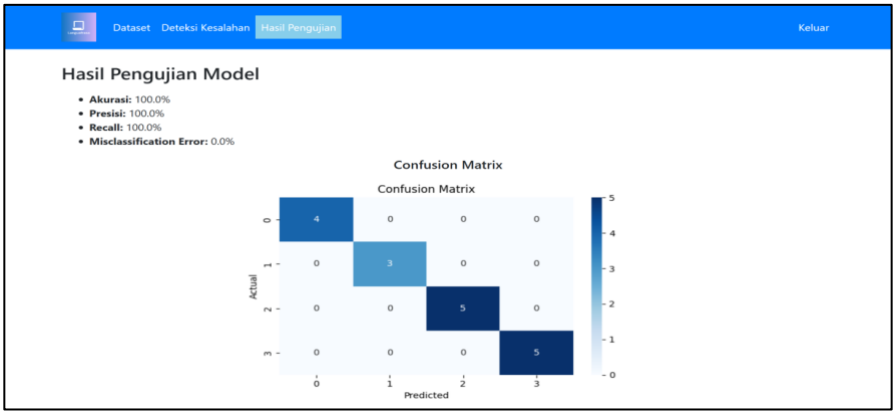

Fig. 15:Testing Result Page

This page displays the results of testing the grammar error detection system. Figure 15 shows the testing results page.

5. Conclusion

The conclusions from this research are as follows:

1. Further exploration of other machine learning algorithms, such as Random Forest, LSTM, or Transformer-based models (e.g., BERT), is recommended, as these may be more adaptive to sentence context and capable of improving performance with more complex data.
2. The dataset should not be limited to dictionary-based sources but should also include texts from various types of writing (articles, social media, essays, conversations). This would enable the model to detect grammatical errors in a broader and more realistic context.
3. It is recommended that the application be developed not only as a web-based tool but also as mobile and desktop applications. This would enhance accessibility and convenience, especially for users who wish to use the application offline or on mobile devices

References

[1] R. Sofiani, S. Rofi'ah, and L. Putriyanti, "Peran Bahasa Indonesia Di Era Globalisasi Saat Ini Untuk Menunjang Prestasi Siswa," in Prosising Sendika, 2023, vol. 4, no. 1, pp. 150–158, [Online]. Available: https://journal.upy.ac.id/index.php/pkn/article/view/4221.

[2] R. Yusuf, S. Budiawan, R. F. Mualafina, and S. Ulfiyani, "Kesalahan Penerapan Kaidah Bahasa Indonesia dalam Karya Tulis Mahasiswa pada Mata Kuliah Bahasa Indonesia di Universitas PGRI Semarang," Transform. J. Bahasa, Sastra, dan Pengajarannya, vol. 3, no. 1, pp. 87–103, 2019,

doi: 10.31002/transformatika.v3i1.1186.
[3] I. S. K. Idris and Y. A. Mustofa, "Typo Checking Menggunakan Algoritma Rabin-Karp," Jambura J. Electr. Electron. Eng., vol. 4, no. 1, pp. 87–91, 2022, doi: 10.37905/jjeee.v4i1.12150.
[4] T. Setiawan, S. Liem, and D. M. R. Pribadi, "Perbandingan Algoritma SVM dan Naïve Bayes dalam Analisis Sentimen Komentar Tiktok pada Produk Skincare," Appl. Inf. Technol. Comput. Sci., vol. 3, no. 2, pp. 28–32, 2024, [Online]. Available: https://jurnal.politap.ac.id/index.php/aicoms.
[5] A. M. B. Ledjap, F. P. Rochmawati, D. A. E. Marsanda, and A. P. Sari, "Pemanfaatan Natural Language Processing Untuk Pengecekan Ejaan Sesuai KBBI," JAMASTIKA, vol. 3, no. 2, pp. 46–56, 2024.
[6] R. A. Rizal, I. S. Girsang, and S. A. Prasetiyo, "Klasifikasi Wajah Menggunakan Support Vector Machine (SVM)," REMIK (Riset dan E-Jurnal Manaj. Inform. Komputer), vol. 3, no. 2, p. 1, 2019, doi: 10.33395/remik.v3i2.10080.