International Journal Of Health, Engineering And Technology (IJHET) *E-ISSN 2829 - 8683*
Volume 5, Number 2, July 2026, *Page. 1941 - 1952*
Email : editorijhess@gmail.com

Development Of A Web-Based Machine Learning Application For Identifying Students' Interests And Talents In Surabaya

Bagus Adianto[1)], Made Kamisutara[2)]
[1,2)] Informatics Engineering, Faculty of Computer Science, Universitas Narotama, Indonesia

*Corresponding Author
E-mail: bagusadianto5@gmail.com

Abstract

This study aims to develop a web-based machine learning application that can identify students' interests and talents in Surabaya in a more accurate and systematic manner. A mixed-method approach was employed, combining quantitative and qualitative methods to obtain comprehensive results in the system development process. Data were collected from 100 high school students, covering academic performance, cognitive abilities, personality traits, and social skills. The data were then processed through preprocessing stages, including normalization, handling missing values, and outlier removal to ensure data quality for modeling. The dataset was split into 50% training data and 50% testing data to build and evaluate the model using the Random Forest algorithm, which operates through bootstrapping, random feature selection, and majority voting. Model evaluation was conducted using accuracy, precision, recall, and F1-score metrics, and further validated through cross-validation to ensure model generalization and reduce overfitting. The research variables include academic data, non-academic data, and psychometric test results as independent variables, while the dependent variables are major recommendations and user satisfaction levels. The evaluation results show that the system performs well with an accuracy of 83%, precision of 87.5%, recall of 80%, and F1-score of 88.52%. The system was developed using Vue.js, Express.js, and Python and can be accessed via both desktop and mobile devices. Overall, the system effectively assists students in identifying their potential and supports more structured educational planning.

Keywords: ***Sea Surface Temperature, warming trend, Indian Ocean Dipole, lag correlation, GLORYS12v1, Southern Java.***

INTRODUCTION

Interest and talent are essential aspects of students' educational development. Both often serve as primary determinants in shaping future pathways, whether in education or career. Research indicates that talent, as an innate ability requiring development through practice, must be supported by interest in order to achieve optimal learning outcomes ((Fitri et al., 2023). In the context of higher education, selecting a field of study that aligns with students' interests and talents is a crucial decision. The development of these aspects requires special attention through continuous experience and exploration, enabling students to grow optimally without coercion, but rather with genuine enthusiasm and motivation (Basri et al., 2021).

Determining students' interests and talents often presents challenges for both students and parents. A study shows that approximately 87 percent of university students in Indonesia feel they have chosen the wrong major due to a lack of understanding of their own potential (Wasilah et al., 2023). Therefore, it is essential to identify and determine students' interests and talents early on so they can develop in accordance with their potential. This process requires sustained exploration and adaptive methods. For example, activities such as brainstorming, pre-test and post-test assessments, and educational programs have proven effective in enhancing students' understanding of their potential, as implemented at SMPN 3 Jatinangor in 2024 (Iceu Amira & Rosidin, 2024). Such strategies enable students to independently identify their interests and talents while reducing reliance on external pressure (Iceu Amira & Rosidin, 2024).

In the era of technological advancement, traditional approaches such as manual counseling or interviews often face limitations in terms of accuracy and time efficiency. Studies indicate that expert system-based approaches using Forward Chaining have been able to accelerate the identification of students' interests and talents, although further development is needed to improve accuracy,

particularly in cases involving multiple talents (Aulia, 2023). This highlights the need for innovative solutions capable of providing more in-depth and accurate analysis. One such approach is machine learning, which can process both academic and non-academic data comprehensively to generate relevant analyses and recommendations (Alfarizi & Al-farish, 2023).

A significant breakthrough is the application of the Profile Matching method within decision support systems. Research conducted at SMA Tunas Budaya Jakarta (2023) demonstrated that this method enhances objectivity in selecting high-achieving students for academic competitions by integrating both intracurricular and extracurricular data (SHELEMO, 2023). Furthermore, the integration of artificial intelligence (AI) in education, as implemented at SMA Negeri 11 Bulukumba (2024), introduced AI-based training to foster students' interest in technology. As a result, students showed increased awareness of their potential through simulations and algorithm-based creative projects (Kaswar et al., 2024).

Machine learning is selected due to its superiority over traditional methods. This technology is capable of recognizing complex data patterns and generating accurate predictions, even with large datasets. Additionally, machine learning enables real-time analysis, providing fast and efficient results. In this study, algorithms such as Random Forest are considered due to their ability to produce reliable predictive models with minimal error (Roihan et al., 2020). In the financial sector, the implementation of Random Forest for motor vehicle credit feasibility assessment in Indonesia achieved an accuracy of 78.60% and an AUC value of 0.907, which falls into the "very good" category (Pahlevi et al., 2023). Meanwhile, research on apartment price prediction during the Covid-19 era using Random Forest produced an R^2 accuracy of 0.977, although the Backpropagation method performed better (Khairani et al., 2022).

By leveraging machine learning, a web-based application can be developed to assist students, teachers, and parents in better understanding students' potential. This application is expected to provide recommendations for academic majors or career paths that align with students' interests and talents, thereby facilitating better decision-making. This solution not only addresses the issue of choosing an unsuitable major but also supports the continuous development of students' potential within the educational landscape.

RESEARCH METHODS

This study employs a mixed-method approach, combining quantitative and qualitative methods to obtain comprehensive results in developing a web-based interest and talent identification application. Data were collected from 100 high school students, covering academic aspects, cognitive abilities, personality traits, and social skills, and then processed through preprocessing stages such as normalization, handling missing data, and removing outliers to ensure data quality. The dataset was divided into 50% training data and 50% testing data to build and evaluate the model using the Random Forest algorithm, which operates through bootstrapping, random feature selection, and majority voting to generate predictions. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics, and further validated באמצעות cross-validation to ensure good generalization and minimize overfitting. Additionally, this study applied control variables to maintain result validity, with independent variables including academic data, non-academic data, and psychometric tests, while the dependent variables consist of major recommendations and user satisfaction levels with the system.

RESULTS AND DISCUSSION

Website Implementation

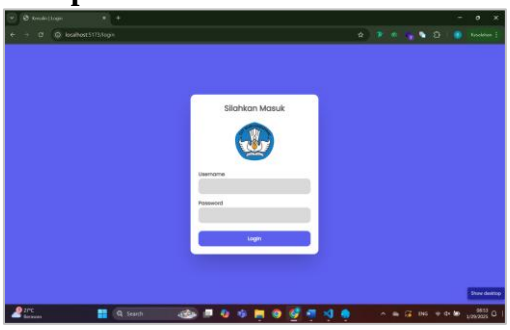

Figure 1. shows the Login Page

Figure 2. display the Dashboard pages.

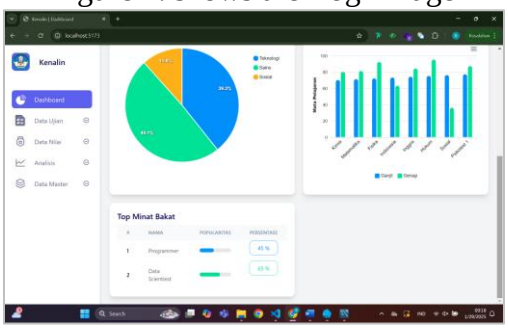

Figure 3. display the Dashboard pages

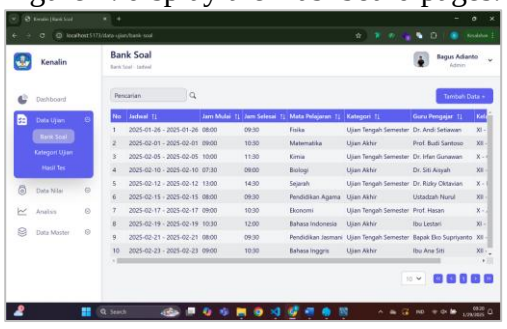

Figure 4. presents the initial Question Bank page.

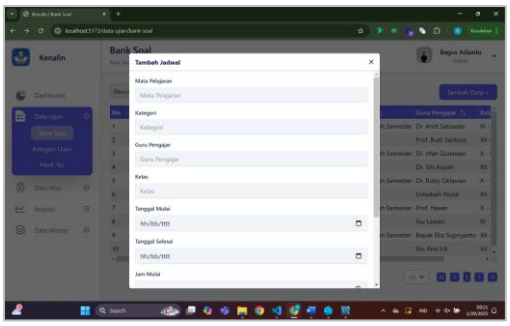

Figure 5. shows the input form for adding an exam schedule.

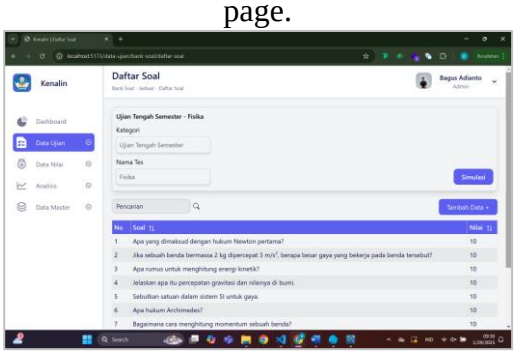

Figure 6. display the question list page and the modal form for adding questions.

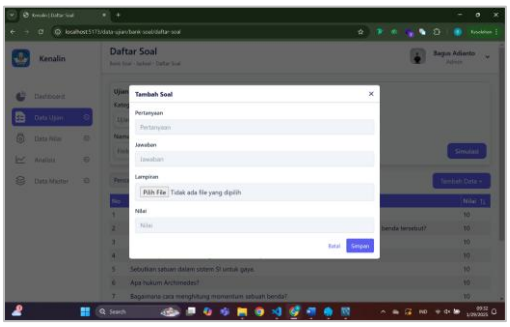

Figure 7. display the question list page and the modal form for adding questions.

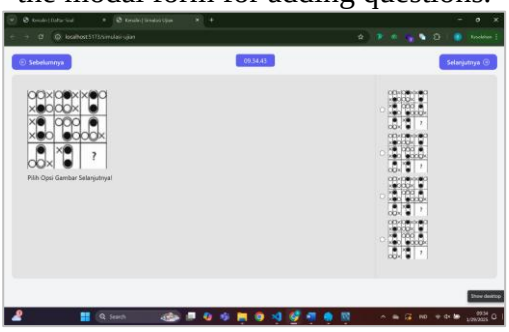

Figure 8. illustrates the exam simulation page.

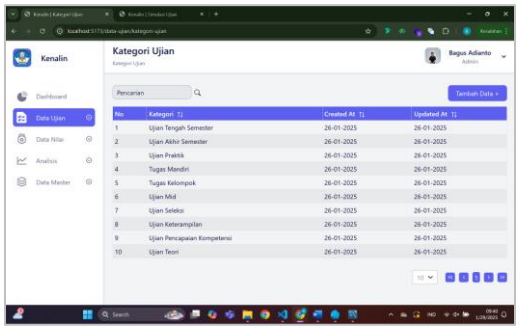

Figure 9. present the exam category data page and the modal form for adding categories.

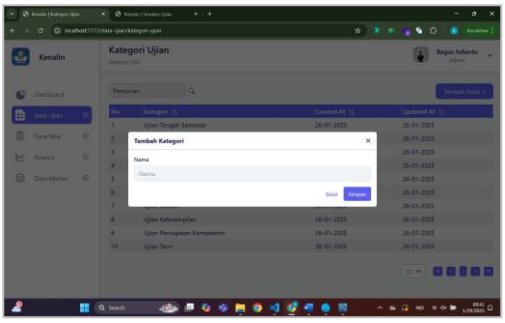

Figure 10. present the exam category data page and the modal form for adding categories.

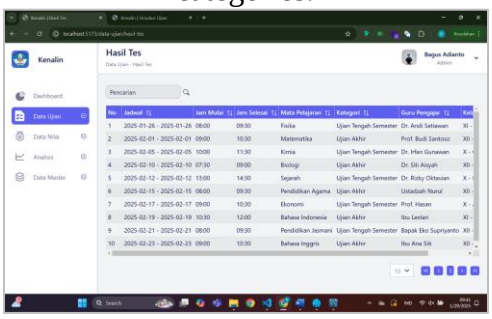

Figure 11. show the test results page and detailed test results.

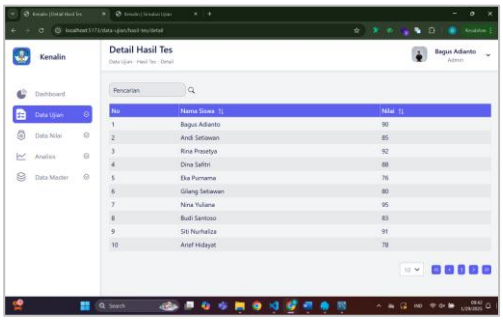

Figure 12. show the test results page and detailed test results.

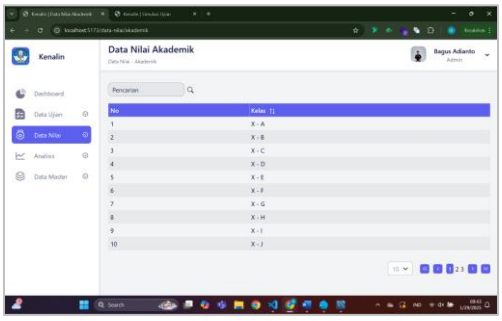

Figure 13. Academic or Non-Academic Score Data Page

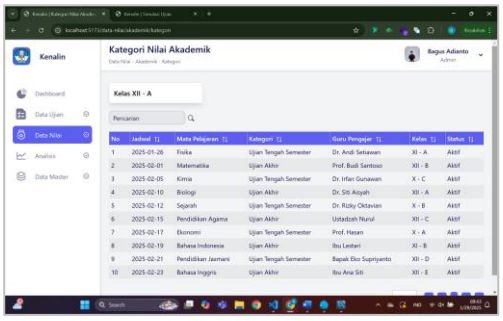

Figure 14.Detail Score Page Based on Class or Category

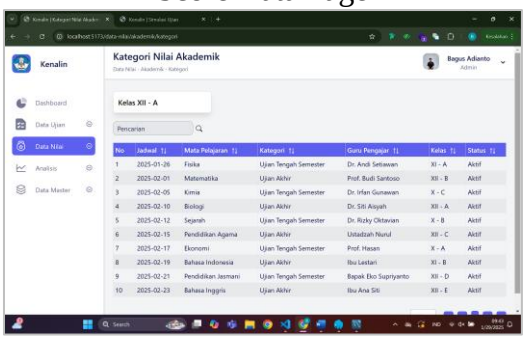

Figure 15. Detail Score Page Based on Schedule

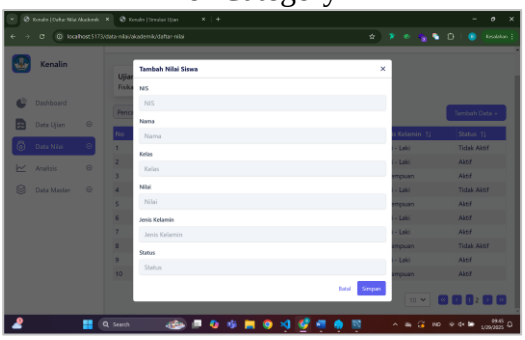

Figure 16. Add Score Data Form Modal

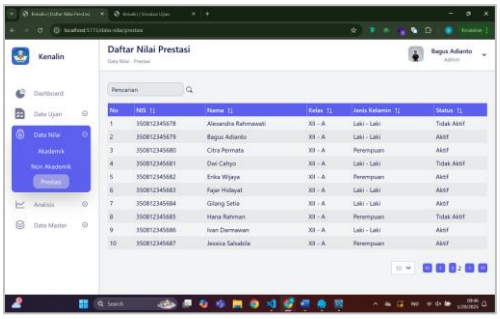

Figure 17. Student Data Page (Achievement Menu)

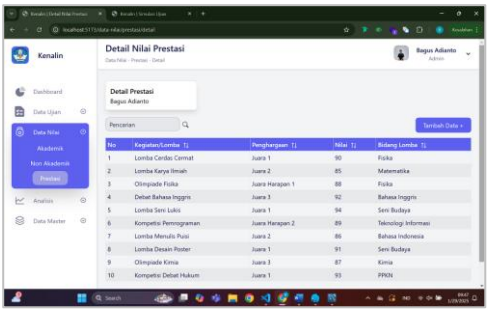

Figure 18. Student Achievement Detail Page

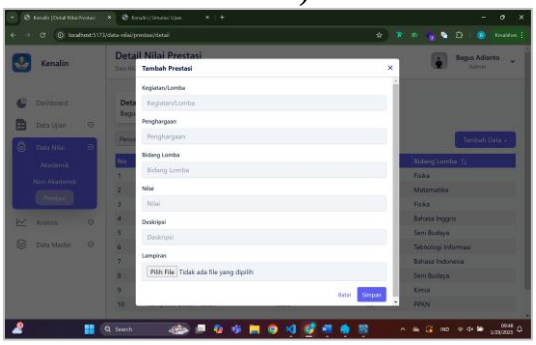

Figure 19. Add Achievement Form Modal

Figure 20. Individual Report Page

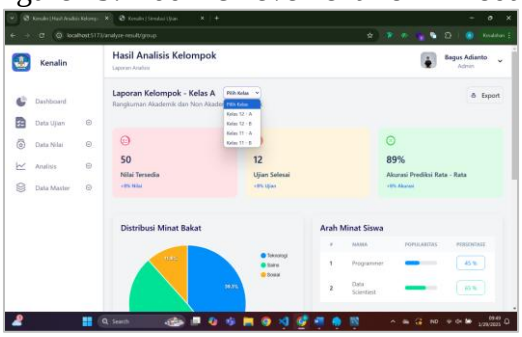

Figure 21. Group Report Page

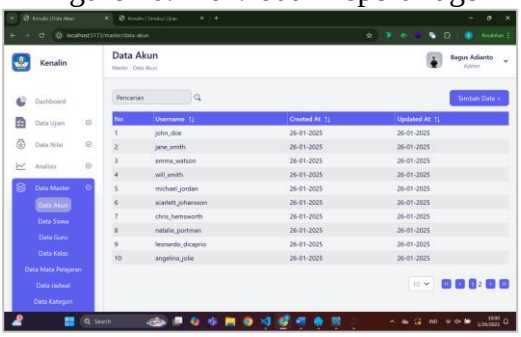

Figure 22. Account Data Page

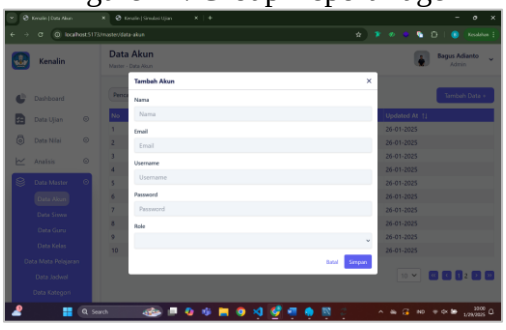

Figure 23. Add Account Form Modal

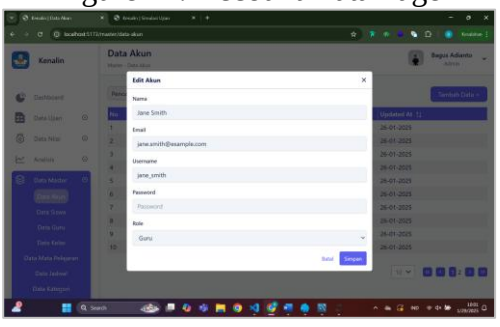

Figure 24. Edit Account Form Modal

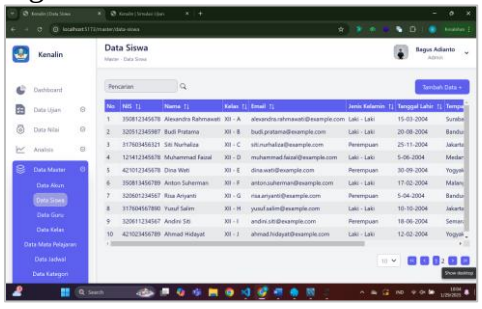

Figure 25. Student Data Page

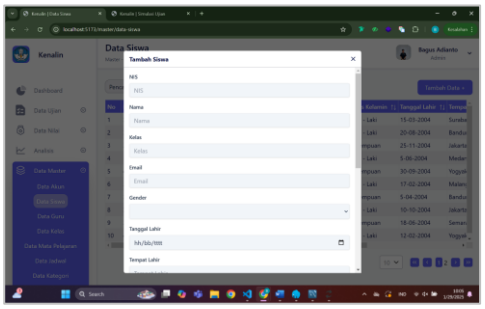

Figure 26. Add Student Form Modal

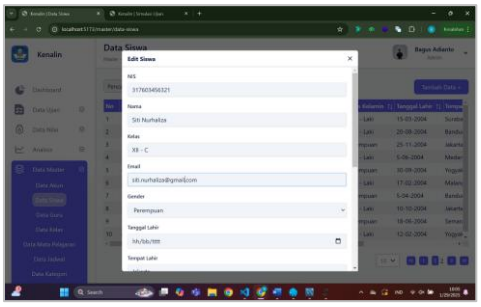

Figure 27. Edit Student Form Modal

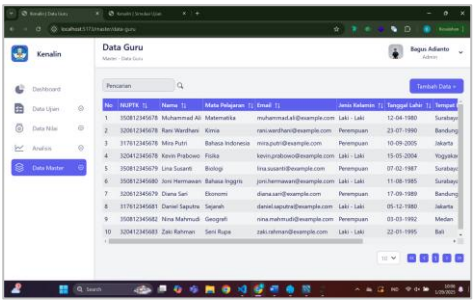

Figure 28. Teacher Data Page

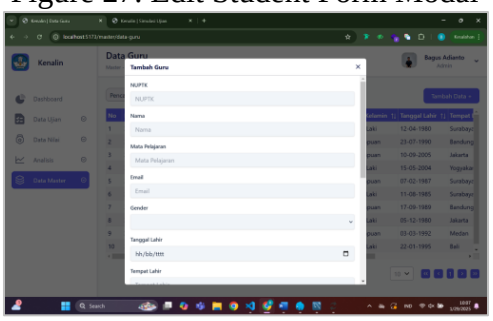

Figure 29. Add Teacher Form Modal

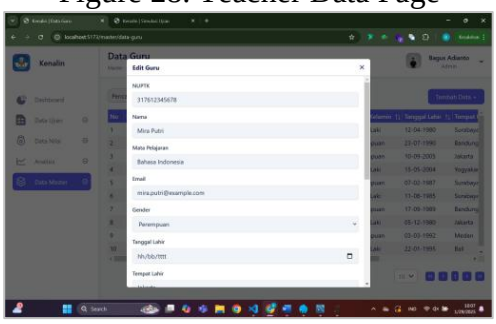

Figure 30. Edit Teacher Form Modal

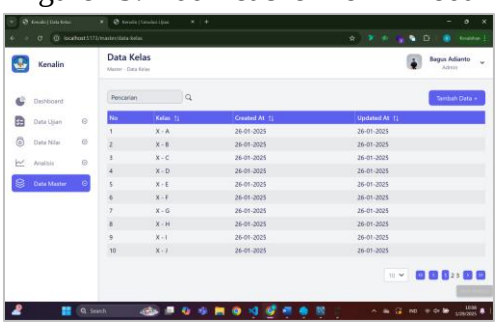

Figure 31. Class Data Page

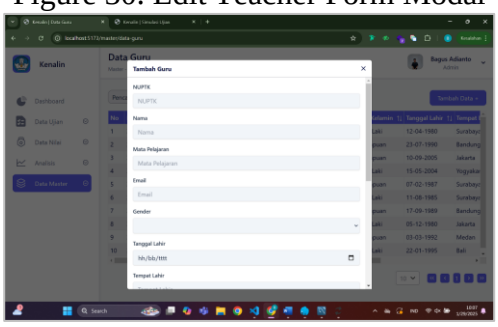

Figure 32. Add Class Form Modal

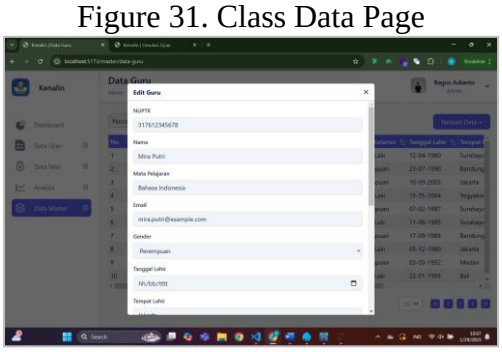

Figure 33. Edit Class Form Modal

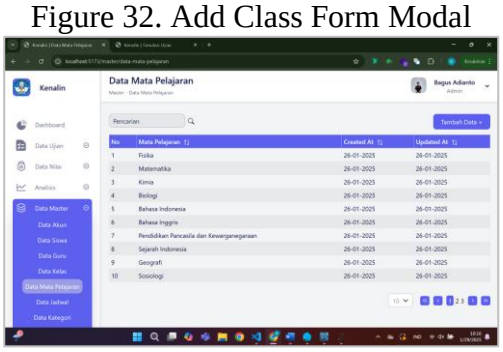

Figure 34. Subject Data Page

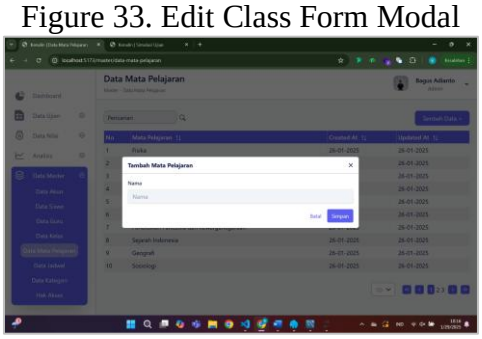

Figure 35. Add Subject Form Modal

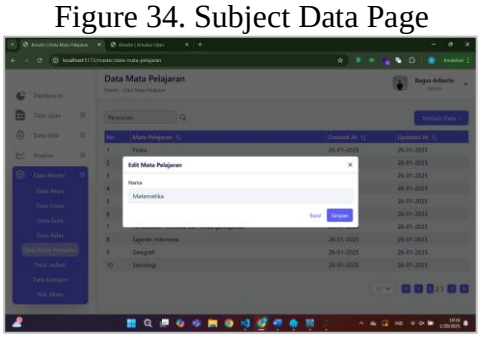

Figure 36. Edit Subject Form Modal

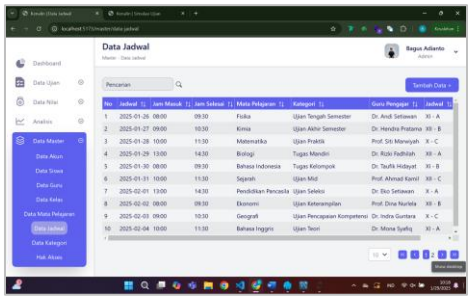

Figure 37. Schedule Data Page

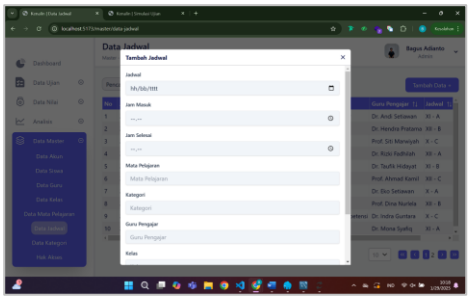

Figure 38. Add Schedule Form Modal

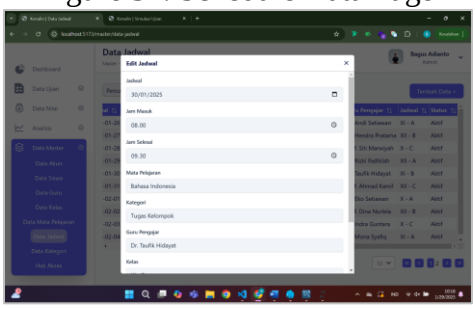

Figure 39. Edit Schedule Form Modal

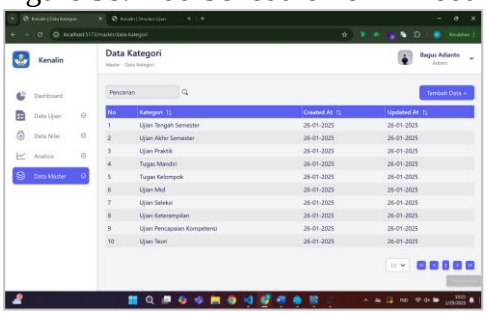

Figure 40. Category Data Page

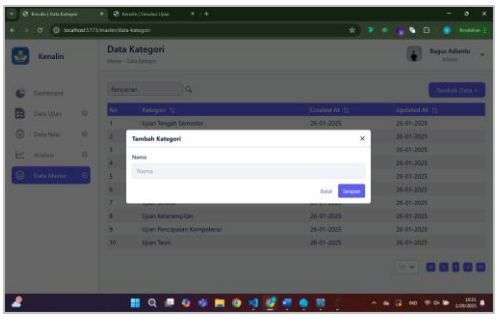

Figure 41. Add Category Form Modal

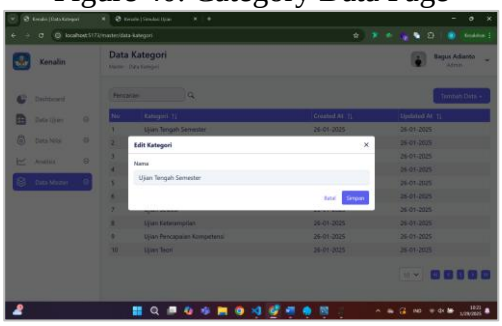

Figure 42. Edit Category Form Modal

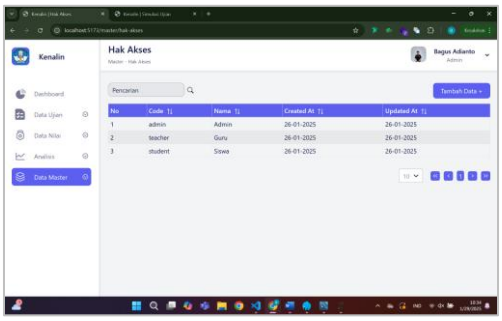

Figure 43. Access Rights Data Page

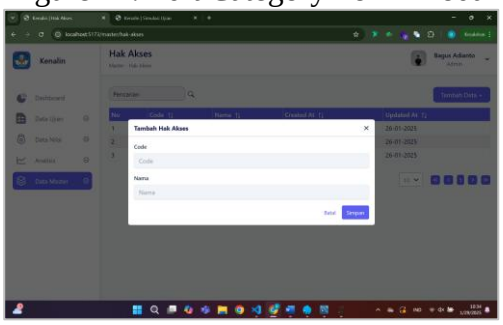

Figure 44. Add Access Rights Form Modal

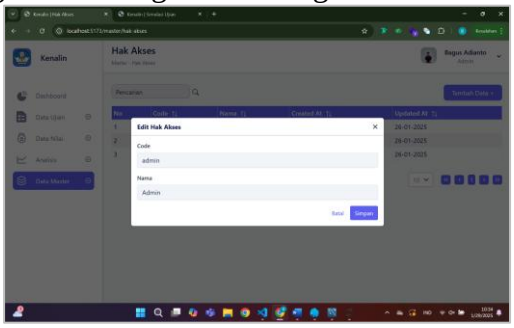

Figure 45. Edit Access Rights Form Modal

Machine Learning Data Analysis

In this section, an in-depth analysis of students' interests and talents data is conducted using machine learning techniques, with the aim of predicting interest and talent categories based on previously selected features. As an initial step, the cleaned and preprocessed data is used to build a prediction model based on the Random Forest algorithm, which is one of the commonly used classification methods due to its high accuracy and ability to handle complex and non-linear data.

To visualize the analysis results and understand the model performance, several diagrams and evaluation metrics are used, as follows:

Interest and Talent Distribution

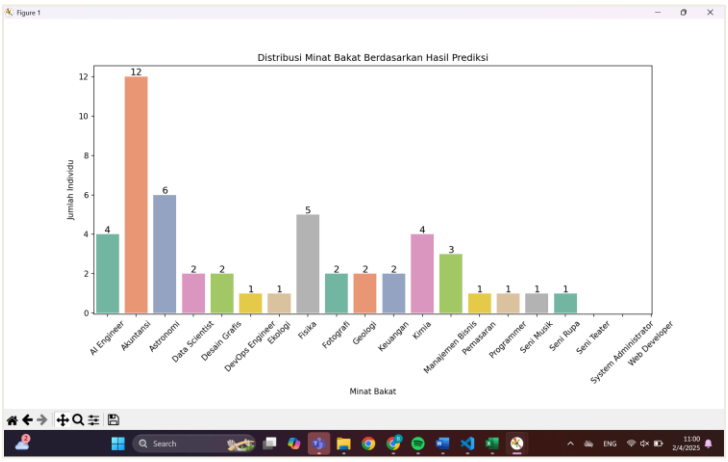

Figure 46. Interest and Talent Distribution Diagram

The diagram above presents the results of students' academic performance data from semester 1 to semester 4, which has been processed and classified using the Random Forest machine learning algorithm. This analysis produces a distribution of interest and talent categories based on patterns identified in students' academic scores. Each category is grouped according to the detected tendencies within the dataset, providing an overview of students' potential and interests based on their academic performance. This information can be used as a basis for determining appropriate educational pathways or extracurricular activities, as well as supporting the development of a more accurate predictive model in the future.

Correlation Heatmap of Numerical Features

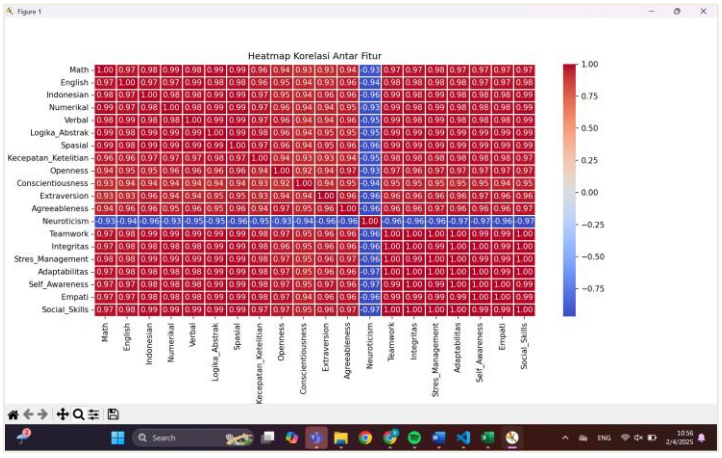

Figure 47. Correlation Heatmap

The correlation heatmap of numerical features illustrates the relationships between various variables within the dataset. Each cell represents the correlation value between two different numerical features, ranging from -1 to +1. A value close to +1 indicates a strong positive relationship, meaning that when one feature increases, the other tends to increase as well. Conversely, a value close to -1

indicates a strong negative relationship, where an increase in one feature is associated with a decrease in the other.

In this heatmap, several meaningful patterns can be observed that help in understanding inter-feature relationships. For example, strong correlations may exist between subjects such as mathematics and science, indicating that students who perform well in mathematics also tend to achieve high performance in science. On the other hand, features with low or near-zero correlation indicate that changes in one variable have little to no effect on others. This heatmap is highly useful for feature selection in machine learning, as highly correlated features may be combined or selected as key variables, while low-correlation features may be removed or adjusted.

Model Performance Metrics

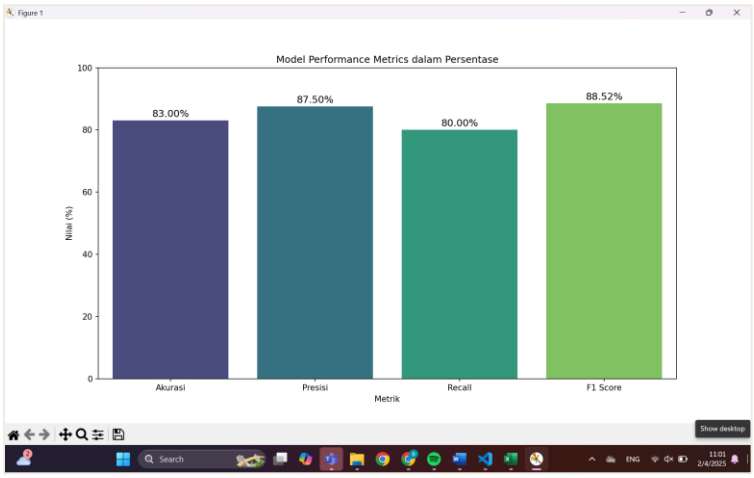

Figure 48. Model Performance Metrics

Accuracy = 83%

Accuracy is a metric that measures how well the model correctly predicts outcomes based on the available data. In this case, an accuracy of 83% means that 83% of all predictions made by the model match the actual results. This metric provides a general overview of model performance, but it can be slightly misleading if the classes in the dataset are imbalanced. Nevertheless, this relatively high accuracy indicates that the model performs fairly well overall in making predictions.

Precision = 87.5%

Precision measures how many of the predicted positive cases are actually positive. In this case, a precision of 87.5% means that 87.5% of all predictions classified as positive truly belong to the positive class. Precision is particularly important when the goal is to minimize false positives, or errors in predicting something that is actually not present. In this context, the model successfully identifies most of the positive instances correctly.

Recall = 80%

Recall measures how many actual positive cases are correctly identified out of all existing positive instances. In other words, a recall of 80% indicates that the model successfully captures 80% of all positive data in the dataset. This metric is important for evaluating how well the model detects all relevant cases and avoids false negatives (failing to identify positive instances). This relatively high recall suggests that the model does not miss many important positive cases.

F1-Score = 88.52%

The F1-Score is a combined metric of precision and recall that reflects the balance between the two. With an F1-Score of 88.52%, the model demonstrates very strong performance, indicating a good balance between minimizing false positives and capturing most of the relevant data. The F1-Score is particularly useful when dealing with imbalanced class distributions, as it provides a more comprehensive evaluation than relying solely on precision or recall.

Overall, this model demonstrates very strong performance, as indicated by the high values across all evaluation metrics. The high accuracy, along with solid precision, recall, and F1-Score, suggests that the model is capable of effectively identifying and predicting outcomes within the given dataset.

Comparison Between Manual Methods and Machine Learning

Machine Learning Method

The data processing and an alysis method using machine learning is implemented with Python as follows:

Data Preparation

Data Used: Student data consisting of academic scores (Mathematics, English, Indonesian, etc.) and results from interest and aptitude tests.

Formulas and Analysis Logic

Data Cleaning

Duplicate data are removed based on the *Name* and *Semester* columns to ensure that each student has unique data.

Formula:

df = df.drop_duplicates(subset=['Name', 'Semester'])

Missing values are handled using the following formulas:

df[numerical_columns] = df[numerical_columns].fillna(df[numerical_columns].mean())df['Interest'] = df['Interest'].fillna('Unknown')

Modeling with Random Forest Classifier

Separating features and target

Features (X) consist of all numerical columns except *Name* and *Interest*, while the target (y) is the *Interest* column.

Formula:

X = df_combined.drop(columns=['Name', 'Interest'])

y = df_combined['Interest']

Splitting data into training and testing sets

The dataset is divided into 75% training data and 25% testing data.

Formula:

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)

Training the model

The Random Forest Classifier is used to predict students' interests based on the provided features.

Formula:

model = RandomForestClassifier(n_estimators=100, random_state=42)

model.fit(X_train, y_train)

Predicting test data

The model is used to predict students' interests on the test dataset.

Formula:

y_pred = model.predict(X_test)

Manual Method

The manual data processing and analysis method is conducted as follows:

Data Preparation

Required Data:

Student names, academic or non-academic scores, and results of interest/aptitude tests (psychological tests).

Create Data Table

Data Cleaning

Removing Duplicates

If duplicate student data exist (e.g., same name and semester), they should be removed.

Handling Missing Data

Missing data are filled with default values, such as 0 for numerical data or "null" for categorical data.

Data Analysis

Academic Performance Analysis

Formula for average student score:

(Score Semester 1 + Semester 2 + …) / Number of Semesters

Formula for class average comparison:

Class Average = Total scores of all students / Number of students

Interest and Talent Analysis

Categorizing interests

Students are grouped based on interest categories (e.g., Technology, Science, Arts, Economics).

Formula:

Number of Students = Count(Interest = Technology)

Correlation Analysis

Identify the relationship between academic performance and student interests.

Formula:

Average = Total score / Number of students

Interest and Talent Prediction (Manual)

Define Logical Rules

Example rule:

If Mathematics score > 80 and Science_Social = 1, then predicted interest = Technology.

Prediction Evaluation

Compare predicted interests with actual student interests.

Formula:

Accuracy = (Number of correct predictions / Total predictions) × 100%

CONCLUSIONS

The interest and talent analysis system using a machine learning approach through a web-based platform has been successfully designed and developed through several stages, involving technologies such as Vue.js, Express.js, and Python. This system aims to assist high school students in Surabaya in identifying their interests and talents more accurately and systematically.

Based on the evaluation results of the machine learning model employed, the system demonstrates excellent performance with the following metrics:

- **Accuracy:** 83%, indicating that the model correctly predicts 83% of the total data.
- **Precision:** 87.5%, suggesting that 87.5% of the positive predictions made by the model are correct.
- **Recall:** 80%, showing that the model successfully captures 80% of all actual positive data.
- **F1-Score:** 88.52%, reflecting a strong balance between precision and recall, as well as optimal overall model performance.

These results indicate that the system is not only accurate in predicting students' interests and talents but also effective in minimizing prediction errors (false positives and false negatives). Therefore, this system can serve as an effective tool for students, teachers, and parents in understanding students' potential and planning more targeted development strategies.

Furthermore, the web-based approach enables the system to be easily accessed by users via both computers and mobile devices. This makes the system more flexible and scalable for implementation across various schools in Surabaya.

Overall, this interest and talent analysis system has successfully achieved its primary objective, which is to provide an innovative, technology-based solution to assist students in identifying their

potential. In the future, the system can be further developed by incorporating additional features, such as recommendations for university majors or career paths based on the analysis results, as well as enhancing the machine learning model to improve accuracy and predictive performance.

REFERENCES

Alfarizi, M. R., & Al-farish, M. Z. (2023). *PENGGUNAAN PYTHON SEBAGAI BAHASA PEMROGRAMAN UNTUK MACHINE LEARNING DAN DEEP LEARNING*. *2*, 1–6.

Aulia, F. N. (2023). Sistem Pakar Dalam Menentukan Minat Dan Bakat Siswa Menggunakan Metode Forward Chaining. *Multidisciplinary Applications of Quantum Information Science (Al-Mantiq)*, *1*(2), 62-68q. https://doi.org/10.32665/almantiq.v1i2.1992

Basri, H., Yusuf, A. M., & Afdal, A. (2021). Kesesuaian Antara Bakat dan Minat dalam Menentukan Jurusan Pendidikan Tinggi Melalui Bimbingan Karir di Sekolah Menengah Atas. *SCHOULID: Indonesian Journal of School Counseling*, *6*(2), 157–163.

Fitri, Y., Mudjiran, M., & Refnywidialistuti, R. (2023). Peranan Bakat dan Minat dalam Belajar. *Jurnal Ilmiah Pendidikan Scholastic*, *7*(3), 62–67. https://doi.org/10.36057/jips.v7i3.637

Iceu Amira, H., & Rosidin, U. (2024). MENGEMBANGKAN MINAT DAN BAKAT PADA SISWA SMPN 3 JATINANGOR. *MENGEMBANGKAN MINAT DAN BAKAT PADA SISWA SMPN 3 JATINANGOR Iceu*, 2667–2676.

Kaswar, A. B., Andayani, D. D., & ... (2024). Pengenalan Artificial Intelligence untuk Pengembangan Minat dan Bakat Siswa SMA Negeri 11 Bulukumba. *Jurnal Sipakatau …*, *1*, 96–102.

Khairani, F., Kurnia, A., Aidi, M. N., & Pramana, S. (2022). Predictions of Indonesia Economic Phenomena Based on Online News Using Random Forest. *SinkrOn*, *7*(2), 532–540. https://doi.org/10.33395/sinkron.v7i2.11401

Pahlevi, O.-, Amrin, A.-, & Handrianto, Y.-. (2023). Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit. *Jurnal Infortech*, *5*(1), 71–76. https://doi.org/10.31294/infortech.v5i1.15829

Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *5*(1), 75–82. https://doi.org/10.31294/ijcit.v5i1.7951

SHELEMO, A. A. (2023). No Titleيايب. *Nucl. Phys.*, *13*(1), 104–116.

Wasilah, Azzaroh, I., Rahmawati, N., & Fitriana, A. Q. Z. (2023). Pengaruh Salah Pilih Jurusan Terhadap Nilai Akademik MahasiswaProgram Studi BKI UIN Khas Jember. *Jurnal Ilmu Sosial, Humaniora Dan Seni (JISHS)*, *1*(3), 615–619.