JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 3, Juni 2024, hlm. 459 - 466 ISSN 2550- 020 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v10i3.3067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

SUPPORT VECTOR MACHINE ANALYSIS FOR INTEREST AND TALENT CLASSIFICATION WITH PYTHON LIBRARY

Devi Sartika[1], Febie Elfaladonna[1*], Andre Mariza Putra[1]
[1]Informatics Management, Politeknik Negeri Sriwijaya
email: *febie_elfaladonna_mi@polsri.a.c.id

Abstract: Recognizing one's interests and talents early on is crucial in guiding an individual toward a prosperous future. While distinct, interests and talents share a close relationship. Interest denotes a genuine attraction to something without external pressure, and when consistently nurtured, it evolves into a skill or talent. Machine learning, specifically utilizing the SVM algorithm with the RBF kernel, can be applied to categorize interests and talents. Prior to SVM modeling, conducting Exploratory Data Analysis (EDA) is imperative for scrutinizing interests and talents. This analysis facilitates the identification of variables, enabling the elimination of missing values and ensuring the selection of appropriate interest and talent variables. The primary objective is to achieve optimal accuracy in modeling the classification of interests and talents. The insights gained from this research contribute to the creation of an application designed for categorizing interests and talents within SDN XYZ school. This application is designed for student use, assisting them in making informed decisions about their future education and career paths

Keywords: exploratory data analysis; interests and talents; machine learning; SVM Algorithm

Abstrak: Mengenali minat dan bakat seseorang sejak dini sangat penting dalam membimbing individu menuju masa depan yang sukses. Meskipun berbeda, minat dan bakat memiliki hubungan yang erat. Minat mengindikasikan ketertarikan yang tulus terhadap sesuatu tanpa tekanan eksternal, dan ketika terus-menerus dibina, berkembang menjadi keterampilan atau bakat. Pembelajaran mesin, khususnya dengan menggunakan algoritma SVM dan kernel RBF, dapat digunakan untuk mengelompokkan minat dan bakat. Sebelum pemodelan SVM, melakukan Analisis Data Eksploratif (EDA) sangat penting untuk mengkaji minat dan bakat. Analisis ini memfasilitasi identifikasi variabel, memungkinkan penghilangan nilai yang hilang, dan memastikan pemilihan variabel minat dan bakat yang tepat. Tujuan utamanya adalah mencapai akurasi optimal dalam pemodelan klasifikasi minat dan bakat. Temuan dari penelitian ini berkontribusi pada pengembangan aplikasi yang ditujukan untuk mengkategorikan minat dan bakat di sekolah SDN XYZ. Aplikasi ini dirancang untuk digunakan oleh siswa, membantu mereka membuat keputusan yang terinformasi mengenai pendidikan dan karier masa depan mereka.

Kata kunci: Algoritma SVM; exploratory data analysis; machine learning; minat dan bakat

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 3, Juni 2024, hlm. 459 - 466 ISSN 2550- 020 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v10i3.3067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

INTRODUCTION

Interests and aptitudes, which are important dimensions in individual development, are often managed by parents to further the future of their children, with education playing a key role in the process [1]. Interests, as a natural expression of interest, reflect self-acceptance of the surrounding environment, while aptitude involves abilities that require specialised training[2]. For example, someone who is interested in singing and takes vocal training demonstrates aptitude in that area.

Implementing a structured classification system for students' interests and talents at SDN XYZ in Palembang City through digitalization can significantly impact the school's progress. It replaces the class teacher's subjective selection method with an objective approach, potentially leading to more successful alumni in line with the school's goals.

Method selection is crucial in developing applications for interest and talent classification. Machine learning, particularly the Support Vector Machine (SVM) algorithm, is commonly chosen. Machine learning facilitates data classification into specific classes or targets by predicting future outcomes based on data, executed on computers. The learning process involves two stages: testing and training. Machine learning comprises three main categories: Supervised Learning, Unsupervised Learning, and Reinforcement Learning[3]. In Supervised Learning algorithms, the system receives a training dataset consisting of inputs and desired outputs. Through this data, the system gains understanding by discovering patterns. These patterns are then used as a guide to process the next dataset [4].

Meanwhile, reinforcement learning utilises methods that can operate in a dynamic environment and achieve goals without relying on direct instructions from a computer. This category often includes both supervised and unsupervised aspects simultaneously. The Support Vector Machine (SVM) algorithm is designed to find the maximal hyperplane, which is a function to separate two classes. In the process, SVM endeavours to increase the margin or distance between the training pattern and the decision boundary. SVMs have proven to be effective on datasets with many attributes and are relatively easy to implement[5]. The advantages of using SVM include superior performance for both small and large datasets, as well as for data that has many attributes [6].

Previous research on the SVM algorithm focuses on providing early childhood education tailored to children's interests and talents from home. One project involves talent mapping through various learning methods and educational games, helping parents feel more at ease. The goal is to develop an AI-based learning application (Hompimpa.id) that uses AI algorithms to identify and nurture students' interests and talents [7]. Another study found that scientific paper abstracts offer specific topic details, but the increasing number of journals makes analysis challenging. An automated system using machine learning can efficiently classify information from these abstracts, such as research problems, solutions, results, discussions, implications, and keywords, enhancing information retrieval[8].

Another study evaluated the SVM method for classifying SMAW weld quality, noting human control's unreliability. Using a quadratic function kernel, the SVM model achieved a 96.2% accuracy in testing[9]. Researchers found that the Gaussian RBF kernel achieved optimal classification accuracy. In the study, 337 training data and 77 out of 82 test data were correctly classified,

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 3, Juni 2024, hlm. 459 - 466 ISSN 2550- 020 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v10i3.3067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

resulting in a prediction accuracy of 93.9%[10].

The study concludes that Naïve Bayes is more effective than SVM in data classification, despite a small accuracy difference of 0.37%. Naïve Bayes achieves a Class Recall of 89.29%, compared to SVM's 92.86%, with an accuracy difference of 3.22%[11]. In Human Form recognition research, ANOVA and forward feature selection were used, with selected features input into SVM. The SVM's performance was consistent across linear, polynomial, and Gaussian RBF kernels. ANOVA was more effective than forward feature selection, confirming SVM's effectiveness as a classification tool[12]. Sentiment analysis of child-free opinions on Twitter shows that using SMOTE in Naive Bayes and SVM models enhances sensitivity for underrepresented classes. The SMOTE-optimized SVM is the best for predicting sentiment. Pro child-free opinions often cite parental unpreparedness, while anti child-free opinions are based on religious beliefs and concerns about old age care[13].

Research at PTIPD highlighted the use of manual processes with Microsoft Word and Excel for managing complaint data. This research aims to classify iRaise system problems into multiclass categories. Using RapidMiner with 1040 complaint records and 10-Fold cross-validation, the model achieved a 95.67% accuracy without feature selection, with parameters C=2 and y=0.09[14].
A study using SVM to forecast house prices in Bandar Lampung included data on price, location, and building details. Using 51 data points and 33 variables, regression and classification were tested with three kernels. The polynomial kernel in regression achieved the highest R^2 of 95.99%, while the linear and Gaussian kernels achieved R^2 of 90.99% and 81.43%. In classification, the Gaussian kernel with eight classes reached the highest accuracy of 91.18%, followed by the linear kernel at 90.20% and the polynomial kernel at 89.90%[15]. A study on classifying Down Syndrome, Autism, and ADHD used SVM with kernels for non-linear data. The highest average accuracy was 63.11% with a polynomial kernel (λ = 10, C = 1, itermax = 200). However, the accuracy was unsatisfactory compared to psychologists' classifications, likely due to data imbalance and the limited dataset [16].

The studies conclude that the SVM algorithm's average accuracy exceeds 80%. Originally for binary classification, SVM now handles multiple classes, regression, and outlier detection. In machine learning, Exploratory Data Analysis (EDA) enhances data understanding through steps like maximizing insights, designing data structures, extracting key variables, detecting anomalies, testing hypotheses, selecting models, and optimizing factors [17].

Python, created by Guido van Rossum in the Netherlands in 1990, has become the preferred programming language for many corporations and developers. Inspired by Monty Python's Flying Circus, Python's simplicity, concise syntax, and extensive library support have led to its widespread use in desktop, web, and mobile applications, making it popular in both industry and education[18]. Python libraries gather modules containing reusable code, streamlining development by eliminating the need to rewrite code for each program. This research uses several libraries provided by Python such as Pandas, NumPy, SciPy, Scikit-learn, and others.

METHOD

There are five stages in the research that are part of the machine

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 3, Juni 2024, hlm. 459 - 466 ISSN 2550- 020 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v10i3.3067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

learning process. The following is an explanation for each stage:

Data Identification and Preparation Stage

Data collection often involves data mining and web scraping. A logical and mathematical approach was used to survey 144 parents of SDN XYZ students, with 15 statements about children's preferences. Responses, considered ordinal data, were analyzed by school staff to classify interests and talents based on the BK teacher's assessment.

Data Acquisition

Derived from a Google Form questionnaire consisting of 15 statements, resulting in the classification of two classes, namely interest and talent. the 15 statements are: student's grade, age of student, critical thinking, likes trading and math, likes certain types of sports, likes to play musical instruments or sing, likes to cook, likes to write, read or tell stories, likes to draw, paint or color, often participate in coloring, drawing or painting competitions, often participate in coloring, drawing or painting competitions, often participate in storytelling, speech competitions. or poetry, often participates in singing or acting competitions, good at singing and playing music, proficient in drawing, coloring, and painting, and proficient in acting activities.

Questions are answered using a Likert scale, scored by the counseling teacher from 0 to 100 to reflect the child's interests or talents. Next, convert the survey data into a tabular format for analysis using Pandas, NumPy, and SciPy in Jupyter Notebook.

Exploratory Data Analysis

In this study, data exploration involved analyzing features of the dataset on primary school students' interests and aptitudes using chi-square and T-tests. Chi-square (x^2) assesses the significance of relationships between dependent and independent variables. The T-test measures the significance of differences between variables, helping to determine if the relationships are statistically acceptable.

```
from sklearn.feature_selection import chi2
x = df.drop(columns=['Prediksi'], axis=1)
y = df['Prediksi']

chi_score = chi2(x, y)
```

Figure 1. Run The Chi Square Test Sklearn Function

The results of the chi-square code execution as shown in the figure 2.

```
(array([ 0.99209243,  0.63224066,  0.69202329,  0.83476502,  0.43042221,
         9.09129968,  4.68226551,  0.29389977,  6.17494196,  6.24648434,
        15.46748504,  2.93558343,  4.31730334,  4.55054406, 17.79284731]),
```

Figure 2. Chi square results

In addition to producing a chi-square value which is an implementation of "chi2", the output also includes T-test results that show how significant the effect of the independent variable is on the dependent variable.

```
array([3.19231504e-01, 4.26534613e-01, 4.05477036e-01, 3.60898283e-01,
       5.11781849e-01, 2.56828298e-03, 3.04755318e-02, 5.87732233e-01,
       1.29572161e-02, 1.24440052e-02, 8.39368115e-05, 8.66473532e-02,
       3.77266641e-02, 3.29082881e-02, 2.46307450e-05]))
```

Figure 3. T-Test Results

In this study, a higher chi-square value indicates a stronger correlation between variables. Chi-square and T-tests were conducted twice, using the same dependent and independent variables. However, in the second round, only the independent variables selected from the first round were used.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 3, Juni 2024, hlm. 459 - 466 ISSN 2550- 020 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v10i3.3067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

```
from sklearn.preprocessing import LabelEncoder
for col in df.columns:
  le = LabelEncoder()
  df[col] = le.fit_transform(df[col])
  df.head()
```

```
from sklearn.feature_selection import chi2
x = df.drop(columns=['Prediksi'], axis=1)
y = df['Prediksi']

chi_score = chi2(x, y)
```

```
chi_score
```

```
(array([ 0.69202329,  0.83476502,  9.09129968,  4.68226551,  6.17494196,
         6.24648434, 15.46748504,  2.93558343,  4.31730334,  4.55054406,
        17.79284731]),
 array([4.05477036e-01, 3.60898283e-01, 2.56828298e-03, 3.04755318e-02,
        1.29572161e-02, 1.24440052e-02, 8.39368115e-05, 8.66473532e-02,
        3.77266641e-02, 3.29082881e-02, 2.46307450e-05]))
```

Figure 4. Chi Square and T-Test Second Results

Anomaly Identification

Anomaly identification recognizes and removes inappropriate data to maintain data quality. This data can be re-examined or deleted before further processing. Chi-square and T-Test results revealed that only nine of the many variables tested were suitable for further processing.

Recommended Variables

Based on Chi-square and T-Test results, two dependent variables ("prediction" labeled as interest and talent) and nine independent variables were identified as suitable. The nine variables are: "good at acting","participates in writing, reading, or storytelling activities","plays musical instruments and sings","participates in coloring, drawing, or painting competitions","draws, paints, or colors", "engages in cooking activities","good at drawing, painting, and coloring","good at singing while playing musical instruments," and "participates in singing and acting competitions".

RESULT AND DISCUSSION

Support Vector Machine with Python

After completing exploratory data analysis and identifying the recommended variables, the next step is to analyze the data using the Support Vector Machine (SVM) algorithm with Python. Four Python libraries were used: pandas, numpy, matplotlib, and seaborn. Matplotlib transforms raw data into valuable information through graphs and diagrams, while seaborn creates graphs and statistical visualizations.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 144 entries, 0 to 143
Data columns (total 12 columns):
 #   Column                                            Non-Null Count  Dtype
---  ------                                            --------------  -----
 0   Berpikir_Kritis                                   144 non-null    object
 1   Suka_berdagang_dan_berhitung                      144 non-null    object
 2   Suka_memainkan_alat_musik_atau_bernyanyi          144 non-null    object
 3   Suka_kegiatan_memasak                             144 non-null    object
 4   Suka_menggambar_atau_melukis_atau_mewarnai        144 non-null    object
 5   Sering_mengikuti_perlombaan_mewarnai_menggambar_melukis 144 non-null    object
 6   Suka_mengikuti_perlombaan_bercerita_berpidato_puisi 144 non-null    object
 7   Suka_mengikuti_perlombaan_menyanyi_akting         144 non-null    object
 8   Cakap_dalam_bernyanyi_bermain_musik               144 non-null    object
 9   Cakap_dalam_menggambar_melukis_mewarnai           144 non-null    object
 10  Cakap_dalam_kegiatan_akting                       144 non-null    object
 11  Prediksi                                          144 non-null    object
dtypes: object(12)
memory usage: 13.6+ KB
```

Figure 5. Data Frame

```
data_kol.shape
```

```
(144, 12)
```

```
data_kol.Prediksi.value_counts()
```

```
B     95
M     49
Name: Prediksi, dtype: int64
```

Figure 6. Collective Data Information

The figure above illustrates the number of rows in the dataset, reaching 144 data with 12 variables. In addition, this data is classified by aptitude prediction, with 95 people, and interest prediction, with 49 people.

Data Preparation

Some variables have an object data type, which were transformed using the label encoder function from the scikit-learn library to facilitate correlation observation. A strong linear pattern in the relationships indicates a correlation close to -1 or 1, while a non-linear relationship indicates a correlation close to 0 or an irregular pattern.

The dataset showed an imbalance in class distribution, which was addressed using the SMOTE (Synthetic Minority Over-Sampling Technique) oversampling technique. SMOTE generates new synthetic samples by finding the nearest neighbor of the minority sample and using it as a reference. After applying SMOTE, the class distribution became balanced, as shown in the following figure:

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. X No 3, Juni 2024, hlm. 459 - 466 ISSN 2550- 020 (Online)
DOI: http://dx.doi.org/10.33330/jurteksi.v10i3.3067
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

```
from imblearn.over_sampling import SMOTE

X = data_kol.iloc[:,1:-1]
y = data_kol.iloc[:, 11]
```

```
smote = SMOTE(random_state=0)
```

```
X_resampled, y_resampled = smote.fit_resample(X, y)
```

```
print(y_resampled.value_counts())
```

```
B    95
M    95
Name: Prediksi, dtype: int64
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X_resampled, y_resampled, test_size=0.20, random_state=0)
print("Nilai X_train:")
X_train
```

Figure 7. SMOTE process

In the early stages of modeling, the dataset was divided into 60% training data and 40% testing data. This larger proportion for training data ensures the model can be trained optimally, allowing it to identify complex patterns and relationships, thus enhancing its ability to make accurate predictions on unseen data.

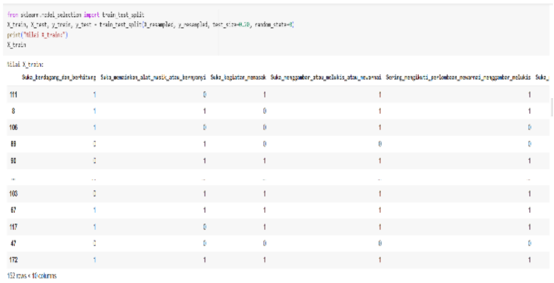

Figure 8. Training Data (x and y)

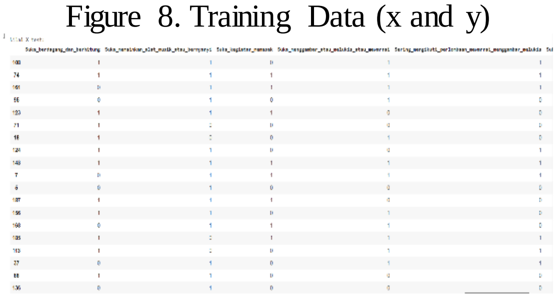

Figure 9. Data Testing (X_test dan y_test)

The dataset was split into training and testing sets using train_test_split from scikit-learn. X_train and X_test store features, while y_train and y_test store targets. An SVM model with an RBF kernel was used for its effectiveness with non-linear data. GridSearchCV was employed to optimize C and gamma values, testing C (0.1, 1, 10, 100, 1000) and gamma (1, 0.1, 0.01, 0.001, 0.0001).

```
Fitting 5 folds for each of 25 candidates, totalling 125 fits
[CV 1/5] END ........C=0.1, gamma=1, kernel=rbf;, score=0.613 total time=   0.0s
[CV 2/5] END ........C=0.1, gamma=1, kernel=rbf;, score=0.774 total time=   0.0s
[CV 3/5] END ........C=0.1, gamma=1, kernel=rbf;, score=0.567 total time=   0.0s
[CV 4/5] END ........C=0.1, gamma=1, kernel=rbf;, score=0.700 total time=   0.0s
[CV 5/5] END ........C=0.1, gamma=1, kernel=rbf;, score=0.700 total time=   0.0s
[CV 1/5] END ......C=0.1, gamma=0.1, kernel=rbf;, score=0.774 total time=   0.0s
[CV 2/5] END ......C=0.1, gamma=0.1, kernel=rbf;, score=0.742 total time=   0.0s
[CV 3/5] END ......C=0.1, gamma=0.1, kernel=rbf;, score=0.700 total time=   0.0s
[CV 4/5] END ......C=0.1, gamma=0.1, kernel=rbf;, score=0.833 total time=   0.0s
[CV 5/5] END ......C=0.1, gamma=0.1, kernel=rbf;, score=0.733 total time=   0.0s
[CV 1/5] END .....C=0.1, gamma=0.01, kernel=rbf;, score=0.516 total time=   0.0s
[CV 2/5] END .....C=0.1, gamma=0.01, kernel=rbf;, score=0.516 total time=   0.0s
[CV 3/5] END .....C=0.1, gamma=0.01, kernel=rbf;, score=0.500 total time=   0.0s
[CV 4/5] END .....C=0.1, gamma=0.01, kernel=rbf;, score=0.500 total time=   0.0s
[CV 5/5] END .....C=0.1, gamma=0.01, kernel=rbf;, score=0.500 total time=   0.0s
[CV 1/5] END ....C=0.1, gamma=0.001, kernel=rbf;, score=0.516 total time=   0.0s
```

Figure 10. Parameter Tuning on SVM

```
print(grid.best_params_)
print(grid.best_estimator_)
```

```
{'C': 1, 'gamma': 1, 'kernel': 'rbf'}
SVC(C=1, gamma=1)
```

Figure 11. Best Parameter Tuning

The next step is to evaluate the performance of the SVM model using Confusion Matrix. Confusion Matrix is a table used to illustrate the predictive quality of a model, by comparing the predicted results with the actual values of the test data.

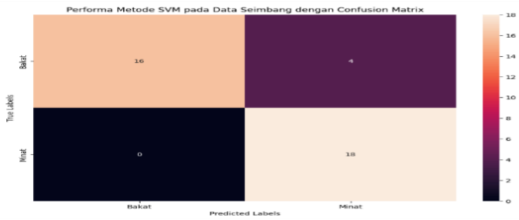

Figure 12. SVM Performance Test Using Confusion Matrix

The results of the confusion matrix in classifying interest and talent status with the Support Vector Machine method are illustrated in Figure 18. Furthermore, classification report is used to obtain a detailed report on the performance of the classification model in predicting interest and talent classes, as shown in the following figure.

```
Classification Report:
              precision    recall  f1-score   support

           B       1.00      0.80      0.89        20
           M       0.82      1.00      0.90        18

    accuracy                           0.89        38
   macro avg       0.91      0.90      0.89        38
weighted avg       0.91      0.89      0.89        38

Accuracy: 89.47368421052632
Sensitivity: 90.0
Specificity: 89.47368421052632
```

Gambar 13. Classification Report of SVM

The precision for the aptitude class is 1.00, indicating a 100% accuracy rate, while for the interest class, it's 0.82, representing an 82% accuracy rate. Recall for both classes shows that all positive samples were correctly identified, with rates of 80% for aptitude and 100% for interest. High recall suggests the model effectively recognizes positive samples. F1-score values were high for all classes, indicating a balanced performance between precision and recall, and a strong ability to make accurate positive predictions while recognizing all positive samples.

Evaluation

The cross-validation method assesses a model's performance on unseen data. Evaluation of the interest and talent model using cross-validation yielded an accuracy of 96.66%. These results affirm SVM's suitability and its potential in developing applications for classifying students' interests and talents at SDN XYZ Palembang, aiding decision-making processes.

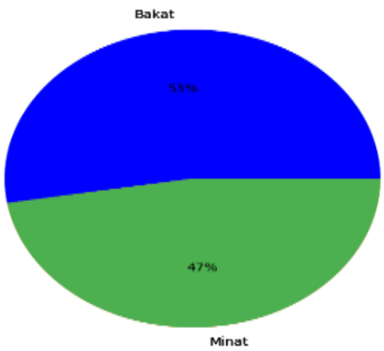

Figure 14. Pie Chart of Interest and Talent Classification

Based on the implementation of Exploratory Data Analysis (EDA) and Support Vector Machine (SVM) algorithm on 144 questionnaire datasets from SDN 204 Palembang students, SVM proves highly suitable for interest and talent classification, achieving a 73% accuracy rate. The pie chart illustrates that aptitude accounts for 53% of the datasets, while interest represents 47%. This suggests that schools can efficiently leverage student competencies based on interests and aptitudes, aligning with the goal of nurturing competent graduates.

CONCLUSION

Utilizing Exploratory Data Analysis (EDA) and the Support Vector Ma-chine (SVM) algorithm on 144 questionnaire datasets for SDN XYZ Palem-bang students, it is concluded that SVM is highly suitable for interest and talent classification with a 73% accuracy rate. From the pie charts, it's evident that among the 144 datasets, aptitude ac-counts for 53%, while interest is at 47%. This facilitates efficient management of student competence in SDN XYZ Pa-lembang based on interests and talents, aligning with the school's goal of producing competent graduates.

BIBLIOGRAPHY

[1] I. A. Anggraini, W. D. Utami, and S. B. Rahma, “Mengidentifikasi Minat Bakat Siswa Sejak Usia Dini di SD Adiwiyata,” *ISLAMIKA*, vol. 2, no. 1, 2020, doi: 10.36088/islamika.v2i1.570.

[2] Y. Fitriyah Ningsih, N. Hariadi, and D. A. Puspitaningrum, “Hubungan Antara Minat dan Bakat Mahasiswa Universitas Jember Kampus Bondowoso Terhadap Fasilitas Olahraga,” *Jurnal Porkes*, vol. 2, no. 2, 2019, doi: 10.29408/porkes.v2i2.1643.

[3] A. Roihan, P. A. Sunarya, and A. S. Rafika, “Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper,” *IJCIT (Indonesian Journal on Computer and Information Technology)*, vol. 5, no. 1, 2020, doi: 10.31294/ijcit.v5i1.7951.

[4] H. Abijono, P. Santoso, and N. L. Anggreini, “Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data,” *Jurnal Teknologi Terapan: G-Tech*, vol. 4, no. 2, 2021, doi: 10.33379/gtech.v4i2.635.

[5] F. Abdusyukur, "Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter," *Komputa : Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, 2023, doi: 10.34010/komputa.v12i1.9418.

[6] Oryza Habibie Rahman, Gunawan Abdillah, and Agus Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, 2021, doi: 10.29207/resti.v5i1.2700.

[7] F. N. Afiana *et al.*, "Aplikasi Pembelajaran Anak Usia Dini Untuk Menentukan Minat Bakat Dengan Teknologi AI," *Indonesian Journal on Software Engineering (IJSE)*, vol. 7, no. 2, 2021.

[8] T. E. Puspitawati, S. Basuki, V. Rahmayanti, and S. Nastiti, "Algoritma Support Vector Machine (SVM) Untuk Identifikasi Komponen Abstrak Pada Jurnal Ilmiah Berbasis Teknik Klasifikasi," *REPOSITOR*, vol. 3, no. 5, 2021.

[9] A. S. Ritonga and E. S. Purwaningsih, "Penerapan Metode Support Vector Machine (SVM) Dalam Klasifikasi Kualitas Pengelasan Smaw (Shield Metal Arc Welding)," *Ilmiah Edutic*, vol. 5, no. 1, 2018.

[10] P. A. Octaviani, Yuciana Wilandari, and D. Ispriyanti, "Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang," *Jurnal Gaussian*, vol. 3, no. 8, 2014.

[11] H. W. Dhany and F. Izhari, "Analisis Algorithms Support Vector Machine Dengan Naive Bayes Kernel Pada Klasifikasi Data," *Jurnal Teknik Dan Informatika* , vol. 6, no. 2, 2019.

[12] A. Wenda, "Support Vector Machine Untuk Pengenalan Bentuk Manusia Menggunakan Kumpulan Fitur Yang Dioptimalkan," *JST (Jurnal Sains dan Teknologi)*, vol. 11, no. 1, 2022, doi: 10.23887/jstundiksha.v11i1.44437.

[13] Dania Siregar, Faroh Ladayya, Naufal Zhafran Albaqi, and Bintang Mahesa Wardana, "Penerapan Metode Support Vector Machines (SVM) dan Metode Naïve Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter," *Jurnal Statistika dan Aplikasinya*, vol. 7, no. 1, 2023, doi: 10.21009/jsa.07109.

[14] Fatmawati and A. Muhammad, "Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) (Studi Kasus : Akun Facebook Group iRaise Helpdesk)," *Jurnal CoreIT*, vol. 3, no. 1, 2017.

[15] F. R. Lumbanraja, R. A. Saputra, K. Muludi, A. Hijriani, and A. Junaidi, "Implementasi Support Vector Machine Dalam Memprediksi Harga Rumah Pada Perumahan Di Kota Bandar Lampung," *Jurnal Pepadun*, vol. 2, no. 3, 2021, doi: 10.23960/pepadun.v2i3.90.

[16] I. P. Monika and M. T. Furqon, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 10, 2018.

[17] M. Radhi, A. Amalia, D. R. H. Sitompul, S. H. Sinurat, and E. Indra, "Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook," *Jurnal Sistem Informasi dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, vol. 4, no. 2, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2475.

[18] Muhammad Romzi and B. Kurniawan, "Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma," *JTIM: Jurnal Teknik Informatika Mahakarya*, vol. 03, no. 2, 2020.