

Pemodelan Prediktif Kinerja Ujian Siswa Berdasarkan Kebiasaan Belajar Menggunakan Pembelajaran Mesin dan Analisis SHAP

Predictive Modeling of Student Exam Performance Based on Learning Habits Using Machine Learning and SHAP Analysis

Ayu Lestari Azis¹; Irmawati²; Firman Aziz^{3,*}

^{1,2} IRMEX Digital Akademi, Makassar 90155, Indonesia

³ Universitas Pancasakti, Makassar 90121, Indonesia

¹ ayulestariazis@irmexdigikal.com; ² irmawati@irmexdigikal.com; ³ firman.aziz@unpacti.ac.id

* Corresponding author

Abstrak

Penelitian ini bertujuan untuk menganalisis dampak penggunaan Artificial Intelligence (AI) terhadap performa akademik mahasiswa menggunakan pendekatan machine learning dengan model Logistic Regression untuk klasifikasi multikelas. Dataset yang digunakan terdiri dari 1000 observasi yang diperoleh dari platform Kaggle dan mencakup variabel terkait intensitas penggunaan AI serta indikator akademik mahasiswa. Tahapan penelitian meliputi preprocessing data, pembagian data 80:20, pelatihan model Logistic Regression, serta evaluasi menggunakan accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model mampu melakukan klasifikasi dengan performa yang stabil pada seluruh kelas. Analisis interpretatif menggunakan SHAP digunakan untuk memahami kontribusi masing-masing variabel terhadap hasil prediksi. Temuan menunjukkan bahwa beberapa faktor penggunaan AI memiliki pengaruh berbeda terhadap kategori performa akademik mahasiswa, dengan pola hubungan yang dapat dijelaskan secara transparan melalui pendekatan Explainable Artificial Intelligence (XAI). Penelitian ini bersifat evaluatif dan interpretatif, serta tidak mengklaim generalisasi di luar konteks dataset yang digunakan. Hasil penelitian diharapkan dapat menjadi dasar pengembangan analisis lanjutan dengan validasi tambahan dan pendekatan model yang lebih kompleks.

Kata Kunci: Machine Learning; Performa Akademik; Kebiasaan Belajar; Explainable AI (SHAP); Educational Data Mining

Abstract

This study aims to analyze the impact of Artificial Intelligence (AI) use on student academic performance using a machine learning approach with a Logistic Regression model for multiclass classification. The dataset used consists of 1000 observations obtained from the Kaggle platform and includes variables related to the intensity of AI use and student academic indicators. The research stages include data preprocessing, 80:20 data splitting, training the Logistic Regression model, and evaluation using accuracy, precision, recall, and F1-score. The results show that the model is capable of classifying with stable performance across classes. Interpretive analysis using SHAP was used to understand the contribution of each variable to the prediction results. The findings indicate that several factors of AI use have different influences on student academic performance categories, with relationship patterns that can be explained transparently using an Explainable Artificial Intelligence (XAI) approach. This research is evaluative and interpretative in nature, and does not claim generalizations beyond the context of the dataset used. The results are expected to form the basis for developing further analysis with additional validation and more complex modeling approaches.

Keywords: Machine Learning; Academic Performance; Learning Habits; Explainable AI (SHAP); Educational Data Mining

Pendahuluan

Performa akademik mahasiswa merupakan indikator utama dalam mengevaluasi keberhasilan sistem pendidikan tinggi. Berbagai faktor diketahui memengaruhi capaian akademik, termasuk kebiasaan belajar, durasi tidur, serta keterlibatan dalam pembelajaran tambahan. Studi terdahulu menunjukkan bahwa kebiasaan belajar memiliki hubungan signifikan terhadap prestasi akademik mahasiswa. Misalnya, penelitian oleh Credé and Kuncel menunjukkan bahwa manajemen waktu dan strategi belajar berkontribusi secara signifikan terhadap peningkatan

performa akademik [1]. Hal ini diperkuat oleh Nonis and Hudson yang menemukan bahwa jumlah waktu belajar memiliki hubungan positif terhadap pencapaian akademik, meskipun dipengaruhi oleh faktor lain seperti motivasi [2].

Selain kebiasaan belajar, faktor fisiologis seperti tidur juga berperan penting dalam mendukung fungsi kognitif. Penelitian oleh Matthew Walker menunjukkan bahwa tidur berperan dalam konsolidasi memori dan peningkatan kemampuan belajar [3]. Studi lain oleh Curcio et al. mengungkapkan bahwa kurang tidur dapat menurunkan konsentrasi dan performa akademik mahasiswa [4].

Seiring perkembangan teknologi, pembelajaran berbasis daring juga menjadi faktor penting dalam proses pendidikan. Means et al. menyatakan bahwa pembelajaran online memiliki potensi meningkatkan hasil belajar jika dimanfaatkan secara optimal [5]. Namun demikian, kontribusinya terhadap performa akademik masih menunjukkan hasil yang bervariasi tergantung tingkat keterlibatan mahasiswa.

Dalam beberapa tahun terakhir, pendekatan machine learning telah banyak digunakan untuk memprediksi performa akademik karena kemampuannya dalam menangkap hubungan kompleks dan non-linear antar variabel. Studi oleh Romero and Ventura menunjukkan bahwa Educational Data Mining mampu memberikan wawasan yang lebih mendalam terkait pola belajar mahasiswa [6], [7]. Penelitian lain juga menegaskan bahwa algoritma seperti Random Forest dan Support Vector Machine mampu meningkatkan akurasi prediksi dibandingkan metode statistik konvensional [8], [9].

Namun, salah satu kelemahan utama dari model machine learning adalah kurangnya interpretabilitas. Untuk mengatasi hal tersebut, pendekatan Explainable Artificial Intelligence (XAI) dikembangkan. Metode SHAP yang diperkenalkan oleh Lundberg and Lee memungkinkan interpretasi kontribusi masing-masing fitur secara konsisten [10]. Penggunaan SHAP dalam analisis pendidikan memberikan transparansi terhadap faktor-faktor yang memengaruhi performa akademik [11].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi performa akademik mahasiswa berbasis machine learning serta menginterpretasikan hasilnya menggunakan pendekatan SHAP. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem analitik pendidikan yang lebih akurat dan interpretatif.

Metode

A. Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental berbasis machine learning untuk memodelkan hubungan antara kebiasaan belajar mahasiswa dan performa akademik. Pendekatan ini dipilih karena mampu menangkap hubungan kompleks dan non-linear antar variabel yang tidak dapat dijelaskan secara optimal oleh metode statistik konvensional [6], [7], [12].

B. Dataset dan Variabel

Dataset yang digunakan berasal dari platform Kaggle yang berisi data perilaku belajar mahasiswa dan performa ujian. Dataset terdiri dari 1000 observasi dengan empat variabel utama, yaitu:

- `daily_study_hours`: jumlah jam belajar per hari
- `online_courses_taken`: jumlah kursus online yang diikuti
- `avg_sleep_hours`: rata-rata jam tidur per hari
- `exam_score`: nilai ujian sebagai variabel target

Variabel independen dalam penelitian ini adalah kebiasaan belajar dan pola tidur, sedangkan variabel dependen adalah performa akademik mahasiswa.

C. Dataset dan Variabel

Tahap preprocessing dilakukan untuk memastikan kualitas data sebelum digunakan dalam proses pemodelan. Langkah-langkah yang dilakukan meliputi:

1. Penghapusan atribut tidak relevan

Variabel `student_id` dihapus karena tidak memiliki kontribusi terhadap proses prediksi.

2. **Normalisasi data**

Data dinormalisasi menggunakan StandardScaler untuk menghindari bias akibat perbedaan skala antar variabel [35].

3. **Pembagian data**

Dataset dibagi menjadi:

- 80% data latih (training set)
- 20% data uji (testing set)

Pembagian ini bertujuan untuk menghindari overfitting dan memastikan generalisasi model [13].

Analisis eksploratif dilakukan untuk memahami karakteristik data dan hubungan antar variabel. Teknik yang digunakan meliputi:

- Analisis korelasi Pearson
- Visualisasi distribusi data
- Scatter plot antar variabel

EDA bertujuan untuk mengidentifikasi pola awal serta potensi hubungan linear maupun non-linear antar variabel [14].

D. Model Machine Learning

Penelitian ini menggunakan beberapa algoritma machine learning untuk membandingkan performa prediksi, yaitu:

1. **Linear Regression**

Digunakan sebagai baseline model untuk memodelkan hubungan linear antar variabel [15].

2. **Random Forest**

Algoritma ensemble berbasis decision tree yang mampu menangkap hubungan non-linear dan mengurangi overfitting [16].

3. **Support Vector Regression (SVR)**

Model berbasis kernel yang efektif dalam memodelkan data dengan dimensi tinggi [17].

4. **XGBoost (Extreme Gradient Boosting)**

Algoritma boosting yang dikenal memiliki performa tinggi dalam berbagai kasus prediksi [18].

Penggunaan beberapa model ini bertujuan untuk menentukan algoritma terbaik dalam memprediksi performa akademik mahasiswa.

E. Explainable Artificial Intelligence (XAI)

Untuk meningkatkan interpretabilitas model, penelitian ini menggunakan pendekatan Explainable Artificial Intelligence (XAI) dengan metode SHAP (SHapley Additive exPlanations).

Metode SHAP digunakan untuk:

- Mengidentifikasi kontribusi masing-masing variabel
- Menentukan fitur paling dominan
- Memberikan interpretasi model secara transparan

Pendekatan ini penting karena model machine learning sering dianggap sebagai “black box”.

F. Explainable Artificial Intelligence (XAI)

Evaluasi model dilakukan menggunakan beberapa metrik untuk mengukur akurasi prediksi, yaitu:

- **Root Mean Square Error (RMSE):** Mengukur rata-rata kesalahan prediksi dalam satuan yang sama dengan data.
- **Mean Absolute Error (MAE):** Mengukur rata-rata absolut kesalahan prediksi.
- **Koefisien Determinasi (R²):** Mengukur kemampuan model dalam menjelaskan variabilitas data.

Semakin kecil nilai RMSE dan MAE serta semakin besar nilai R^2 , maka semakin baik performa model [19].

Hasil dan Diskusi

Hasil implementasi model machine learning menunjukkan adanya perbedaan performa yang signifikan antar algoritma yang digunakan. Evaluasi dilakukan menggunakan metrik Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R^2), sebagaimana disajikan pada Tabel 1.

Tabel 1. Performa Model Machine Learning

<i>Model</i>	<i>RMSE</i>	<i>MAE</i>	<i>R²</i>
Linear Regression	8.45	6.72	0.71
Random Forest	5.12	3.98	0.88
SVR	6.37	4.85	0.82
XGBoost	4.76	3.65	0.90

Berdasarkan Tabel 1, model XGBoost menunjukkan performa terbaik dengan nilai RMSE terendah dan R^2 tertinggi, yang mengindikasikan bahwa model ini memiliki kemampuan yang sangat baik dalam memprediksi nilai ujian mahasiswa. Performa tinggi yang ditunjukkan oleh model berbasis boosting ini menguatkan bahwa hubungan antara variabel kebiasaan belajar dan performa akademik bersifat kompleks dan non-linear. Sebaliknya, model Linear Regression menunjukkan performa paling rendah, yang mengindikasikan keterbatasannya dalam menangkap pola hubungan yang tidak linear.

Analisis korelasi yang divisualisasikan melalui correlation matrix menunjukkan bahwa variabel `daily_study_hours` memiliki hubungan positif yang paling kuat terhadap `exam_score`. Hal ini menegaskan bahwa intensitas belajar merupakan faktor utama dalam menentukan capaian akademik mahasiswa. Variabel `avg_sleep_hours` menunjukkan korelasi moderat, yang mengindikasikan bahwa faktor fisiologis juga memiliki kontribusi penting, terutama dalam mendukung fungsi kognitif. Sementara itu, variabel `online_courses_taken` memiliki korelasi yang relatif lebih rendah, yang menunjukkan bahwa kontribusinya bersifat pelengkap.

Temuan tersebut semakin diperkuat oleh hasil analisis feature importance dari model Random Forest, yang menunjukkan bahwa variabel `daily_study_hours` memiliki kontribusi paling dominan, diikuti oleh `avg_sleep_hours` dan `online_courses_taken`. Dominasi variabel jam belajar menunjukkan bahwa peningkatan durasi belajar secara konsisten berperan besar dalam meningkatkan performa akademik. Namun demikian, keberadaan variabel durasi tidur sebagai faktor kedua menunjukkan bahwa keseimbangan antara aktivitas belajar dan istirahat menjadi aspek penting yang tidak dapat diabaikan.

Analisis lebih mendalam menggunakan metode SHAP memberikan interpretasi yang lebih komprehensif terhadap model. Hasil analisis menunjukkan bahwa `daily_study_hours` memiliki kontribusi paling besar terhadap prediksi nilai ujian, dengan nilai SHAP yang meningkat seiring dengan bertambahnya jam belajar. Selain itu, `avg_sleep_hours` menunjukkan pola hubungan yang tidak sepenuhnya linear, yang mengindikasikan adanya rentang durasi tidur optimal dalam mendukung performa akademik. Durasi tidur yang terlalu rendah maupun terlalu tinggi cenderung memberikan dampak yang kurang optimal terhadap hasil belajar. Sementara itu, `online_courses_taken` memberikan kontribusi positif terhadap model, namun dengan tingkat pengaruh yang relatif lebih kecil dibandingkan variabel lainnya.

Berdasarkan keseluruhan hasil yang diperoleh, terdapat beberapa temuan penting yang dapat diidentifikasi. Pertama, intensitas belajar yang direpresentasikan oleh `daily_study_hours` merupakan faktor paling dominan dalam menentukan performa akademik mahasiswa. Kedua, faktor fisiologis seperti durasi tidur memiliki pengaruh signifikan dan menunjukkan hubungan yang tidak linear, yang menandakan adanya kondisi optimal dalam mendukung fungsi kognitif. Ketiga, model machine learning berbasis ensemble seperti XGBoost dan Random Forest terbukti lebih efektif dibandingkan model linier dalam menangkap pola hubungan kompleks antar variabel. Keempat, penggunaan metode Explainable Artificial Intelligence (XAI) melalui SHAP memberikan nilai tambah dalam hal interpretasi model, sehingga hasil prediksi dapat dijelaskan secara transparan.

Meskipun penelitian ini memberikan hasil yang signifikan, terdapat beberapa keterbatasan yang perlu diperhatikan. Dataset yang digunakan masih terbatas pada variabel kebiasaan belajar dasar dan belum mencakup faktor psikologis seperti motivasi, tingkat stres, serta gaya belajar mahasiswa yang juga berpotensi memengaruhi

performa akademik. Selain itu, dataset yang digunakan bersifat generik dan tidak berasal dari institusi pendidikan tertentu, sehingga generalisasi hasil penelitian masih perlu diuji lebih lanjut pada konteks yang lebih spesifik. Dari sisi metodologi, penelitian ini belum mengintegrasikan pendekatan deep learning atau teknik optimasi lanjutan yang berpotensi meningkatkan performa model secara lebih signifikan.

Oleh karena itu, penelitian selanjutnya disarankan untuk mengintegrasikan variabel tambahan yang lebih kompleks, seperti faktor psikologis dan lingkungan belajar, serta menggunakan dataset yang lebih besar dan beragam. Selain itu, penerapan model deep learning dan teknik optimasi seperti hyperparameter tuning secara lebih mendalam dapat dilakukan untuk meningkatkan akurasi prediksi. Integrasi data multimodal, seperti data perilaku digital atau aktivitas pembelajaran berbasis platform online, juga dapat menjadi arah penelitian yang menjanjikan untuk menghasilkan model prediktif yang lebih komprehensif dan robust.

Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi performa akademik mahasiswa berbasis pendekatan machine learning dengan mempertimbangkan variabel kebiasaan belajar dan faktor pendukung lainnya. Hasil penelitian menunjukkan bahwa model berbasis ensemble, khususnya XGBoost dan Random Forest, memiliki performa terbaik dibandingkan model lainnya, yang ditunjukkan oleh nilai RMSE yang rendah dan koefisien determinasi (R^2) yang tinggi. Hal ini mengindikasikan bahwa pendekatan machine learning mampu menangkap hubungan kompleks dan non-linear antar variabel secara lebih efektif dibandingkan metode statistik konvensional. Analisis lebih lanjut mengungkap bahwa variabel `daily_study_hours` merupakan faktor paling dominan dalam menentukan performa akademik mahasiswa. Temuan ini menegaskan bahwa intensitas belajar memiliki pengaruh langsung dan signifikan terhadap pencapaian nilai ujian. Selain itu, variabel `avg_sleep_hours` juga menunjukkan kontribusi yang penting dengan pola hubungan yang tidak linear, yang mengindikasikan adanya durasi tidur optimal dalam mendukung fungsi kognitif dan proses pembelajaran. Sementara itu, variabel `online_courses_taken` memberikan kontribusi tambahan, namun dengan pengaruh yang relatif lebih kecil dibandingkan variabel lainnya. Penggunaan metode Explainable Artificial Intelligence (XAI) melalui SHAP memberikan nilai tambah yang signifikan dalam penelitian ini, karena mampu menjelaskan kontribusi masing-masing variabel secara transparan dan interpretatif. Dengan demikian, model yang dihasilkan tidak hanya memiliki tingkat akurasi yang tinggi, tetapi juga dapat digunakan sebagai dasar pengambilan keputusan yang lebih informatif dalam konteks pendidikan. Secara keseluruhan, penelitian ini memberikan kontribusi dalam pengembangan bidang Educational Data Mining dengan menghadirkan pendekatan integratif antara machine learning dan interpretabilitas model. Temuan yang dihasilkan diharapkan dapat menjadi dasar bagi institusi pendidikan dalam merancang strategi pembelajaran yang lebih efektif, khususnya dalam mendorong peningkatan kebiasaan belajar dan menjaga keseimbangan antara aktivitas akademik dan kesehatan mahasiswa. Untuk pengembangan lebih lanjut, disarankan agar penelitian selanjutnya mengintegrasikan variabel yang lebih kompleks, seperti faktor psikologis dan lingkungan belajar, serta menggunakan pendekatan model yang lebih canggih guna meningkatkan akurasi dan generalisasi hasil penelitian.

Deklarasi Kepentingan

Seluruh penulis menyatakan bahwa penelitian ini dilaksanakan tanpa adanya kepentingan pribadi, profesional, maupun finansial yang dapat memengaruhi hasil penelitian atau proses publikasi artikel.

Ucapan Terima Kasih

Penulis ingin mengucapkan terima kasih kepada seluruh pihak yang terlibat dalam penelitian ini. Pelaksanaan penelitian dilakukan secara mandiri dan tidak memperoleh dukungan pendanaan khusus dari institusi pemerintah, sektor industri, maupun lembaga pendanaan penelitian lainnya.

Ketersediaan Data

Data yang digunakan dalam penelitian ini bersumber dari dataset publik yang tersedia pada platform Kaggle. Dataset tersebut dapat diakses kembali sesuai ketentuan penyedia data atau melalui penulis korespondensi untuk kebutuhan akademik yang relevan.

Penggunaan Ai Dan Deklarasi Penggunaan Ai Generatif

Dalam proses penyusunan artikel, penulis memanfaatkan teknologi AI generatif sebagai alat bantu pada tahap pengembangan draft awal, penyempurnaan struktur penulisan, dan pemeriksaan kebahasaan. Seluruh substansi ilmiah, hasil analisis, interpretasi data, serta isi akhir manuskrip telah diperiksa secara menyeluruh dan sepenuhnya menjadi tanggung jawab penulis.

Daftar Pustaka

- [1] M. Credé and N. R. Kuncel, "Study Habits, Skills, and Attitudes: The Third Pillar Supporting Collegiate Academic Performance," *Perspect. Psychol. Sci.*, vol. 3, no. 6, pp. 425–453, 2008, doi: 10.1111/J.1745-6924.2008.00089.X;ISSUE:ISSUE:DOI.
- [2] S. A. Nonis and G. I. Hudson, "Performance of College Students: Impact of Study Time and Study Habits," *J. Educ. Bus.*, vol. 85, no. 4, pp. 229–238, Mar. 2010, doi: 10.1080/08832320903449550.
- [3] "Why We Sleep: Unlocking the Power of Sleep and Dreams - Matthew Walker - Google Books." https://books.google.co.id/books?hl=en&lr=&id=ZIU3DwAAQBAJ&oi=fnd&pg=PA3&dq=M.+Walker,+Why+We+Sleep,+Scribner,+2017&ots=MXRdbH0YnY&sig=WdtAWy80DM_119o6X6wTnCNg6N0&redir_esc=y#v=onepage&q=M.+Walker%2C+Why+We+Sleep%2C+Scribner%2C+2017&f=false (accessed Apr. 20, 2026).
- [4] G. Curcio, M. Ferrara, and L. De Gennaro, "Sleep loss, learning capacity and academic performance," *Sleep Med. Rev.*, vol. 10, no. 5, pp. 323–337, Oct. 2006, doi: 10.1016/J.SMRV.2005.11.001.
- [5] B. Means, Y. Toyama, R. Murphy, and M. Baki, "The effectiveness of online and blended learning: A meta-analysis of the empirical literature," *Teach. Coll. Rec.*, vol. 115, no. 3, 2013, doi: 10.1177/016146811311500307;PAGEGROUP:STRING:PUBLICATION.
- [6] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 6, pp. 601–618, Nov. 2010, doi: 10.1109/TSMCC.2010.2053532.
- [7] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 135–146, Jul. 2007, doi: 10.1016/J.ESWA.2006.04.005.
- [8] B. Albreiki, N. Zaki, and H. Alashwal, "A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques," *Educ. Sci. 2021, Vol. 11, Page 552*, vol. 11, no. 9, p. 552, Sep. 2021, doi: 10.3390/EDUCSCII1090552.
- [9] E. Ahmed and A. Al-Omari, "Student Performance Prediction Using Machine Learning Algorithms," *Appl. Comput. Intell. Soft Comput.*, vol. 2024, no. 1, p. 4067721, Jan. 2024, doi: 10.1155/2024/4067721.
- [10] S. M. Lundberg, P. G. Allen, and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, Accessed: Feb. 05, 2026. [Online]. Available: <https://github.com/slundberg/shap>.
- [11] C. Molnar, "Interpretable Machine Learning - Google Books." https://books.google.co.id/books?hl=en&lr=&id=jBm3DwAAQBAJ&oi=fnd&pg=PP1&dq=R.+Molnar,+Interpretable+Machine+Learning,+2020.&ots=EhyV-nKCT0&sig=ZDLtxUO08WWJKDi6oDt6cao0KM&redir_esc=y#v=onepage&q&f=false (accessed Apr. 20, 2026).
- [12] A. Cutler, D. R. Cutler, and J. R. Stevens, "Random Forests," *Ensemble Mach. Learn.*, pp. 157–175, 2012, doi: 10.1007/978-1-4419-9326-7_5.
- [13] "Machine Learning: A Probabilistic Perspective - Kevin P. Murphy - Google Books." https://books.google.co.id/books?hl=en&lr=&id=RC43AgAAQBAJ&oi=fnd&pg=PR7&dq=K.+Murphy,+Machine+Learning,+MIT+Press,+2012.&ots=unmAfBTv1d&sig=SG6jIABIQMYdxQhFX5-rCFKTFx0&redir_esc=y#v=onepage&q&f=false (accessed Apr. 20, 2026).
- [14] M. Chen, J. Han, P. Y.-I. T. on Knowledge, and undefined 1996, "Data mining: an overview from a database perspective," *ieeexplore.ieee.org*, Accessed: Jun. 15, 2022. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/553155/>.
- [15] D. H. Jr, S. Lemeshow, and R. Sturdivant, *Applied logistic regression*. 2013.
- [16] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 4, Jul. 2018, doi: 10.1002/WIDM.1249.
- [17] E. Byvatov, G. S.-A. bioinformatics, and undefined 2003, "Support vector machine applications in bioinformatics.," *europemc.org*, Accessed: Aug. 06, 2022. [Online]. Available:

<https://europepmc.org/article/med/15130823>.

- [18] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-August-2016, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785;CSUBTYPE:STRING:CONFERENCE.
- [19] A. Salih, A. A.-J. of S. C. and, and undefined 2021, “Evaluation of classification algorithms for intrusion detection system: A review,” *publisher.uthm.edu.my*, vol. 2, no. 1, pp. 31–40, 2021, doi: 10.30880/jscdm.2021.02.01.004.