
Prediksi *Student Performance* Pada Hasil Penilaian Proses Pembelajaran Online Mata Pelajaran Informatika Di SMA

Sasra Dipa¹, Joan Santoso², Francisca H. Chandra³

^{1,2,3} Pasca Sarjana, ISTTS, Jl. Ngagel Jaya Tengah No.73-77, Surabaya, Jawa Timur, 60284, Indonesia

***Email Korespondensi :**
Sasra_d20@mhs.istts.ac.id

Abstrak

Di Endemi Corona, kita tidak sekedar kembali ke pola pendidikan offline namun sudah mengarah pada edukasi 5.0. Pola pembelajaran online, normal, blended sudah menjadi hal yang biasa. Penilaian pembelajaran Online membutuhkan prediksi student performance yang cepat dan tepat (akurasi tinggi). Penyebabnya adalah pertama, karena keterbatasan interaksi langsung. Kedua, pembelajaran normal biasanya ada penilaian proses belajar dan penilaian karakter untuk bisa memberikan penilaian akhir akurat itu sulit dapat dilaksanakan pada pembelajaran Online secara akurat. Ketiga, banyaknya data untuk diolah cepat dan tepat agar bisa dilaporkan kepada Institusi pendidikan dan pada keluarga peserta didik. Keempat, Informatika adalah pelajaran yang 80 % praktek dan 20 % teori sehingga instrument penilaian yang digunakan adalah 80 % instrument unjuk kerja (taksonomi bloom : C2, C3, C4, C5) dan 20 % instrument sejenis pilihan ganda (C1). Koreksi dan penilaian Informatika lebih membutuhkan banyak waktu karena 80 % tidak bisa dinilai secara otomatis. Penelitian ini bertujuan memprediksi student performance (Lulus (1) atau Intervensi (0)) pada hasil penilaian proses pembelajaran online mata pelajaran informatika di sma. Apabila hasil prediksi student performance menghasilkan Intervensi maka segera ditindaklanjuti dengan memberikan strategi intervensi supaya terjadi peningkatan student performance. Target hasil penelitian mencapai akurasi > 70 % terhadap dataset yang diolah. Penelitian ini menggunakan metode ensemble learning random Forest Classification dan XG Boosting classification. Hasil penelitian Prediksi Student Performance menggunakan XG Boost Classification menghasilkan akurasi lebih tinggi daripada RF Classification yang memiliki nilai akurasi rata-rata = 93 % sedangkan RF Classification memiliki hasil akurasi rata-rata = 92 %. Tujuan penelitian sudah tercapai karena hasil 2 metode yang digunakan sudah sesuai target yang diinginkan.

Kata Kunci : Informatika; Random Forest; Student Performance; SMA; XG Boosting

Abstract

In the Corona Endemic, we are not just returning to offline education patterns but are already moving towards education 5.0. Online, normal, blended learning patterns have become commonplace. Online learning assessment requires fast and precise predictions of student performance (high accuracy). The reason is first, due to limited direct interaction. Second, normal learning usually involves an assessment of the learning process and character assessment to be able to provide an accurate final assessment, which is difficult to implement in online learning accurately. Third, there is a lot of data to be processed quickly and precisely so that it can be reported to educational institutions and to students' families. Fourth, Informatics is a lesson that is 80% practical and 20% theory so that the assessment instruments used are 80% performance instruments (Bloom's taxonomy: C2, C3, C4, C5) and 20% multiple choice instruments (C1). Informatics correction and assessment requires more time because 80% cannot be assessed automatically. This research aims to predict student performance (Pass (1) or Intervention (0)) on the results of the online learning process assessment for informatics subjects in high school. If the student performance prediction results in an intervention, it will be immediately followed up by providing an intervention strategy to increase student performance. The target of the research results is to achieve > 70% accuracy on the processed dataset. This research uses the ensemble learning method random Forest Classification and XG Boosting classification. The research results of Student Performance Prediction using XG Boost Classification produce higher accuracy than RF Classification which has an average accuracy value = 93% while

©2024 J-Intech. Published by LPPM STIKI Malang

This is an open access article under the CC BY SA license. (<https://creativecommons.org/licenses/by-sa/4.0/>)

RF Classification has an average accuracy result = 92%. The research objectives have been achieved because the results of the 2 methods used have met the desired targets.

Keywords: Informatics; Random Forest; Student Performance; SENIOR HIGH SCHOOL; XG Boosting

1. Pendahuluan

Pembelajaran online tidak akan ditinggalkan, justru jadi bagian penting pada edukasi 5.0 sesuai perkembangan iptek dan perubahan pola hidup manusia. Sudah banyak teknologi pembelajaran online seperti MOOCs, LMS, VLE dan berbagai teknologi pendidikan berbasis AI, yang membuka peluang untuk bisa mendapat akses terhadap Iptek secara lebih mudah, murah, terbuka kepada masyarakat luas. Di Endemi Corona, kita tidak sekedar kembali ke pola pendidikan offline namun sudah mengarah pada edukasi 5.0.(Ahmad et al., 2023)

Saat ini pembelajaran online, normal, blended sudah menjadi hal yang biasa. Bisa digunakan sesuai kebutuhan proses pembelajaran di SMA. Dan disesuaikan oleh guru pengampu dan pelajaran yang di ampuh. Beberapa hal yang masih tetap dipertahankan adalah kecepatan dan ketepatan dalam membuat dan mengirim laporan pada institusi pendidikan, orangtua didik serta peserta didik. Karena selama pembelajaran online hal itulah yang disukai oleh orangtua peserta didik untuk bisa memantau progres putra-putrinya dengan cepat dan tepat. Karena itulah pengukuran performanya juga tetap membutuhkan prediksi student performance yang cepat tepat. Agar peserta didik yang dibawah rata-rata segera ditindaklanjuti dengan cepat dan tepat.

Penilaian adalah salah satu hal yang penting untuk mengukur Student Performance. Penilaian pembelajaran Online membutuhkan prediksi student performance yang cepat dan tepat (akurasi tinggi). Beberapa hal yang menyebabkan penilaian pembelajaran online khususnya butuh dianalisa dengan cepat dan tepat adalah pertama, karena keterbatasan interaksi langsung. Kedua, pembelajaran normal biasanya ada penilaian proses belajar dan penilaian karakter untuk bisa memberikan penilaian akhir akurat dimana hal itu sulit untuk dapat dilaksanakan pada pembelajaran Online. Karena beberapa hal antara lain kendala teknis, contoh : layar zoom tidak bisa aktif karena rusak, sinyal tidak bagus, hemat bandwidth. Baterai drop. Semua itu merupakan pengalaman observasi peneliti pada waktu proses pembelajaran online. Ketiga, banyaknya data yang butuh dinilai dengan cepat dan tepat untuk segera dilaporkan kepada Institusi pendidikan dan pada keluarga peserta didik. Jika dilaksanakan penilaian manual normal maka tidak mencukupi waktu dan resource untuk melaksanakannya. Keempat, Informatika adalah pelajaran yang 80 % praktek dan 20 % teori sehingga instrument penilaian yang digunakan adalah 80 % instrument unjuk kerja (taksonomi bloom : C2, C3, C4, C5) dan 20 % instrument sejenis pilihan ganda (C1). Proses penilaian Informatika lebih membutuhkan banyak waktu karena 80 % tidak bisa dinilai secara otomatis. Apabila pengukuran student performance bisa dilakukan untuk pembelajaran Online maka hal itu juga akan sangat membantu pada pembelajaran normal ataupun blended.

Pada beberapa penelitian yang sudah dilakukan di bidang pengukuran student performance dengan menggunakan machine learning memberikan sebuah peluang penelitian yang baru. Penelitian yang dilakukan Nabil, A. dkk, peneliti menggunakan machine learning, Deep Learning untuk menilai student performance di Universitas Menggunakan jumlah 500 data berisi 16 feature yang umum institusional(Nabil et al., 2021). Ghorbani, R., dkk, peneliti menggunakan machine learning, deep learning untuk menilai student performance di universitas menggunakan jumlah 340 data berisi 19 feature yang umum dan institusional (Ghorbani & Ghousi, 2020). Saidani, dkk, peneliti menggunakan esembly learning : boosting, deep learning untuk menilai student performance di kursus persiapan kerja setelah universitas menggunakan jumlah 50 data berisi 18 feature yang umum dan institusional (Saidani et al., 2022). Ada celah penelitian yang membuat peneliti tertarik yaitu memprediksi student performance di satu SMA sebagai sumber data utama.

Penelitian ini bertujuan memprediksi student performance (Lulus (1) atau Intervensi (0)) pada hasil penilaian proses pembelajaran online mata pelajaran informatika di sma. Apabila hasil prediksi student performance menghasilkan Intervensi maka segera ditindaklanjuti dengan memberikan strategi intervensi supaya terjadi

peningkatan student performance. Target hasil penelitian mencapai akurasi > 70 % terhadap dataset yang diolah. Penelitian ini menggunakan metode ensemble learning random Forest Classification dan XG Boosting classification. Sehingga hasil akhir yang ingin diraih adalah student performance (tingkat kelulusan) khususnya mata pelajaran Informatika meningkat (Mienye & Sun, 2022).

Saat ini, evaluasi prediksi student performance dalam suatu organisasi memegang peranan penting dalam pembelajaran siswa. Prediksi tidak hanya bermanfaat untuk intervensi untuk peningkatan student performance. Namun juga bisa untuk membuat rekomendasi. STIKI Malang adalah perguruan tinggi yang bergerak di bidang Informatika, pada perguruan tinggi ini terdapat 4 program studi, yaitu program studi Teknik Informatika, Sistem Informasi, Manajemen Informatika, dan Desain Komunikasi Visual. Siswa/i yang telah lulus dari SMA/K dan ingin melanjutkan pendidikan di perguruan tinggi STIKI Malang akan memutuskan untuk memilih program studi yang ada di STIKI Malang. Sering kali terjadi tamu yang datang ke ruangan Penerimaan Mahasiswa Baru STIKI Malang/calon mahasiswa baru di STIKI Malang masih bingung untuk menentukan jurusan Oleh karena itu dibutuhkan system Rekomendasi yang sesuai (Perkasa & Eka Purwiantono, 2023).

Machine Learning adalah salah satu cabang ilmu komputer yang membuat komputer belajar sendiri tanpa bantuan program eksternal. Teknik Machine Learning ini dapat digunakan untuk memprediksi hasil tertentu (Kumar et al., 2020). Machine learning adalah dasar dari ensemble learning dan deep learning. Ensemble learning memperbaiki kinerja dari machine learning. *Ensemble learning methods* melatih banyak single learning dan menggabungkan prediksi mereka untuk mencapai kinerja yang lebih baik dan kemampuan generalisasi yang lebih baik daripada single learning. Ide dasar di balik *Ensemble learning methods* adalah pengakuan bahwa machine learning models memiliki keterbatasan dan dapat membuat kesalahan. Oleh karena itu, Ensemble learning methods bertujuan untuk meningkatkan kinerja klasifikasi dengan memanfaatkan kekuatan kumpulan beberapa single learning. Di sisi lain, algoritma ML memiliki beberapa keterbatasan. menghasilkan model dengan varian tinggi, bias tinggi, dan akurasi rendah. Namun, beberapa penelitian menunjukkan bahwa *Ensemble learning methods* sering kali mencapai akurasi yang lebih tinggi dibandingkan model ML individual. mengurangi varians tanpa meningkatkan bias yang harus dilakukan. Secara keseluruhan, *Ensemble learning methods* lebih kuat dan berkinerja lebih baik daripada single learning (Sahlaoui et al., 2021). *Ensemble learning methods* secara garis besar dapat diklasifikasikan menjadi *boosting*, *bagging*, dan *stacking*. *Gradient Boosting*, *XG Boost*, dan *Ada Boost* adalah contoh algoritma peningkatan (*boosting*), dan *Random Forest Classifier* dan *Extra Trees* adalah algoritma *bagging* yang terkenal. Contoh *stacking framework* adalah *super ensemble* dan *blending techniques*. Ensemble learning methods ini telah mencapai performa yang baik dalam berbagai aplikasi ML. Ia juga dikenal untuk digunakan dalam banyak aplikasi dunia nyata. Algoritma GBDT (Gradient Boosting Decision Tree) secara umum memiliki kinerja terbaik bila dibandingkan pada algoritme lainnya (Ahmed et al., 2021). XGBoost dikembangkan hanya untuk meningkatkan kinerja model dan kecepatan proses data (Ogunleye & Wang, 2020). Metode Ensemble adalah algoritme pilihan di banyak kompetisi ML karena performanya yang tinggi.

Pengambilan Sampel Berdasarkan SMOTE (Synthetic Minority Oversampling Technique) berfungsi mengurangi overfitting karena dataset yang tidak seimbang (Bujang et al., 2021), (Alwarthan et al., 2022).

Data set dibagi menjadi training set dan test set dengan menggunakan metode K-fold cross validation dengan nilai k set menjadi 10. Bagi data set menjadi k set dengan menggunakan metode k-fold transisi cross-validation. Satu set digunakan untuk pelatihan model dan set lainnya untuk pengujian model, yaitu untuk mengukur performa prediktif model pada data tak terlihat yang tidak digunakan selama pelatihan model (Adnan et al., 2021).

Sebagian besar mesin machine learning modern memiliki plugin yang perlu dimodifikasi sebelum dapat dijalankan. Parameter ini disebut hyperparameter. Seringkali kinerja algoritma untuk tugas pembelajaran bergantung pada pengaturan parameternya. Untuk mencapai performa terbaik, pakar machine learning dapat menyesuaikan parameternya (Weerts et al., 2020). Peneliti menggunakan algoritma pencarian grid untuk menentukan parameter yang relevan untuk setiap model klasifikasi (Yao et al., 2021). Grid search dirancang

©2024 J-Intech. Published by LPPM STIKI Malang

This is an open access article under the CC BY SA license. (<https://creativecommons.org/licenses/by-sa/4.0/>)

dengan sekumpulan tetap parameter yang diperlukan untuk akurasi optimal berdasar validasi silang n-fold. Algoritme Grid search memberikan parameter terbaik seperti jumlah fitur yang perlu dipertimbangkan dalam setiap distribusi, contoh : jumlah pohon di hutan, kedalaman maksimum pohon, dan lain sebagainya (Shekar & Dagneu, 2019).

Peneliti menggunakan empat rumus metrik untuk mengevaluasi kinerja model: Confusion Matrix, accuracy, precision, recall, F1-score. Sampel siswa dibagi menjadi empat kelompok: true positif (TP), true negative (TN), false positif (FP) dan false negative (FN).

$$Confusion\ Matrix = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \quad (1)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$F1\ Score = 2 * \left(\frac{precision*recall}{precision+recall} \right) \quad (5)$$

Tune up atau optimasi model algoritma pembelajaran yang tepat untuk semua hyper-parameter nya membutuhkan pengetahuan tentang algoritma, latihan, dan biasanya check and error. Optimasi untuk memperoleh solusi potensial terbaik secara sistematis dan efektif, agar prediktif pengklasifikasi kinerja meningkat. Beberapa optimasi hyper-parameter seperti Manual, Pencarian Grid (Grid Search), Pencarian Acak (Random Search) dan Optimasi Bayesian, terbukti efektif dalam memecahkan masalah penyetelan hyper-parameter dengan keunggulan dan kelemahan masing masing.

Banyak parameter algoritma machine learning didapatkan melalui training model. Selain itu, sebagian besar algoritma pembelajaran mesin modern memiliki parameter yang perlu disetel sebelum dijalankan. Parameter itu disebut sebagai hyperparameter. Dalam berbagai kasus, kinerja suatu algoritma pada tugas pembelajaran yang diberikan tergantung pada pengaturan hyperparameternya

2. Metode Penelitian

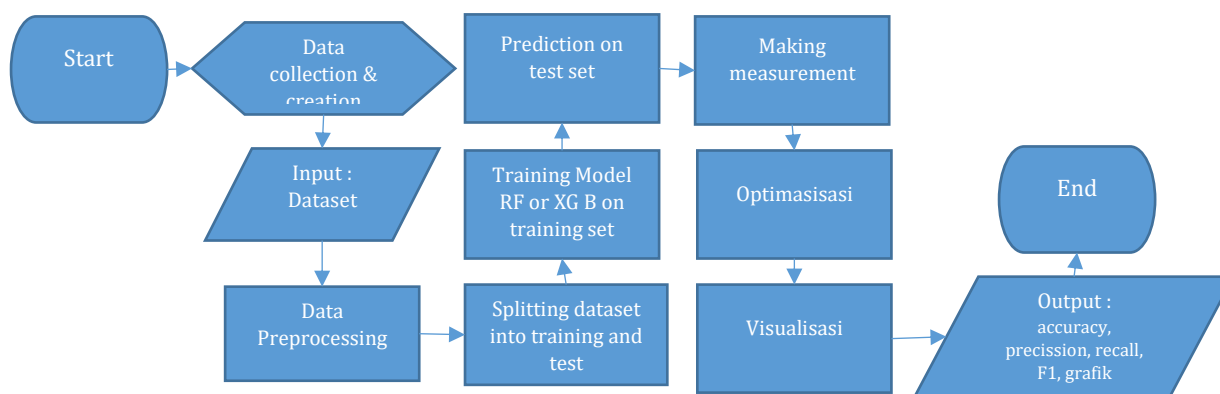
Batasan penelitian ini adalah dataset diambil hanya dari 1 SMA. Nilai proses pembelajaran Online mapel Informatika. Sumber data berupa dataset yang digunakan dikumpulkan hanya dari GCR (Google Classroom). Penelitian menggunakan algoritma Random Forest dan XG Boost. Penelitian ini menggunakan teknik optimalisasi algoritma yaitu K-fold cross-validation dan Grid Search Hyper parameter Optimization. Dataset yang digunakan hanya 3 tahun ajaran yaitu 2020-2021, 2021-2022, 2022-2023. Bahasa pemrograman yang digunakan adalah python. Software yang digunakan adalah google collaborator atau Jupiter Notebook.

Dataset di dapatkan dari download dari GCR berupa data-data nilai proses pembelajaran Informatika sejak tahun ajaran 2020-2021, 2021-2022, 2022-2023. Kemudian diolah dan disimpan dengan format csv. Total data adalah 1373, terbagi menjadi Data X = 454 (11 feature), Data XI (9 feature) = 387, Data XII (4 feature) = 532, Data kombinasi =1373(8 feature). Semua dataset di save dengan format .csv. Feature berdasarkan nilai TP (Tujuan pembelajaran) semua mempunyai tipe data numerik. Dan feature target merupakan jenis kategori yaitu Lulus (1) dan intervensi (0).

Data collection adalah pengumpulan data dari GCR (google classroom) dan pembuatan dataset. Data pre-processing dilakukan sesuai kebutuhan kondisi dataset, yaitu data cleaning adalah pembersihan data yang

NaN. Kemudian *data discretization* adalah mengubah data numerik menjadi nominal di bagian target. Dilanjutkan *feature encoding* yaitu mengubah data kategorikal menjadi bentuk numerik sebelum memasukkan fitur input ke dalam model. Kemudian, *handling imbalanced dataset* yaitu menggunakan resampling method seperti SMOOTH dan *feature scaling*.

Pada model *validation* menggunakan random *hold-out* method yaitu mengubah dataset menjadi 80 % data train dan 20 % data test. Ensemble learning models menggunakan random Forest dan XG Boost. Optimalisasi Model menggunakan *K-fold cross-validation* dan *Grid Search Hiper parameter Optimization*. *K-fold cross-validation* adalah metode yang ditingkatkan dari *cross-validation* untuk peningkatan performa *machine learning*. Untuk *evaluation measures* menggunakan *Accuracy, Precision, Recall, F1-score*.



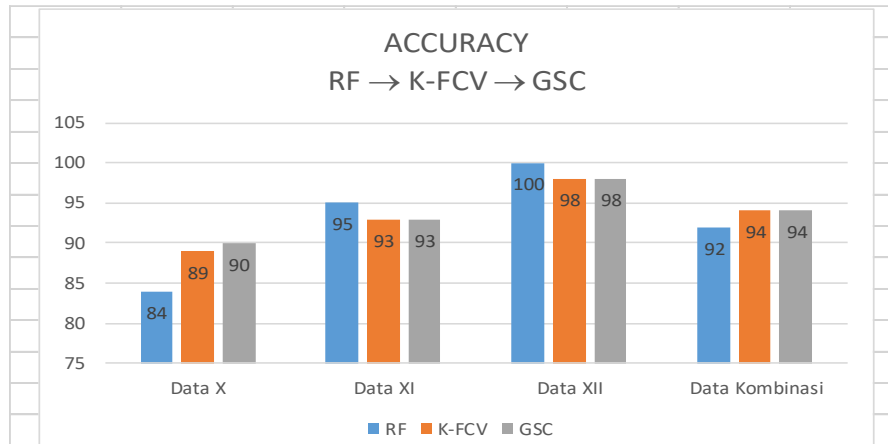
Gambar 1. Flowchart Ensemble Learning

3. Hasil

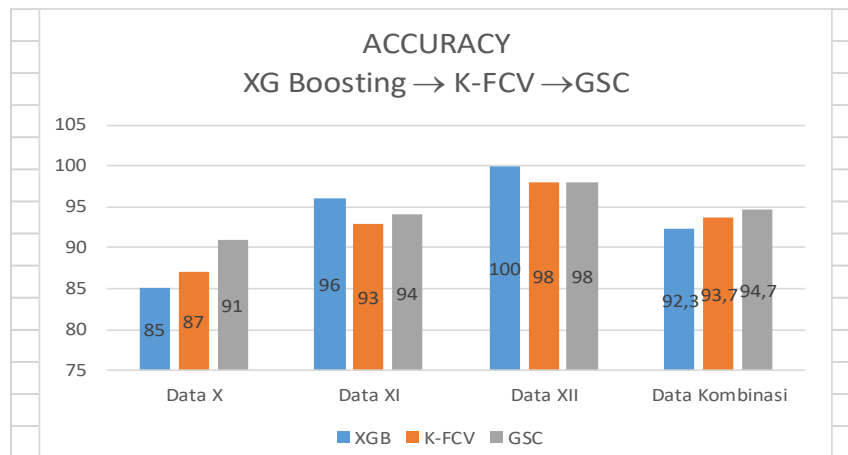
Uji coba dilaksanakan dalam 8 x menggunakan 4 dataset berbeda yang dipakai pada 2 model ensemble learning yaitu Random Forest Classification dan XG Boosting Clasification. Eksperimen dilaksanakan memakai google collaboration. Hasil ujicoba semua dataset untuk random forest classification maupun XG boosting classification (Tabel1).

Tabel 1. Data Diolah oleh Peneliti 2024

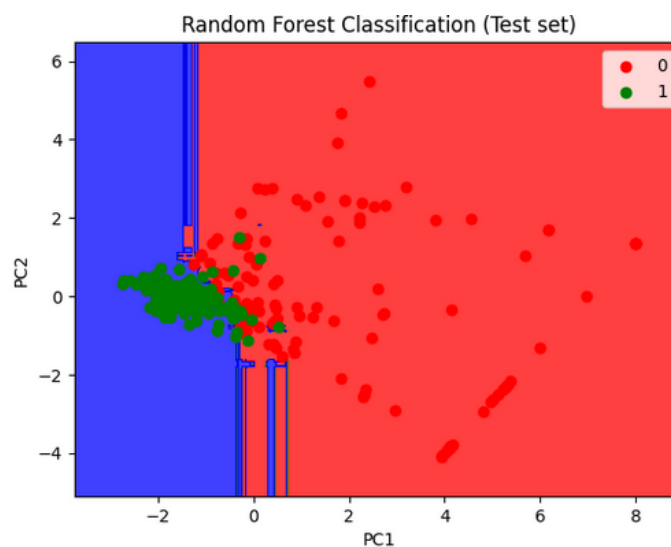
Uji coba	Dataset	Random Forest Classification				k-FCV	GridSearchCV
		<i>accuracy</i>	<i>precision</i>	<i>recall</i>	<i>f1</i>	<i>accuracy</i>	<i>accuracy</i>
1	Data X	0,84	0,84	0,84	0,84	0,89	0,9
2	Data XI	0,95	0,95	0,95	0,95	0,93	0,93
3	Data XII	1	1	1	1	0,98	0,98
4	Data Kombinasi	0,921	0,921	0,921	0,921	0,94	0,94
Uji coba	Dataset	XGB Classification				k-FCV	GridSearchCV
		<i>accuracy</i>	<i>precision</i>	<i>recall</i>	<i>f1</i>	<i>accuracy</i>	<i>accuracy</i>
5	Data X	0,85	0,85	0,85	0,85	0,87	0,91
6	Data XI	0,96	0,96	0,96	0,96	0,93	0,94
7	Data XII	1	1	1	1	0,98	0,98
8	Data Kombinasi	0,923	0,923	0,923	0,923	0,93	0,94



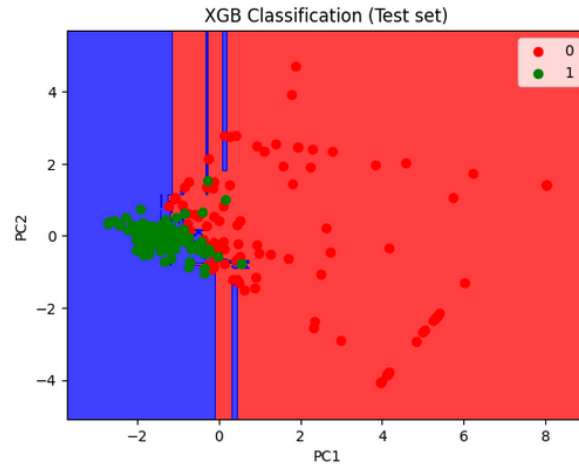
Gambar 2. Hasil Accuracy Data Kombinasi Random forest (data diolah oleh peneliti, 2024)



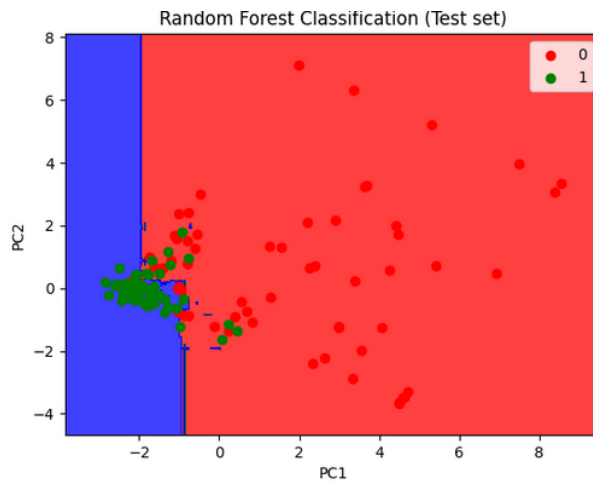
Gambar 3. Hasil Accuracy XG Boosting (data diolah oleh peneliti, 2024)



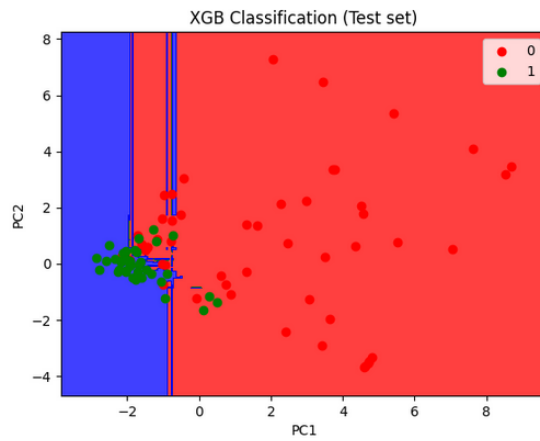
Gambar 4. Hasil ujicoba Data Kombinasi – data test - RF



Gambar 5. Hasil ujicoba Data Kombinasi-data test - XG Boosting



Gambar 6. Hasil ujicoba Data X-data test - RF



Gambar 7. Hasil ujicoba Data X-data test - XG Boosting

Pada gambar 4 dan gambar 5 adalah tampilan Data Kombinasi berupa data test yang dijalankan dalam model *Random Forest Classification* dan dalam model XG Boosting. Dimana data Kombinasi adalah dataset yang memiliki nilai akurasi terbaik dari semua dataset yaitu yang mengalami kenaikan nilai akurasi secara konsisten. Sedangkan gambar 6 dan gambar 7 adalah tampilan Data X berupa data test yang dijalankan dalam model *Random Forest Classification* dan dalam model XG Boosting. Dimana data X adalah data yang memiliki akurasi paling rendah dari semua dataset. Pada gambar tersebut, tersebar 2 kategori data yang beda yaitu bulatan warna hijau yang mewakili kategori 1 yaitu lulus. Sedangkan bulatan warna merah mewakili kategori 0 yaitu intervensi. Area latar belakang biru adalah area kategori 1 sedangkan area latar belakang merah adalah area kategori 0. Jumlah kategori 1 yang masuk area kategori 0 adalah *outlier* yang merupakan indikator besar kecilnya akurasi yang diraih. Semakin banyak kategori 1 yang berada di area kategori 0 maka semakin kecil akurasinya.

4. Pembahasan

Dari hasil ujicoba (tabel 1)(gambar 2)(gambar 3) Algoritma XG Boosting memiliki rata-rata akurasi lebih tinggi daripada *Random Forest*. Nilai akurasi terbaik ada pada dataset Kombinasi dengan XG Boosting karena performa nilai akurasi yang terus meningkat dari algoritma XG boosting lalu dilanjutkan K-fold cross-validation kemudian *Grid Search Hiper parameter Optimization*. Nilai *accuracy* terendah ada pada dataset X dengan *Random Forest*. Terjadi *overfitting* pada dataset XII dengan *random Forest* ataupun XG Boosting, hal ini terjadi karena feature Data XII paling sedikit yaitu 4 feature.

Optimalisasi K-fold cross-validation dan *Grid Search Hiper parameter Optimization* bekerja akumulatif meningkatkan hasil akurasi algoritma secara rata-rata. Sedangkan pada *overfitting*, bisa memperbaiki performa menjadi tidak lagi *overfitting*.

Hasil dari penelitian ini adalah solusi untuk mengatasi masalah prediksi *student performance* pembelajaran online Informatika di SMA. Dengan menerapkan metode prediksi *student performance* menggunakan *random forest* dan XG Boosting cepat dan tepat dalam prediksi nilai pembelajaran online Informatika. Kemudian segera membuat laporan kepada institusi, orangtua didik. Sehingga segera ada intervensi-penanganan kolaboratif terhadap peserta didik yang memiliki *student performance* di bawah rata-rata oleh pihak institusi, orangtua didik, guru pelajaran. Hasil akhirnya terjadi peningkatan *student performance* yaitu tingkat kelulusan informatika meningkat.

5. Kesimpulan

Tujuan penelitian sudah terpenuhi dengan hasil Prediksi *Student Performance* menggunakan XGB Classification menghasilkan akurasi lebih tinggi daripada RF Classification dengan nilai akurasi rata-rata = 93 % sedangkan RF Classification memiliki hasil akurasi rata-rata = 92 %. Hal ini melebihi target yang diinginkan yaitu > 70 % terhadap data yang diolah.

Saran bagi penelitian berikutnya adalah menggunakan algoritma ensemble learning jenis voting dan stacking. Menggunakan optimasi Bayesian. Menggunakan data karakter sebagai salah satu feature independent. Menggunakan data nilai proses pembelajaran offline dikombinasikan dengan data online.

Referensi

- Adnan, M., Habib, A., Ashraf, J., Mussadiq, S., Raza, A. A., Abid, M., Bashir, M., & Khan, S. U. (2021). Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models. *IEEE Access*, 9, 7519–7539. <https://doi.org/10.1109/ACCESS.2021.3049446>
- Ahmad, S., Umirzakova, S., Mujtaba, G., Amin, M. S., & Whangbo, T. (2023). *Education 5.0: Requirements, Enabling Technologies, and Future Directions*. <http://arxiv.org/abs/2307.15846>
- Ahmed, D. M., Abdulazeez, A. M., Zeebaree, D. Q., & Ahmed, F. Y. H. (2021). Predicting University's Students Performance Based on Machine Learning Techniques. *2021 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2021 - Proceedings*, 276–281. <https://doi.org/10.1109/I2CACIS52118.2021.9495862>

- Alwarthan, S., Aslam, N., & Khan, I. U. (2022). An Explainable Model for Identifying At-Risk Student at Higher Education. *IEEE Access*, 10, 107649–107668. <https://doi.org/10.1109/ACCESS.2022.3211070>
- Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. *IEEE Access*, 9, 95608–95621. <https://doi.org/10.1109/ACCESS.2021.3093563>
- Ghorbani, R., & Ghousi, R. (2020). Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. *IEEE Access*, 8, 67899–67911. <https://doi.org/10.1109/ACCESS.2020.2986809>
- Kumar, V. U., Krishna, A., Neelakanteswara, P., & Basha, C. Z. (2020). Advanced Prediction of Performance of a Student in an University using Machine Learning Techniques. *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, 121–126. <https://doi.org/10.1109/ICESC48915.2020.9155557>
- Mienye, I. D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. In *IEEE Access* (Vol. 10, pp. 99129–99149). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2022.3207287>
- Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks. *IEEE Access*, 9, 140731–140746. <https://doi.org/10.1109/ACCESS.2021.3119596>
- Ogunleye, A., & Wang, Q. G. (2020). XGBoost Model for Chronic Kidney Disease Diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(6), 2131–2140. <https://doi.org/10.1109/TCBB.2019.2911071>
- Perkasa, K. B. P. Y., & Eka Purwiantono, F. (2023). Sistem Rekomendasi Jurusan Menggunakan Algoritma Naïve Bayes Gaussian Berbasis Web. *J-INTECH*, 11(2), 361–370. <https://doi.org/10.32664/j-intech.v11i2.1090>
- Sahlaoui, H., Alaoui, E. A. A., Nayyar, A., Agoujil, S., & Jaber, M. M. (2021). Predicting and Interpreting Student Performance Using Ensemble Models and Shapley Additive Explanations. *IEEE Access*, 9, 152688–152703. <https://doi.org/10.1109/ACCESS.2021.3124270>
- Saidani, O., Menzli, L. J., Ksibi, A., Alturki, N., & Alluhaidan, A. S. (2022). Predicting Student Employability Through the Internship Context Using Gradient Boosting Models. *IEEE Access*, 10, 46472–46489. <https://doi.org/10.1109/ACCESS.2022.3170421>
- Shekar, B. H., & Dagnew, G. (2019). Grid Search-Based Hyperparameter Tuning and Classification of Microarray Cancer Data. *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, 1–8. <https://doi.org/10.1109/ICACCP.2019.8882943>
- Weerts, H. J. P., Mueller, A. C., & Vanschoren, J. (2020). *Importance of Tuning Hyperparameters of Machine Learning Algorithms*. <http://arxiv.org/abs/2007.07588>
- Yao, J., Zheng, Y., & Jiang, H. (2021). An Ensemble Model for Fake Online Review Detection Based on Data Resampling, Feature Pruning, and Parameter Optimization. In *IEEE Access* (Vol. 9, pp. 16914–16927). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2021.3051174>