Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa

Uce Indahyanti [1)*], Nuril Lutvi Azizah[2)] , Hamzah Setiawan[3)]
[1)2)3)] Program Studi Informatika, Fakultas Sains & Teknologi, Universitas Muhammadiyah Jl. Mojopahit 666B, Sidoarjo
[1)]*uceindahyanti@umsida.ac.id
[2)]nurillutviazizah@umsida.ac.id
[3)]hamzah@umsida.ac.id

Abstrak

Penelitian ini bertujuan untuk meningkatkan akurasi prediksi kinerja mahasiswa dalam sistem pembelajaran *virtual* atau *elearning* menggunakan pendekatan *ensemble learning*. Dataset penelitian merupakan data publik berupa data log aktifitas elearning. Dataset yang telah melalui tahap *preprocessing*, dimasukkan ke dalam pemodelan prediksi menggunakan gabungan beberapa algoritma pengklasifikasi yaitu *Decision Tree*, *Random Forest*, dan *AdaBoost* (*ensemble learning*). Tahap berikutnya mengevaluasi kinerja model dan menganalisis hasil prediksi menggunakan teknik *root mean square error* (RMSE). Output pemodelan berupa tiga *level* prediksi kinerja akademik (kelulusan mahasiswa) dalam sebuah course/semester, yaitu *low-level*, *middle-level*, dan *high-level*. Hasil pemodelan menunjukkan bahwa algoritma RF menghasilkan prediksi yang lebih akurat dibandingkan algoritma *Decision Tree* dan *AdaBoost*, yaitu sebesar 75.79%, dengan RMSE mendekati 0 yaitu 0.44. Dampak penelitian ini dapat memberikan tambahan kajian sekaligus konfirmasi penerapan teknik *ensemble learning* dalam pemodelan prediksi. Selain itu hasil prediksi kinerja akademik mahasiswa dapat dijadikan sebagai bahan evaluasi dalam proses pembelajaran.

Kata kunci: prediksi kinerja akademik mahasiswa, *ensemble learning*, RMSE

Abstract

This study aims to improve the prediction accuracy of student performance in a virtual learning system or e-learning using an ensemble learning approach. The research dataset is public data in the form of e-learning activity log data. The dataset that has gone through the pre-processing stage is entered into predictive modeling using a combination of several classification algorithms, namely Decision Tree, Random Forest, and AdaBoost (ensemble learning). The next stage is evaluating the performance of the model and analyzing the prediction results using the root mean square error (RMSE) technique. The modeling output is in the form of three levels of prediction of academic performance (student graduation) in a course/semester, namely low-level, middle-level, and high-level. The modeling results show that the RF algorithm produces more accurate predictions than others, equal to 75.79, with RMSE close to 0 which is 0.44. The impact of this research can provide additional studies and confirmation of the application of ensemble learning techniques in predictive modeling. The results of predicting student academic performance can also be used as an evaluation in the learning process.

Keywords: *student academic performance prediction, ensemble learning, RMSE*

1. PENDAHULUAN

Salah satu aspek penting dalam akademik adalah keberhasilan studi mahasiswa. Mendeteksi resiko kegagalan mahasiswa atau kinerja akademik mereka dalam mengikuti perkuliahan berbasis *elearning*, merupakan upaya penting sebagai strategi pencegahan dan umpan balik dalam mengevaluasi keberhasilan akademik [1], [2]. Tetapi upaya tersebut tidak mudah, mengingat beragamnya faktor atau karakteristik yang mempengaruhi kegagalan maupun keberhasilan mahasiswa [3]–[5]. Rekam jejak aktifitas mahasiswa yang menggambarkan kinerjanya dalam

DOI: 10.34128/jsi.v8i2.459 *Received: 6 Agustus 2022*
Accepted: 14 Desember 2022

elearning dapat dilihat dari beberapa data log aktifitas elearning, antara lain atribut *assignment*, *courses*, *resources*, *forums*, *quizzes*, *choices*, dan urls [6].

Sejumlah penelitian yang berfokus pada prediksi kinerja mahasiswa telah banyak dilakukan, baik dalam pembelajaran berbasis *elearning* maupun pembelajaran tradisional. Teknik data mining menggunakan pendekatan *machine learning* maupun statistik telah banyak digunakan untuk menganalisis dan memprediksi kinerja akademik mahasiswa, termasuk menemukan fitur atau atribut yang mempengaruhi hasil studi [7]. Beberapa penelitian sejenis menyebutkan pendekatan *ensemble learning* mampu menghasilkan prediksi yang lebih akurat dibandingkan model dasar atau pengklasifikasi tunggal [4], [7], [8].

Berdasarkan hal tersebut di atas, diusulkan penelitian ini yang bertujuan untuk meningkatkan akurasi prediksi kinerja mahasiswa dalam sistem pembelajaran virtual atau *elearning* menggunakan pendekatan ensemble learning. Dampak dari penelitian ini diharapkan dapat memberikan tambahan referensi bagi penelitian terkait penerapan teknik *ensemble* dalam pemodelan prediksi, serta memberikan bahan evaluasi proses pembelajaran bagi pemangku kepentingan. Teknik *ensemble* pada penelitian ini menggabungkan algoritma *Random Forest*, *AdaBoost*, dan *Decision Tree* (satu pohon keputusan atau pengklasifikasi tunggal). Sesuai namanya *Random Forest* merupakan kumpulan dari banyak pohon keputusan yaitu salah satu algoritma *bagging* yang sederhana namun efektif dan telah diterapkan pada banyak aplikasi [9]. *AdaBoost* merupakan salah satu algoritma *boosting* yang dapat meningkatkan kinerja pengklasifikasi dalam banyak situasi, termasuk ketika data tidak seimbang [10]. Teknik pengambilan nilai terbaik pada penelitian ini menggunakan *majority vote*, yaitu prediksi akhir diambil dari nilai akurasi tertinggi yang dihasilkan oleh ketiga algoritma. Dataset yang digunakan berbasis *log* aktifitas elearning, dan atribut yang diteliti meliputi yaitu *viewing, discussion groups, absence day, raised hand,* dan *visited resources*. Atribut-atribut tersebut dipilih karena terkait langsung dan merupakan representasi aktifitas mahasiswa dalam proses *elearning*. Sebagai contoh *log* aktifitas akses materi (*visited resources*).

2. TINJAUAN PUSTAKA

Penambangan data (data mining) dapat menggunakan pendekatan *machine learning* atau *statistic*. Teknik ini telah banyak digunakan untuk menganalisis dan memprediksi kinerja akademik mahasiswa, termasuk menemukan fitur atau atribut yang mempengaruhi hasil studi mahasiswa [7].

Decision tree merupakan sebuah algoritma pengklasifikasi dalam machine learning, dimana setiap cabangnya menunjukkan alternatif pilihan, dan setiap daunnya menunjukkan keputusan yang dipilih. Sedangkan *ensemble learning* merupakan sebuah pendekatan berbasis pohon keputusan (*decision tree*) yang menggabungkan beberapa algoritma pengklasifikasi. Hasil prediksi dari masing-masing algoritma pengklasifikasi tersebut akan dipilih berdasarkan majority vote untuk selanjutnya dijadikan hasil akhir pemodelan. Metode *ensemble learning* terdiri dari teknik *bagging* dan *boosting*. *Ensemble Random Forest* merupakan salah satu algoritma bagging yang sederhana namun efektif dan telah diterapkan untuk banyak aplikasi di dunia nyata [9]. *Bagging* merupakan metode *ensemble* yang banyak diterapkan pada algoritma klasifikasi, dengan tujuan untuk meningkatkan akurasi pengklasifikasi dengan menggabungkan pengklasifikasi tunggal, dan hasilnya lebih baik daripada *random sampling* [11].

Ensemble AdaBoost merupakan salah satu algoritma *boosting*. Secara umum algoritma boosting lebih baik dari pada *bagging*, tetapi tidak merata. *Boosting* dapat meningkatkan kinerja pengklasifikasi dalam banyak situasi, termasuk ketika data tidak seimbang [10].

3. METODE PENELITIAN

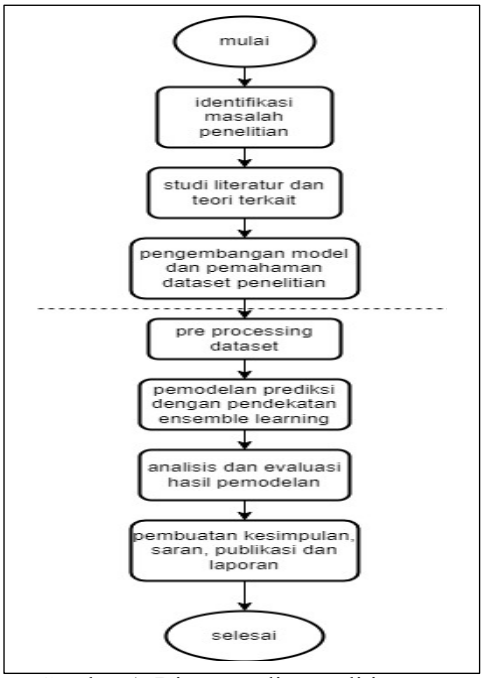
Diagram alir dan rancangan dimulai dari langkah-langkah berurutan sebagai berikut: identifikasi masalah penelitian, studi literatur, pemodelan, analisis hasil, kesimpulan, seperti yang ditampilkan pada Gambar 1.

Pemodelan prediksi yang diusulkan menggunakan gabungan teknik *ensemble Random Forest*, *AdaBoost*, dan *Decision Tree*, dengan teknik *majority vote* untuk memperoleh hasil prediksi akhir.

p-ISSN: 2460-173X
e-ISSN: 2598-5841

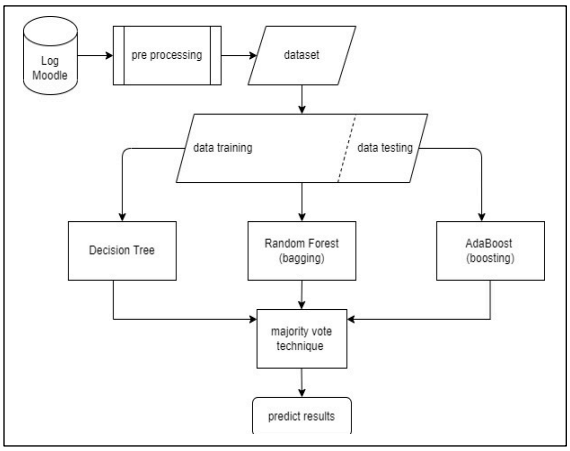
Sebelumnya dataset melalui tahap preprocessing dan pembagian menjadi data training dan data testing dengan rasio 80% : 20%, seperti yang ditampilkan pada Gambar 2.

Dataset penelitian merupakan data publik diunduh dari Kaggle.com, yang terdiri dari 480 data (record) dan 17 atribut (variable). Pada penelitian ini atribut intrinsik mahasiswa diabaikan, seperti gender dan identitas pribadi lainnya, dan dipilih atribut yang terkait langsung dengan aktifitas elearning (sesuai yang ada pada dataset) yaitu *viewing, discussion groups, absence day, raised hand, visited resources*, dan *class* sebagai label prediksi kinerja mahasiswa atau tingkat kelulusan. Seperti halnya [6] yang memilih atribut-atribut yang terkait langsung dengan aktifitas elearning sesuai dengan dataset yang digunakan. Sebelum dimasukkan ke pemodelan, dataset harus melalui tahap persiapan. Dimulai dari pembersihan data dengan cara menghilangkan data redundan dan missing value, mengkonversi *dataset* ke dalam format csv, sampai dengan memilah dataset menjadi data training dan data *testing*.



Gambar 1. Diagram alir penelitian

Masing- masing algoritma dalam pemodelan prediksi ini memproses dataset yang sama dan menghasilkan nilai prediksi berbeda-beda berupa klasifikasi kelulusan. Selanjutnya hasil akhir ditentukan melalui teknik *majority vote*. Hasil pemodelan berupa tiga level klasifikasi prediksi kinerja akademik (kelulusan) mahasiswa, yaitu Low-Level (0-69), Middle-Level (70-89), dan High-Level (90-100).



Gambar 2. Pemodelan prediksi kinerja akademik mahasiswa

p-ISSN: 2460-173X
e-ISSN: 2598-5841

Kinerja algoritma pengklasifikasi diukur menggunakan acuan confusion matrix, yang merepresentasikan hasil prediksi dan kondisi aktual berupa akurasi, presisi, dan recall. Akurasi merupakan rasio prediksi benar positif dan negatif, presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positf, sedangkan recall merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

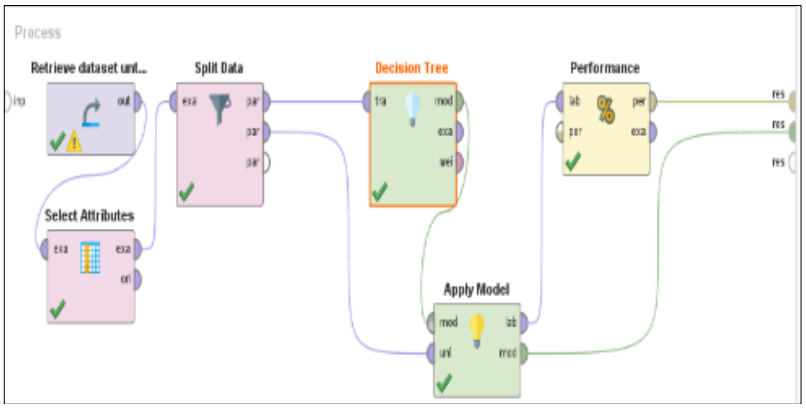
Hasil kinerja pemodelan berupa akurasi, presisi, dan recall selanjutnya dievaluasi menggunakan teknik RMSE (*root mean square error*). RMSE merupakan besarnya tingkat kesalahan hasil prediksi, jika nilainya semakin kecil (mendekati 0), maka hasil prediksi akan semakin akurat. Nilai RMSE dapat dihitung melalui Persamaan 1, dimana At adalah nilai aktual atau pengamatan, Ft adalah nilai prediksi, a adalah banyaknya data, dan lambang summary untuk keseluruhan nilai.

$$\text{RMSE}=\sqrt{\frac{\sum_{t=1}^{n}(A_t-F_t)^2}{n}} \quad (1)$$

4. PEMBAHASAN

Dataset yang dimasukkan ke dalam pemodelan merupakan data yang telah melalui proses pembersihan, pelabelan, pemilahan dan pembagian (split data) menjadi data training dan data testing dengan rasio 80 : 20. Sebelum model dijalankan, beberapa parameter dipilih untuk mengukur kinerja model. Parameter tersebut antara lain kedalaman pohon, nilai kepercayaan, dan pembobotan atribut (*information gain*). Kedalaman pohon dipilih 10 dengan melakukan proses pruning (pemangkasan pohon) dengan nilai confidence 0.5. Pohon keputusan yang dihasilkan dapat berukuran besar, sehingga dapat disederhanakan dengan proses pruning berdasarkan nilai kepercayaan (*confident level*) [12].

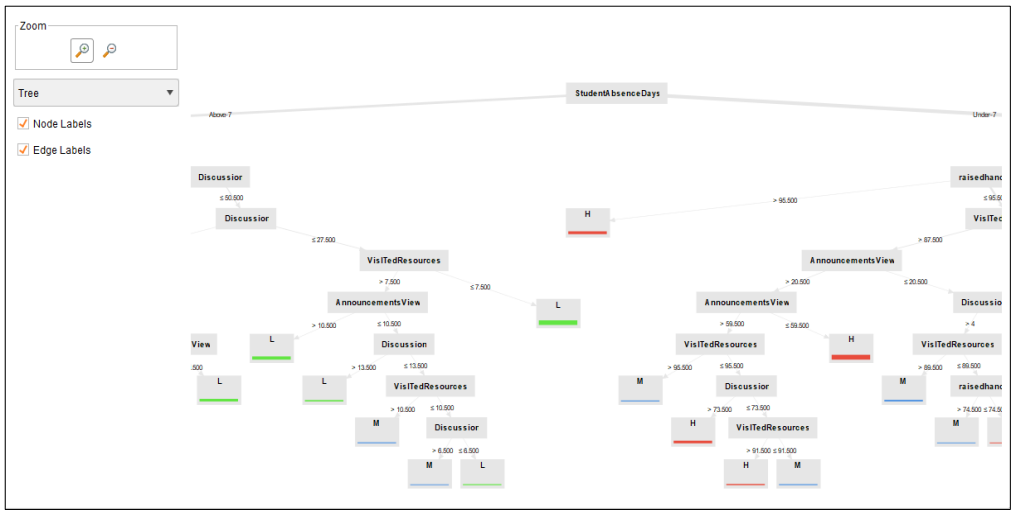
Gambar 3 menampilkan desain model pohon keputusan menggunakan algoritma pengklasifikasi tunggal (Decision Tree) dengan alat bantu Rapid Miner, dimulai dari pengambilan dataset, pemilihan atribut, pembagian data training dan testing, penerapan algoritma, sampai dengan pengujian kinerja.



Gambar 3. Pemodelan DT

Setelah model dijalankan, muncul output model dalam bentuk struktur root seperti yang diilustrasikan pada Gambar 4. Setiap node dalam pohon merepresentasikan atribut dan cabangnya merepresentasikan nilai dari atribut, sedangkan daunnya digunakan untuk merepresentasikan kelas. Node teratas dari pohon keputusan disebut dengan root.

p-ISSN: 2460-173X
e-ISSN: 2598-5841



Gambar 4. Struktur *root* pohon tunggal DT

Sedangkan bentuk deskripsi percabangan (if-then-else) dari pohon keputusan ditampilkan pada Gambar 5, yang memberikan gambaran aturan if-then-else pada hasil pemodelan menggunakan algoritma DT dan diperoleh StudentAbsenceDays sebagai root.

Tree

```
StudentAbsenceDays = Above-7
|   VisITedResources > 26.500
|   |   AnnouncementsView > 11
|   |   |   VisITedResources > 76.500
|   |   |   |   Discussion > 78
|   |   |   |   |   raisedhands > 72.500: H {M=0, L=0, H=1}
|   |   |   |   |   raisedhands ≤ 72.500: M {M=2, L=0, H=0}
|   |   |   |   Discussion ≤ 78: M {M=22, L=0, H=0}
|   |   |   VisITedResources ≤ 76.500
|   |   |   |   Discussion > 18
|   |   |   |   |   VisITedResources > 50.500
|   |   |   |   |   |   Discussion > 41.500
|   |   |   |   |   |   |   VisITedResources > 69.500
|   |   |   |   |   |   |   |   raisedhands > 40.500: H {M=0, L=0, H=1}
|   |   |   |   |   |   |   |   raisedhands ≤ 40.500: M {M=1, L=0, H=0}
|   |   |   |   |   |   |   VisITedResources ≤ 69.500: M {M=7, L=0, H=0}
|   |   |   |   |   |   Discussion ≤ 41.500
|   |   |   |   |   |   |   Discussion > 31: L {M=0, L=5, H=0}
|   |   |   |   |   |   |   Discussion ≤ 31
|   |   |   |   |   |   |   |   raisedhands > 55.500: H {M=0, L=0, H=1}
|   |   |   |   |   |   |   |   raisedhands ≤ 55.500: M {M=2, L=0, H=0}
|   |   |   |   |   VisITedResources ≤ 50.500: M {M=13, L=1, H=0}
|   |   |   |   Discussion ≤ 18
|   |   |   |   |   AnnouncementsView > 66.500: M {M=1, L=0, H=0}
|   |   |   |   |   AnnouncementsView ≤ 66.500
|   |   |   |   |   |   VisITedResources > 67.500
|   |   |   |   |   |   |   raisedhands > 34.500: H {M=0, L=0, H=1}
|   |   |   |   |   |   |   raisedhands ≤ 34.500: L {M=0, L=1, H=0}
|   |   |   |   |   |   VisITedResources ≤ 67.500: L {M=0, L=5, H=0}
```

Gambar 5. Deskripsi percabangan DT

Hasil kinerja model menggunakan algoritma Decision Tree (DT) ditampilkan dalam bentuk tabel kinerja dapat dilihat pada Gambar 6.

p-ISSN: 2460-173X
e-ISSN: 2598-5841

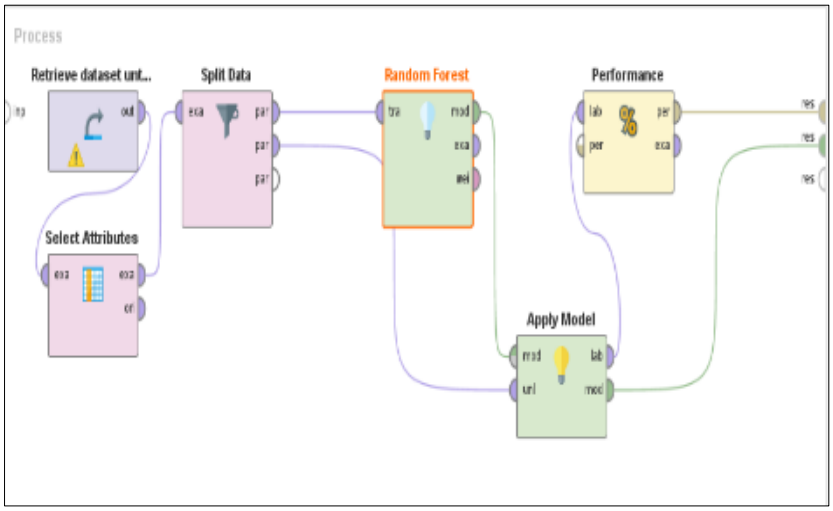
accuracy: 74.74%

	true M	true L	true H	class precision
pred. M	36	4	13	67.92%
pred. L	1	20	0	95.24%
pred. H	5	1	15	71.43%
class recall	85.71%	80.00%	53.57%	

Gambar 6. Kinerja Model DT

Pada tabel kinerja ditampilkan nilai *accuracy*, nilai *recall*, dan *precision* dari setiap class (klasifikasi tingkat kelulusan) dari model. Kinerja model ditunjukkan dalam tabel perbandingan kombinasi nilai prediksi dan nilai aktual (true) dari class atau pelabelannya, dalam hal ini ada 3 class yaitu Low-level (L), Middle-Level (M), High-level (H).

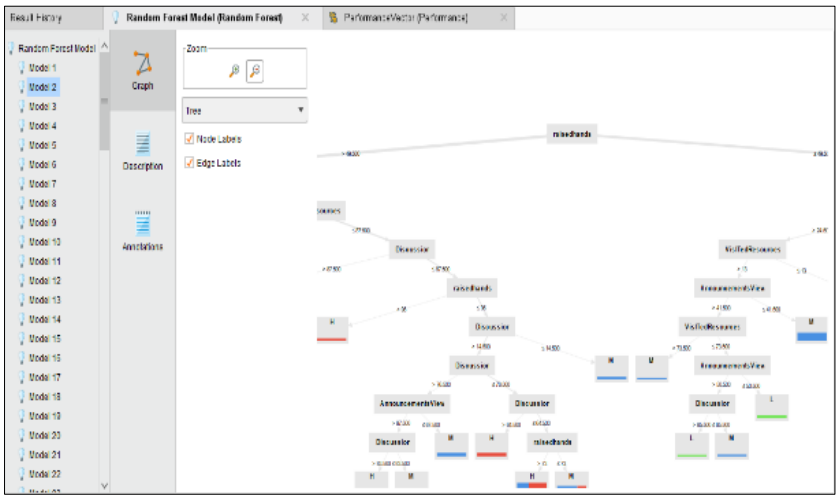
Berikutnya dilakukan tahapan pemodelan dengan tahapan yang sama seperti di atas, menggunakan algoritma *Random Forest* (RF). Algoritma RF merupakan penerapan metode *bagging* dari *ensemble learning*. Pemodelan RF ditampilkan pada Gambar 7.



Gambar 7. Pemodelan RF

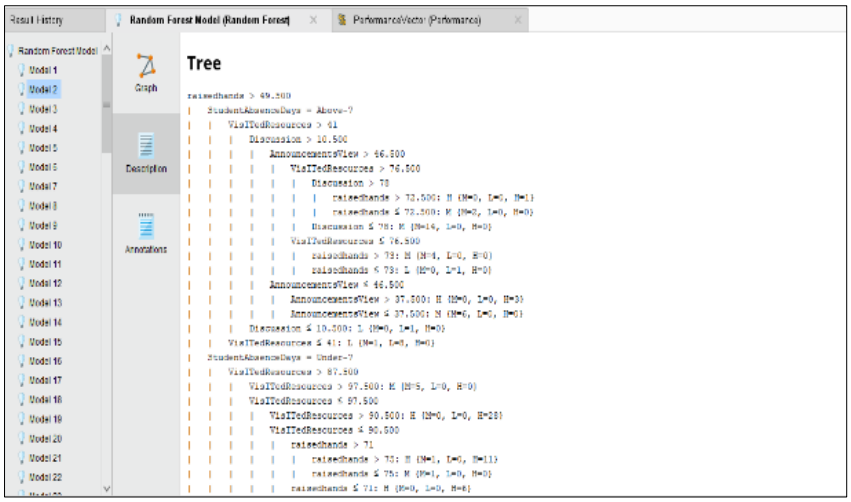
Algoritma Random Forest (RF) menghasilkan banyak pohon keputusan dengan teknik *bagging* sesuai jumlah pohon yang dipilih pada saat penentuan parameter. Gambar 8 menampilkan salah satu model dalam bentuk struktur root yang dihasilkan oleh algoritma RF. Setiap model yang dihasilkan oleh algoritma RF memiliki struktur root dan deskripsi percabangan yang berbeda-beda. Contoh model 2 yang disorot pada Gambar 8 memiliki akar (root) atribut *raise hands*.

p-ISSN: 2460-173X
e-ISSN: 2598-5841



Gambar 8. Salah satu struktur root RF

Sedangkan bentuk deskripsi percabangan dari struktur root di atas ditampilkan pada Gambar 9.



Gambar 9. Salah satu deskripsi percabangan RF

Selanjutnya dapat diketahui hasil kinerja keseluruhan model RF yang ditampilkan dalam bentuk tabel performance pada Gambar 10.

accuracy: 75.79%

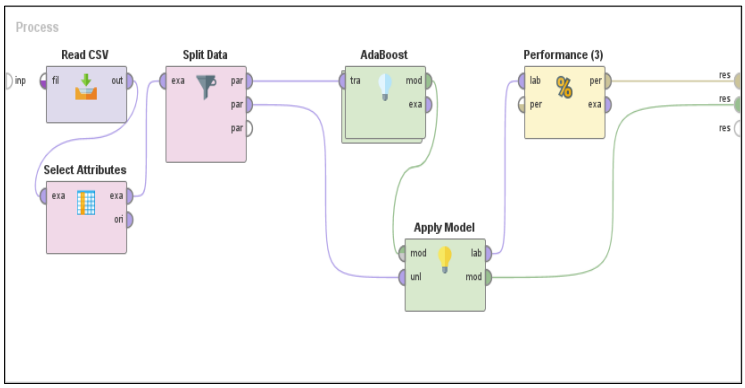
	true M	true L	true H	class precision
pred. M	31	4	8	72.09%
pred. L	1	21	0	95.45%
pred. H	10	0	20	66.67%
class recall	73.81%	84.00%	71.43%	

Gambar 10. Kinerja Keseluruhan Model RF

Tahapan pemodelan menggunakan algoritma AdaBoost juga mengikuti tahapan di atas. AdaBoost merupakan ensemble learning, yang juga menghasilkan beberapa pohon keputusan

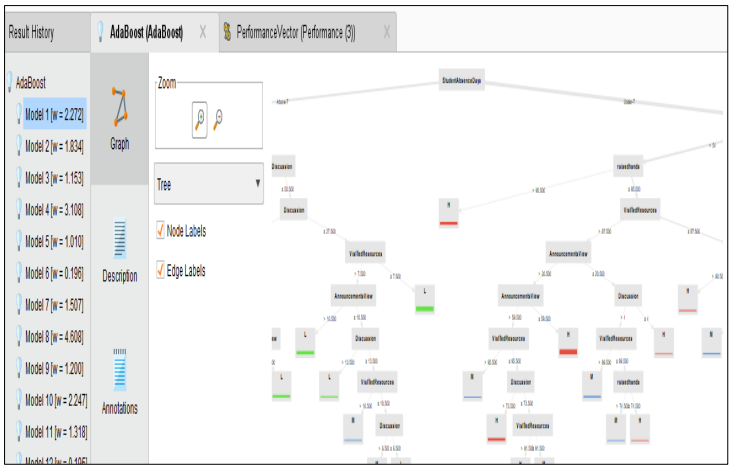
p-ISSN: 2460-173X
e-ISSN: 2598-5841

(model) dengan teknik *boosting*. Pemodelan AdaBoost ditampilkan pada Gambar 11.



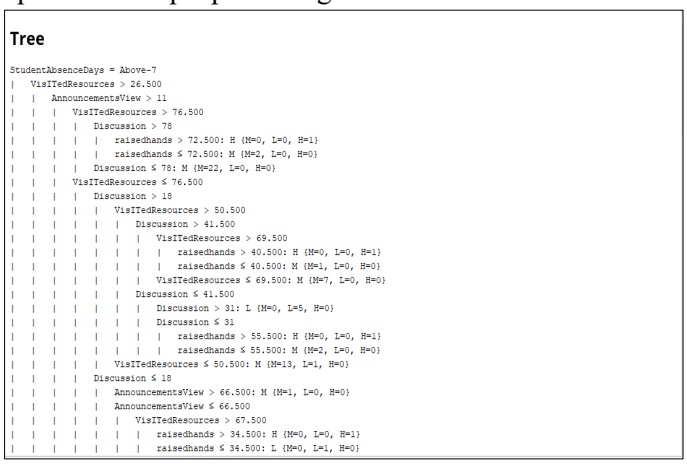
Gambar 11. Pemodelan AdaBoost

Gambar 12 menampilkan salah satu struktur root yang dihasilkan AdaBoost. Contoh model 1 yang disorot pada Gambar 12 memiliki akar (root) atribut *raise hands*.



Gambar 12. Salah satu struktur root AdaBoost

Gambar 13 menampilkan deskripsi percabangan dari struktur root model 1 di atas.

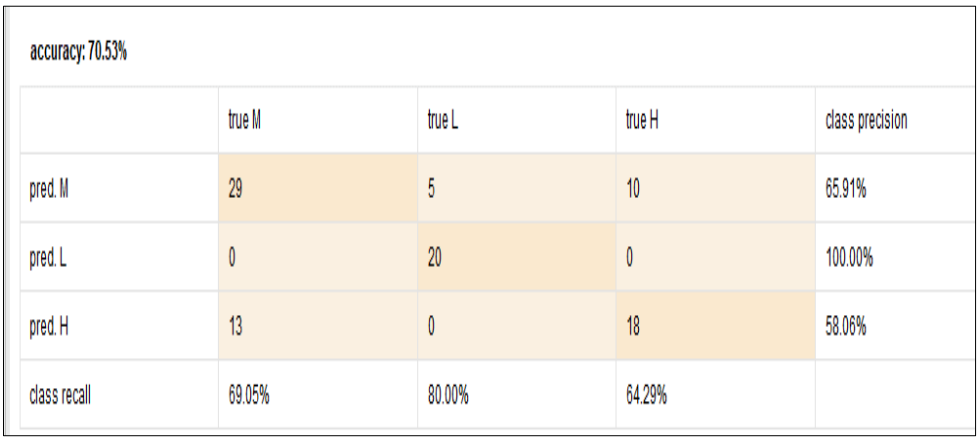


Tree

```
StudentAbsenceDays = Above-7
|   VisITedResources > 26.500
|   |   AnnouncementsView > 11
|   |   |   VisITedResources > 76.500
|   |   |   |   Discussion > 78
|   |   |   |   |   raisedhands > 72.500: H {M=0, L=0, H=1}
|   |   |   |   |   raisedhands ≤ 72.500: M {M=2, L=0, H=0}
|   |   |   |   Discussion ≤ 78: M {M=22, L=0, H=0}
|   |   |   VisITedResources ≤ 76.500
|   |   |   |   Discussion > 18
|   |   |   |   |   VisITedResources > 50.500
|   |   |   |   |   |   Discussion > 41.500
|   |   |   |   |   |   |   VisITedResources > 69.500
|   |   |   |   |   |   |   |   raisedhands > 40.500: H {M=0, L=0, H=1}
|   |   |   |   |   |   |   |   raisedhands ≤ 40.500: M {M=1, L=0, H=0}
|   |   |   |   |   |   |   VisITedResources ≤ 69.500: M {M=7, L=0, H=0}
|   |   |   |   |   |   Discussion ≤ 41.500
|   |   |   |   |   |   |   Discussion > 31: L {M=0, L=5, H=0}
|   |   |   |   |   |   |   Discussion ≤ 31
|   |   |   |   |   |   |   |   raisedhands > 55.500: H {M=0, L=0, H=1}
|   |   |   |   |   |   |   |   raisedhands ≤ 55.500: M {M=2, L=0, H=0}
|   |   |   |   |   VisITedResources ≤ 50.500: M {M=13, L=1, H=0}
|   |   |   |   Discussion ≤ 18
|   |   |   |   |   AnnouncementsView > 66.500: M {M=1, L=0, H=0}
|   |   |   |   |   AnnouncementsView ≤ 66.500
|   |   |   |   |   |   VisITedResources > 67.500
|   |   |   |   |   |   |   raisedhands > 34.500: H {M=0, L=0, H=1}
|   |   |   |   |   |   |   raisedhands ≤ 34.500: L {M=0, L=1, H=0}
```

Gambar 13. Deskripsi AdaBoost

Hasil kinerja keseluruhan model AdaBoost dalam bentuk tabel performance yang memuat akurasi, presisi, dan recall ditampilkan pada Gambar 14.



accuracy: 70.53%

	true M	true L	true H	class precision
pred. M	29	5	10	65.91%
pred. L	0	20	0	100.00%
pred. H	13	0	18	58.06%
class recall	69.05%	80.00%	64.29%	

Gambar 14. Hasil Performance AdaBoost

Selanjutnya model dievaluasi menggunakan menggunakan teknik RMSE untuk mengukur tingkat kesalahan hasil prediksi. Nilai RMSE yang dihasilkan masing-masing algoritma di atas dapat dilihat pada fitur Performance aplikasi Rapid Miner. Tabel 1 menunjukkan perbandingan performance yaitu tingkat akurasi dan nilai RMSE dari masing-masing algoritma pengklasifikasi yang digunakan. Melalui Teknik majority vote diperoleh nilai akurasi tertinggi sebesar 75.79% yang dihasilkan oleh algoritma RF dengan nilai RMSE terkecil sebesar 0.440.

Tabel 1. Perbandingan Performance

Algoritma	Akurasi	RMSE
RF	75.79%	0.440
AdaBoost	70.53%	0.489
DT	74.74%	0.531

5. KESIMPULAN

Berdasarkan hasil pengolahan dan pengujian data dapat disimpulkan bahwa pemodelan prediksi dalam penelitian ini dapat meningkatkan akurasi prediksi kinerja akademik mahasiswa menjadi sebesar 75.79%. Melalui teknik ensemble learning yang menerapkan beberapa algoritma ke dalam sebuah pemodelan (3 algoritma dalam penelitian ini), dimana tiap algoritma menghasilkan nilai prediksinya masing-masing, kemudian dengan teknik majority vote diambil nilai prediksi yang terbaik. Nilai terbaik tersebut diperoleh dari algoritma Random Forest melalui teknik majority vote. Algoritma Random Forest juga menghasilkan nilai RMSE yang lebih kecil dibandingkan algoritma DT dan AdaBoost, yaitu sebesar 0.440. Nilai RMSE semakin kecil mendekati 0, maka hasil prediksi akan semakin akurat.

Penelitian ini diharapkan dapat memberikan konfirmasi mengenai penerapan teknik ensemble learning dalam pemodelan prediksi. Jika dibandingkan dengan hanya menerapkan satu algoritma, teknik ensemble dapat memberikan hasil perbandingan beberapa algoritma untuk diperoleh nilai prediksi yang terbaik. Tetapi penelitian ini masih perlu disempurnakan, hasil pemodelan memang menunjukkan tingkat akurasi yang lebih baik, walaupun belum terlalu signifikan. Saran pengembangan ke depan dapat digunakan data privat berupa data log aktifitas elearning perguruan tinggi yang cukup besar, dengan menambahkan atribut kinerja akademik mahasiswa seperti tugas, UTS, dan UAS, serta menggali semua parameter dan mengkombinasikan semua fitur performance yang tersedia pada tahap running performance model.

DAFTAR PUSTAKA

[1] D. Monllaó Olivé, D. Q. Huynh, M. Reynolds, M. Dougiamas, and D. Wiese, “A supervised learning framework: using assessment to identify students at risk of dropping out of a MOOC,” *J. Comput. High. Educ.*, vol. 32, no. 1, pp. 9–26, 2020, doi: 10.1007/s12528-019-09230-1.

[2] R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, "Predicting student performance from LMS data: A comparison of 17 blended courses using moodle LMS," *IEEE Trans. Learn. Technol.*, vol. 10, no. 1, pp. 17–29, 2017, doi: 10.1109/TLT.2016.2616312.

[3] S. Rakic, N. Tasic, U. Marjanovic, S. Softic, E. Lüftenegger, and I. Turcin, "Student performance on an e-learning platform: Mixed method approach," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 2, pp. 187–203, 2020, doi: 10.3991/ijet.v15i02.11646.

[4] Y. Abubakar, N. Bahiah, and H. Ahmad, "Prediction of Students' Performance in E-Learning Environment Using Random Forest," *Int. J. Innov. Comput.*, vol. 7, no. 2, pp. 1–5, 2017, [Online]. Available: http://se.fsksm.utm.my/ijic/index.php/ijic

[5] S. A. Salloum, A. Qasim Mohammad Alhamad, M. Al-Emran, A. Abdel Monem, and K. Shaalan, "Exploring students' acceptance of e-learning through the development of a comprehensive technology acceptance model," *IEEE Access*, vol. 7, pp. 128445–128462, 2019, doi: 10.1109/ACCESS.2019.2939467.

[6] J. López-Zambrano, J. A. Lara, and C. Romero, "Towards portability of models for predicting students' final performance in university courses starting from moodle logs," *Appl. Sci.*, vol. 10, no. 1, 2020, doi: 10.3390/app10010354.

[7] M. N. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Systematic ensemble model selection approach for educational data mining," *Knowledge-Based Syst.*, vol. 200, p. 105992, 2020, doi: 10.1016/j.knosys.2020.105992.

[8] S. M. S. Samuel-Soma M. Ajibade, Nor Bahiah Ahmad, *A Data Mining Approach to Predict Academic Performance of Students Using Ensemble Techniques*. Springer International Publishing, 2020. doi: 10.1007/978-3-030-16657-1.

[9] G. Liang and C. Zhang, "An Empirical Evaluation of Bagging with Different," pp. 339–352, 2011.

[10] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "RUSBoost: Improving classification performance when training data is skewed," *Proc. - Int. Conf. Pattern Recognit.*, no. December, 2008, doi: 10.1109/icpr.2008.4761297.

[11] E. Alfaro, M. Gáamez, and N. García, "Adabag: An R package for classification with boosting and bagging," *J. Stat. Softw.*, vol. 54, no. 2, 2013, doi: 10.18637/jss.v054.i02.

[12] N. H. Harani, "Penerapan Adaboost Berbasis Pohon Keputusan Guna Menentukan Pola Masuknya Calon Mahasiswa Baru," *J. Transform.*, vol. 18, no. 1, p. 123, 2020, doi: 10.26623/transformatika.v18i1.1606.

Biodata Penulis

Uce Indahyanti, Lahir di Situbondo - Jawa Timur, pada bulan Mei 1971. Menyelesaikan pendidikan S1 Prodi Manajemen Informatika di STIKOM Surabaya (sekarang UNDIKA, 1990 – 1996), dan menyelesaikan pendidikan S2 Prodi Sistem Informasi di ITS Surabaya (2010-2012). Saat ini, sedang menempuh pendidikan program doktoral Ilmu Komputer di ITS Surabaya, dan aktif berkarya sebagai dosen tetap Prodi Informatika Fakultas Sains dan Teknologi Universitas Muhammadiyah Sidoarjo (UMSIDA).

Nuril Lutvi Azizah, lahir di Lumajang - Jawa Timur, pada bulan April 1989. Menyelesaikan pendidikan S1 Prodi Matematika di ITS Surabaya (2007-2011) dan menyelesaikan pendidikan S2 Prodi Matematika di ITS Surabaya (2011-2013). Sampai saat ini aktif sebagai dosen tetap di Universitas Muhammadiyah Sidoarjo pada Prodi Informatika Fakultas Sains dan Teknologi dan aktif melakukan kegiatan riset serta abdimas.

Hamzah Setiawan, lahir di Sidoarjo - Jawa Timur, pada bulan April 1986. Menyelesaikan pendidikan S1 pada prodi Teknik Informatika di Universitas Trunojoya lulus tahun 2009 dan menyelesaikan pendidikan S2 pada prodi Teknologi Informasi di Institut Sains dan Teknologi Terpadu Surabaya lulus tahun 2019, dan saat ini sedang menempuh pendidikan program doktoral Ilmu Komputer di ITS Surabaya, serta saat ini sebagai dosen di program studi Informarika di Fakultas Sains dan Teknologi Universitas Muhammadiyah Sidoarjo.