

Analisis Performa Akademik Mahasiswa Menggunakan Distributed Random Forest

Herman^{1*}, Yefta Christian^{2*}

* Sistem Informasi, Universitas Internasional Batam
herman@uib.ac.id¹, yefta@uib.ac.id²

Article Info

Article history:

Received 2022-11-22

Revised 2022-12-06

Accepted 2022-12-08

Keyword:

Academic Performance
Machine Learning,
Online Learning,
Algorithms.

ABSTRACT

Student academic performance is one of the important factors for graduation. Therefore, many studies have been conducted in the education field to identify factors that affect student performance. This research focuses on academic performance in online learning conditions by a case study at XYZ university. Data were collected using Machine Learning techniques with Distributed Random Forest model, Naïve Bayes, Generalized Linear Model, and Gradient Boosting Machine algorithms. The results of this study indicate that the Distributed Random Forest and Gradient Boosting Machine models have an average accuracy of 99.83%. Researcher found variables that affect student learning performance especially in online learning, are final exam scores, midterm scores, attendance, assignment scores, amount of material given, number of assignments given, and number of clicks on material. From these findings, Researcher recommends that in order to improve the next learning performance, it should focus on improving the implementation of the Final Exams and learning material on the platform.



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

I. PENDAHULUAN

Performa akademik mahasiswa menjadi salah satu faktor penting untuk kelulusan mahasiswa. Banyak penelitian yang dilakukan dibidang pendidikan untuk mengidentifikasi faktor yang mempengaruhi performa pembelajaran mahasiswa. Dengan mengembangkan metode untuk menganalisis, mendeteksi pola dan menyimpulkan perubahan pada meningkatkan pembelajaran mahasiswa, dan dapat diterapkan untuk memprediksi mahasiswa yang berisiko maupun pencegahan mahasiswa putus kuliah [1] dan pada penelitian [2] peneliti menggunakan analisis performa pembelajaran mahasiswa untuk mendorong mahasiswa agar mengubah perilaku kearah yang lebih baik sehingga performa pembelajaran mahasiswa dapat meningkat. Begitu juga dengan penelitian [3], yang menganalisa performa pembelajaran siswa untuk membantu siswa sekolah menengah atas yang berada pada tahun ketiga agar dapat meningkatkan nilai setidaknya cukup untuk minimal kelulusan, sehingga jalur akademik siswa berjalan tanpa ada gangguan dan pada penelitian [4], peneliti memprediksi performa akademik mahasiswa setelah terdaftar pada

universita, sehingga dapat menentukan mahasiswa yang akan berhasil begitu juga mahasiswa yang akan menerima skor buruk dan dari hasil prediksi tersebut dosen dapat memilih strategi pembelajaran yang cocok untuk mahasiswa.

Dalam melakukan pembelajaran di era modern, mayoritas Universitas sudah mengadopsi sistem informasi dalam manajemen pembelajarannya. Sistem tersebut dikenal dengan istilah *Learning Management System* (LMS). LMS dapat digunakan untuk menyimpan materi, mengumpulkan tugas dan melakukan asesmen sesuai dengan rubrik penilaian. Karena banyaknya penggunaan LMS, data pembelajaran menjadi tersimpan dengan rapi dan terstruktur, karena tidak lagi diatur dengan sistem konvensional oleh bidang akademik [5]. Penelitian pada bidang studi ini perlu untuk mengumpulkan data pembelajaran mahasiswa pada universitas, yang menjadi data terbaik untuk dikumpulkan tanpa hambatan. Data yang telah dikumpulkan, dapat digunakan untuk beberapa teknik seperti studi statistik atau teknik *Machine Learning*.

Penggunaan teknik *Machine Learning* memungkinkan untuk mempelajari sampel data untuk mengidentifikasi variable yang paling mempengaruhi performa akademik

mahasiswa, pada penelitian [6], peneliti menerapkan teknik *Machine Learning* untuk mengidentifikasi variable yang mempengaruhi performa akademik membantu universitas untuk menghilangkan atau meminimalkan faktor-faktor negatif bagi mahasiswa.

Pada penelitian [6] telah ditemukan beberapa faktor-faktor yang mempengaruhi performa akademik mahasiswa akan tetapi tidak melakukan analisis pada faktor jumlah tugas dan jumlah materi yang diberikan, dan penelitian hanya memberikan nilai akurasi tidak nilai AUC (*Area Under Curve*). Pada Universitas XYZ mulai aktif melaksanakan pendidikan daring saat konfirmasi kasus Covid-19 pada tanggal 2 Maret 2020, maka dari itu penting untuk Universitas XYZ untuk melakukan analisis pada faktor yang mempengaruhi performa akademik mahasiswa, sehingga Universitas XYZ dapat meningkatkan kualitas dari pembelajaran dimasa depan. Penelitian ini menerapkan penggunaan *Distributed Random Forest*, *Naïve Bayes*, *Generalized Linear Model*, dan *Gradient Boosting Machine* untuk mempelajari faktor-faktor yang mempengaruhi performa akademik mahasiswa pada Universitas XYZ dengan tambah faktor jumlah tugas, nilai tugas, dan jumlah materi yang diberikan, sehingga jumlah lulusan dapat ditingkatkan melalui peningkatan pada faktor-faktor yang mempengaruhi performa akademik mahasiswa.

II. METODE PELAKSANAAN

A. Knowledge Discovery in Databases

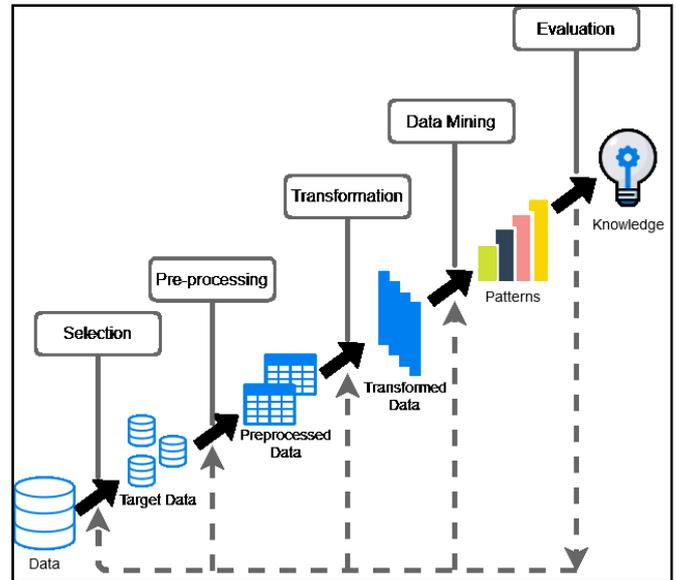
Dalam studi ini, peneliti menggunakan *knowledge discovery in database* untuk mengidentifikasi struktur pengetahuan yang valid, baru, berpotensi bermanfaat, dan pada akhirnya dapat dipahami. *Knowledge Discovery in Databases* (KDD) adalah proses mengidentifikasi struktur pengetahuan yang valid, baru, berpotensi bermanfaat dan pada akhirnya dapat dipahami, sedangkan *Knowledge Mining* mengacu pada langkah-langkah spesifik dalam proses ini [7]. Metode KDD adalah analisis data eksplorasi otomatis dari database besar [8].

Data mining adalah aplikasi khusus dari algoritma untuk mengekstraksi pola dari data [7]. Langkah-langkah tambahan dalam proses KDD yang dijelaskan pada Gambar 1, seperti pemrosesan data, pemilihan data, transformasi, penambangan data, dan interpretasi hasil penambangan data yang tepat, diperlukan untuk memastikan bahwa informasi yang berguna diekstraksi dari data.

Proses-proses KDD adalah sebagai berikut [7]:

1) *Data Selection*: Berdasarkan penelitian [7], *Data Selection* didefinisikan sebagai proses pemilihan set data, atau subset variabel atau sampel data, di mana penemuan harus dilakukan, sehingga menghasilkan target data yang akan digunakan pada proses selanjutnya. Penelitian [8] menyatakan bahwa *data selection* termasuk mencari tahu data apa yang tersedia, memperoleh data tambahan yang diperlukan dan kemudian mengintegrasikan semua data untuk penemuan pengetahuan ke dalam satu dataset, termasuk

atribut yang akan dipertimbangkan untuk proses tersebut. Data penelitian ini berasal dari data pembelajaran Universitas XYZ selama semester Genap 2019 / 2020 yang mulai aktif pembelajaran secara daring pada tanggal 2 Maret 2020 menggunakan Sistem Moodle, delapan (8) matakuliah yang dipilih merupakan matakuliah yang aktif dalam pembelajaran daring menggunakan Sistem Moodle.



Gambar 1. Fayyad's Knowledge Discovery in Databases [7].

Berikut tabel deskripsi variable yang digunakan pada penelitian ini:

TABEL I
DESKRIPSI VARIABEL

| Tipe Variabel | Variabel | Kategori | Deskripsi |
|-----------------|--------------------------|----------|---|
| Demographic | Jenis Kelamin | Nominal | Jenis kelamin pelajar (Pria = L, Wanita = P). |
| | Jurusan | Nominal | Jurusan / Program Studi pelajar. |
| | Umur | Rasio | Umur pelajar. |
| Courses | Jumlah Tugas Matakuliah | Rasio | Jumlah tugas yang diberikan pada mata pelajaran. |
| | Jumlah Materi Matakuliah | Rasio | Jumlah materi digital yang diberikan pada mata pelajaran. |
| Activity Record | Presensi | Rasio | Jumlah kehadiran pelajaran dalam pembelajaran. |
| | Jumlah Klik VLE | Rasio | Total jumlah klik materi digital pada |

| | | | |
|-----------------------------|-------------|---------|--|
| | | | mata pelajaran yang diambil pelajar. |
| | Nilai Tugas | Rasio | Rata – rata nilai tugas pelajar. |
| | Nilai UTS | Rasio | Nilai ujian tengah semester pelajar. |
| | Nilai UAS | Rasio | Nilai ujian akhir semester pelajar. |
| Academic Performance | Hasil Akhir | Nominal | Hasil lulus atau tidak dalam mata pelajaran yang diambil pelajar (lulus = Pass, tidak lulus = Fail). |

2) *Data Preprocessing*: Berdasarkan Penelitian [7], *Preprocessing* didefinisikan sebagai proses operasi dasar termasuk menghilangkan *noise*, mengumpulkan informasi yang diperlukan untuk memodelkan atau melakukan penghitungan, memutuskan strategi untuk menangani data yang hilang. Begitu juga dengan penelitian [8], yang menyatakan bahwa *preprocessing* merupakan tahap peningkatan keandalan data, yang meliputi pembersihan data, seperti menangani nilai yang hilang dan menghilangkan *noise* atau *outlier*. Proses *preprocessing* pada penelitian ini dilaksanakan menggunakan *tools node Knime*, berikut proses pada tahap *preprocessing*:

- Menghitung jumlah klik pada VLE
- Cek nilai kosong / *missing value check*
- Penggabungan / *Joiner*
- Penggantian nama kolom / *Column Rename*

3) *Data Transformation*: Berdasarkan Penelitian [7], *Transformation* didefinisikan sebagai proses menemukan fitur yang berguna untuk merepresentasikan data sesuai dengan tujuan. Menurut penelitian [8], *data transformation* adalah proses generasi data yang lebih baik untuk data mining disiapkan dan dikembangkan. Salah satu metode yang dapat digunakan pada tahap *transformation* adalah pengurangan dimensi, seperti pemilihan fitur dan ekstraksi serta pencatatan sampel. Pada penelitian ini tahapan *transformation* akan dimulai dengan data dari tahap *Preprocessing Data* akan dipilih kolom yang digunakan dan dinormalkan nilai-nilai datasetnya, dataset yang telah selesai diproses akan dipisahkan menjadi 2 bagian yaitu *training dataset* dan *testing dataset*, pembagian *training dataset* dan *testing dataset* akan diproses node *H2O Cross Validation Loop Start* dengan penerapan *N-Fold Crossvalidation* sebanyak *10-Fold*, tujuan dari *N-Fold* adalah untuk meningkatkan performa dari *Model Learning*.

4) *Data Mining*: Berdasarkan Penelitian [7], *Data Mining* didefinisikan sebagai proses untuk metode *data mining* tertentu. Misalnya, *summarization, classification, regression, clustering*, dll. Menurut penelitian [8], menyatakan bahwa tahap ini merupakan proses pemilihan metode khusus yang akan digunakan untuk mencari pola dan proses pengujian algoritma hingga hasil yang memuaskan diperoleh. Tahapan *Data Mining* pada penelitian ini, dimulai dengan *training dataset* yang telah tersedia akan diproses oleh masing-masing algoritma seperti model *Distributed Random Forest* dengan menggunakan node *H2O Random Forest Learner*, untuk model *Naive Bayes* menggunakan node *H2O Naive Bayes Learner*, untuk model *Generalized Linear Model* menggunakan node *H2O Generalized Linear Model Learner*, dan untuk model *Gradient Boosting Machine* menggunakan node *H2O Gradient Boosting Machine Learner*. Hasil *H2O classification* digunakan pada tahap *Model Testing* sebagai *classification* model dan node *H2O Random Forest Learner* juga menghasilkan *variable importance measures* yang merupakan hasil perhitungan *variable* yang menjadi faktor penting dalam proses *Model Learning*.

5) *Evaluation*: Tahap *evaluation* adalah proses mengevaluasi dan menafsirkan pola yang diekstraksi (aturan, keandalan, dll.) [8]. Tahap ini berfokus pada kelengkapan dan kegunaan dari model yang diinduksi. Pada titik ini, perlunya mendokumentasikan pengetahuan yang ditemukan untuk penggunaan lebih lanjut. *Evaluation* pada penelitian ini dimulai dengan menginterpretasi hasil proses klasifikasi dari model *Distributed Random Forest, Naive Bayes, Generalized Linear Model*, dan *Gradient Boosting Machine*. Evaluasi dilakukan dengan tujuan agar hasil dari tahap modelling sesuai dan menjawab faktor – faktor yang mempengaruhi performa akademik mahasiswa. Hasil proses dievaluasi melalui pendekatan *Precision, Recall*, dan perhitungan persentase dari AUC (*Area under the Receiver Operating Characteristics Curve*) merupakan model yang digunakan untuk mengukur kinerja pada hasil proses klasifikasi dari model *Distributed Random Forest, Naive Bayes, Generalized Linear Model*, dan *Gradient Boosting Machine*. Berdasarkan penelitian [9], [10] AUC menangkap kemampuan dari klasifikasi biner untuk memisahkan antara kelas sebagai ambang batas diskriminasi bervariasi.

B. Metode Analisis Hasil Pemodelan Data

Metode-metode yang digunakan pada penelitian ini bertujuan untuk menemukan indikator mahasiswa yang lulus atau gagal pada mata kuliah dan variabel yang paling mempengaruhi lulus atau gagalnya pelajar dalam melaksanakan pembelajaran, metode *preprocessing, tranformation, modeling*, dan *evaluation* sudah tersedia pada *analysis tools* yang digunakan yaitu aplikasi *Knime*. Untuk meningkatkan akurasi modeling penelitian ini menggunakan metode *k-Fold Cross Validation* dengan tingkat akurasi prediksi yang diukur dari nilai *Recall, Precision, Accuracy*,

F-Measure dan *AUC (Area under the Receiver Operating Characteristics Curve)*.

III. HASIL DAN PEMBAHASAN

Tujuan penelitian ini adalah untuk membuat model klasifikasi performa akademik mahasiswa pada pembelajaran khususnya pembelajaran secara daring dengan penerapan model klasifikasi *Distributed Random Forest*, *Naïve Bayes*, *Generalized Linear Model*, dan *Gradient Boosting Machine*. Data yang digunakan berasal dari data pembelajaran Universitas XYZ yang telah diperoleh dari log pembelajaran dengan Sistem Moodle yang berjumlah 1,792 data, pada proses *N-fold Crossvalidation* dataset akan dibagi menjadi *training dataset* sebanyak 1621 data (90%) dan *testing dataset* sebanyak 171 data (10%).

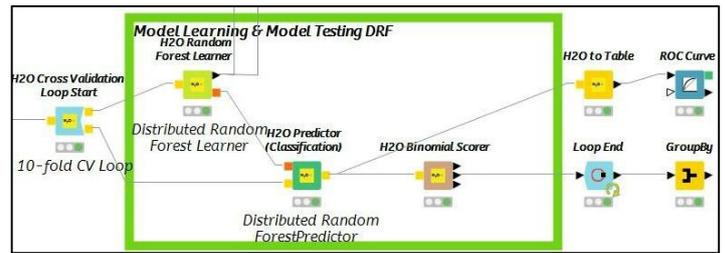
A. Deskripsi Data

Data dari pembelajaran Universitas XYZ terdiri dari 3 dataset yaitu: dataset matakuliah, dataset pembelajaran, dan dataset log pembelajaran. Data hasil akhir pada penelitian ini terbagi menjadi 2 (dua) yaitu: Lulus (*Pass*) dan Tidak Lulus (*Fail*), hasil pembelajaran dipengaruhi oleh beberapa faktor dan pada penelitian ini terdapat 10 faktor yang digunakan, faktor-faktor tersebut adalah: jenis kelamin, umur, jurusan, jumlah tugas matakuliah, jumlah materi matakuliah, presensi, jumlah klik VLE, nilai tugas, nilai UTS, dan nilai UAS.

B. Pengujian Model

Metode evaluasi pada penelitian ini menggunakan *N-Fold Cross-validation*[11]. Prosedur pada *Crossvalidation* adalah *resampling* yang dibagi menjadi N himpunan yang memiliki ukuran yang sama, yang disebut sebagai "*folds*". *N-Fold Cross-validation* kemudian jalankan eksperimen N. Di setiap pengujiannya salah satu himpunan bagian N dihilangkan, sehingga hanya akan digunakan untuk pengujian (ini menjamin bahwa, dalam setiap menjalankan, set pengujian yang digunakan berbeda-beda). *N-Fold Cross-validation* dapat digunakan untuk mengevaluasi kinerja *machine learning* dengan pembagian dataset menjadi *training dataset* dan *testing dataset*, dengan perbandingan 90% training dataset dan 10% testing dataset. Jumlah data yang digunakan sebanyak 1792 dengan 10 *folds*. Peneliti menggunakan 10 *folds* karena berdasarkan penelitian [12], angka optimal pada jenis penelitian sejenis adalah 10 *folds*.

6) *Model Distributed Random Forest*, pengujian model *Distributed Random Forest* dimulai dengan node *H2O Random Forest Learner* memproses *training dataset* dengan algoritma *Random Forest*. Berikut node *Knime* pada model *Distributed Random Forest*:



Gambar 2. Node Knime Distributed Random Forest Model

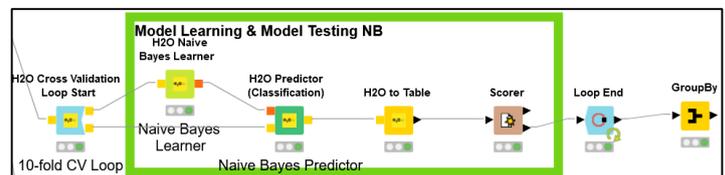
Berdasarkan data hasil pengujian model *Distributed Random Forest*, berikut nilai *recall*, *precision*, *accuracy* dan *AUC* dari *N-Fold Crossvalidation* sebanyak 10 *folds*, seperti yang di tunjukkan pada Tabel 2.

TABEL II
HASIL PENGUJIAN MODEL DISTRIBUTED RANDOM FOREST

| Fold | Accuracy | AUC | f1-score | Precision | Recall |
|----------------|---------------|----------|--------------|--------------|--------------|
| 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 | 1 | 1 |
| 4 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1 | 1 | 1 | 1 | 1 |
| 6 | 1 | 1 | 1 | 1 | 1 |
| 7 | 0.995 | 1 | 0.899 | 0.978 | 0.834 |
| 8 | 0.989 | 1 | 0.914 | 0.994 | 0.857 |
| 9 | 1 | 1 | 1 | 1 | 1 |
| 10 | 1 | 1 | 1 | 1 | 1 |
| Average | 0.9983 | 1 | 0.981 | 0.997 | 0.969 |

Berdasarkan Tabel 2. model *Distributed Random Forest* mendapatkan nilai rata – rata *Accuracy* sebesar 99,8% dan nilai rata – rata *AUC (Area under the Receiver Operating Characteristics Curve)* sebesar 96,3%.

7) *Model Naïve Bayes*, pengujian model *Naïve Bayes* dimulai dengan node *H2O Naïve Bayes Learner* memproses *training dataset* dengan algoritma *Naïve Bayes*. Berikut node *Knime* pada model *Naïve Bayes*:



Gambar 3. Node Knime Naïve Bayes

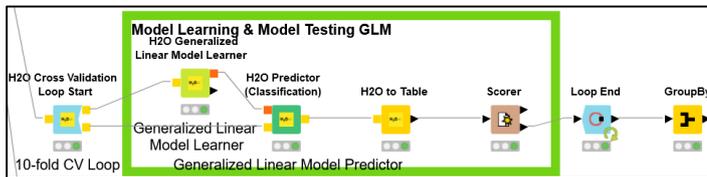
Berdasarkan data hasil pengujian model *Naïve Bayes*, berikut nilai *recall*, *precision*, *accuracy* dan *AUC* dari *N-Fold Crossvalidation* sebanyak 10 *folds*, seperti yang di tunjukkan pada Tabel 3.

TABEL III
HASIL PENGUJIAN MODEL NAÏVE BAYES

| Fold | Accuracy | AUC | f1-score | Precision | Recall |
|----------------|--------------|----------|--------------|--------------|--------------|
| 1 | 0.994 | 1 | 0.927 | 0.997 | 0.875 |
| 2 | 1 | 1 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 | 1 | 1 |
| 4 | 0.995 | 1 | 0.927 | 0.997 | 0.875 |
| 5 | 1 | 1 | 1 | 1 | 1 |
| 6 | 0.988 | 1 | 0.897 | 0.994 | 0.834 |
| 7 | 1 | 1 | 1 | 1 | 1 |
| 8 | 0.989 | 1 | 0.897 | 0.995 | 0.834 |
| 9 | 1 | 1 | 1 | 1 | 1 |
| 10 | 1 | 1 | 1 | 1 | 1 |
| Average | 0.997 | 1 | 0.965 | 0.998 | 0.942 |

Berdasarkan Tabel 3. model *Naïve Bayes* mendapatkan nilai rata – rata *Accuracy* sebesar 99,7% dan nilai rata – rata *AUC (Area under the Receiver Operating Characteristics Curve)* sebesar 93%.

8) *Model Generalized Linear Model*, pengujian model *Generalized Linear Model* dimulai dengan node *H2O Generalized Linear Model Learner* memproses *training dataset* dengan algoritma *Generalized Linear Model*. Berikut node *Knime* pada model *Generalized Linear Model*:



Gambar 4. Node Knime Generalized Linear Model

Berdasarkan data hasil pengujian model *Generalized Linear Model*, berikut nilai *recall*, *precision*, *accuracy* dan *AUC* dari *N-Fold Crossvalidation* sebanyak 10 *folds*, seperti yang di tunjukkan pada Tabel 4.

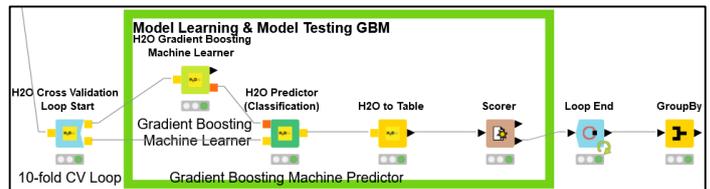
TABEL IV
HASIL PENGUJIAN MODEL GENERALIZED LINEAR MODEL

| Fold | Accuracy | AUC | f1-score | Precision | Recall |
|------|----------|-----|----------|-----------|--------|
| 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 | 1 | 1 |
| 4 | 0.995 | 1 | 0.927 | 0.997 | 0.875 |
| 5 | 0.994 | 1 | 0.960 | 0.997 | 0.929 |
| 6 | 0.988 | 1 | 0.897 | 0.994 | 0.834 |
| 7 | 1 | 1 | 1 | 1 | 1 |
| 8 | 0.983 | 1 | 0.829 | 0.992 | 0.750 |

| | | | | | |
|----------------|--------------|----------|--------------|--------------|--------------|
| 9 | 1 | 1 | 1 | 1 | 1 |
| 10 | 1 | 1 | 1 | 1 | 1 |
| Average | 0.996 | 1 | 0.961 | 0.998 | 0.939 |

Berdasarkan Tabel 4. model *Generalized Linear Model* mendapatkan nilai rata – rata *Accuracy* sebesar 99,6% dan nilai rata – rata *AUC (Area under the Receiver Operating Characteristics Curve)* sebesar 92,3%.

9) *Model Gradient Boosting Machine*, pengujian model *Gradient Boosting Machine* dimulai dengan node *H2O Gradient Boosting Machine Learner* memproses *training dataset* dengan algoritma *Gradient Boosting Machine*. Berikut node *Knime* pada model *Gradient Boosting Machine*:



Gambar 5. Node Knime Gradient Boosting Machine

Berdasarkan data hasil pengujian model *Gradient Boosting Machine*, berikut nilai *recall*, *precision*, *accuracy* dan *AUC* dari *N-Fold Crossvalidation* sebanyak 10 *folds*, seperti yang di tunjukkan pada Tabel 5.

TABEL V
HASIL PENGUJIAN MODEL GRADIENT BOOSTING MACHINE

| Fold | Accuracy | AUC | f1-score | Precision | Recall |
|----------------|---------------|----------|--------------|--------------|--------------|
| 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 | 1 | 1 |
| 4 | 1 | 1 | 1 | 1 | 1 |
| 5 | 1 | 1 | 1 | 1 | 1 |
| 6 | 0.994 | 1 | 0.953 | 0.997 | 0.917 |
| 7 | 1 | 1 | 1 | 1 | 1 |
| 8 | 0.989 | 1 | 0.897 | 0.995 | 0.834 |
| 9 | 1 | 1 | 1 | 1 | 1 |
| 10 | 1 | 1 | 1 | 1 | 1 |
| Average | 0.9983 | 1 | 0.985 | 0.999 | 0.975 |

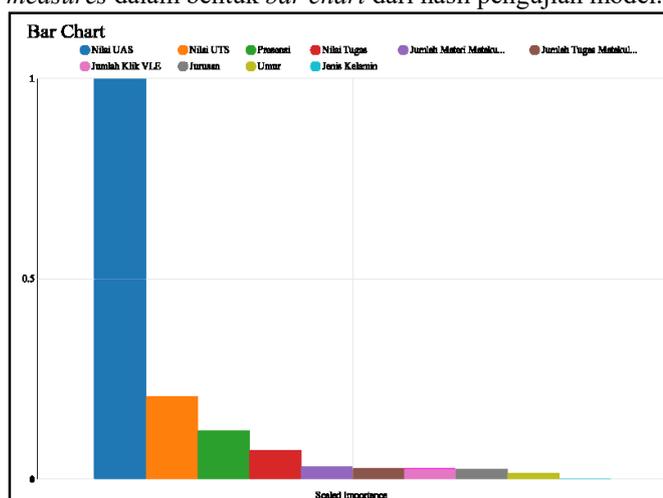
Berdasarkan Tabel 5. model *Gradient Boosting Machine* mendapatkan nilai rata – rata *Accuracy* sebesar 99,83% dan nilai rata – rata *AUC (Area under the Receiver Operating Characteristics Curve)* sebesar 97%.

C. *Variable Importance Measures*

Dari hasil pengujian yang dilakukan *weight of parameters* atau *variable importance measures* yang ditemukan sebagai berikut:

1. Nilai UAS
2. Nilai UTS
3. Presensi
4. Nilai Tugas
5. Jumlah Materi Matakuliah
6. Jumlah Tugas Matakuliah
7. Jumlah Klik VLE
8. Jurusan
9. Umur
10. Jenis Kelamin

Pada Gambar 6 menggambarkan hasil *variable importance measures* dalam bentuk *bar chart* dari hasil pengujian model.



Gambar 6. Variable Importance Measures Chart

IV. SIMPULAN DAN SARAN

A. Kesimpulan

Hasil Penelitian ini menunjukkan bahwa performa pembelajaran menurut analisis data yang diperoleh secara signifikan berasal dari nilai UAS dan oleh karena itu dalam kondisi daring, sangat perlu diperhatikan pelaksanaan UAS itu sendiri. Perlu adanya strategi untuk menghindari kecurangan dalam pelaksanaan UAS agar hasil belajar yang didapatkan sungguh-sungguh menggambarkan kondisi hasil pembelajaran yang telah dilakukan. Selain itu juga presensi juga menjadi faktor yang cukup mempengaruhi performa, hal ini seperti yang didapatkan pada hasil penelitian [1].

Hasil berikutnya adalah ketika variabel Nilai UAS, Nilai UTS, Presensi, dan Nilai Tugas dihapus, maka ditemukan variabel signifikan pada keaktifan mahasiswa dalam berinteraksi dengan kegiatan pembelajaran. Pembelajaran dengan kondisi daring membuat kegiatan pembelajaran oleh mahasiswa tidak dapat dipantau oleh dosen secara langsung, sehingga tingkat kemandirian dalam belajar menjadi faktor yang sangat penting dalam pembelajaran secara daring [6]. Melihat faktor Jumlah Klik VLE yang sangat dominan menunjukkan bahwa keaktifan mahasiswa dalam berinteraksi dengan platform pembelajaran online melalui materi dan kegiatan yang tersedia menjadi sangat penting dalam

membangun performa akademik yang baik. Student engagement memang menjadi salah satu faktor kunci keberhasilan pembelajaran secara online menurut beberapa ahli [6], [13], [14]. namun dalam hal ini tingkat ketertarikan mahasiswa terhadap platform menjadi lebih dominan.

Hasil pengujian *AUC* dan *Accuracy* pada model *Distributed Random Forest*, *Naïve Bayes*, *Generalized Linear Model*, dan *Gradient Boosting Machine* terbukti dapat mengklasifikasikan performa pembelajaran berdasarkan 10 variabel yang telah di uji dan mendapatkan tingkat akurasi dan nilai *AUC* (*Area under the Receiver Operating Characteristics Curve*) yang tinggi pada model *Distributed Random Forest* dan *Gradient Boosting Machine*, yaitu *accuracy* dengan nilai 99,83% dan *AUC* dengan nilai 100%.

Penelitian ini membuktikan bahwa model *Distributed Random Forest* memberikan hasil yang baik untuk *Machine Learning* terhadap dataset performa pembelajaran sebagaimana penelitian yang dilakukan pada penelitian [10]. Penelitian ini juga menunjukkan bahwa Jumlah Klik VLE menjadi variabel yang sangat berpengaruh apabila Nilai UAS, Nilai UTS, Presensi, dan Nilai Tugas dihilangkan dari *Learner node*, sejalan dengan penelitian sebelumnya [6].

Ada beberapa saran untuk penelitian lebih lanjut, yaitu:

1. Model *Distributed Random Forest* dan *Gradient Boosting Machine* dapat digunakan pada dataset hasil performa pembelajaran lainnya, terutama pada dataset performa pembelajaran pada SPADA (Sistem Pembelajaran Daring Indonesia)
2. Penerapan platform *Hadoop* untuk penyimpanan dataset dan penambahan secara real time, sehingga implementasi *machine learning* dapat dijalankan secara periodik tanpa kendala pada server.
3. Rekomendasi untuk pelaksanaan pembelajaran secara daring yang akan datang adalah memperhatikan proses pelaksanaan Ujian Akhir Semester dan pengaturan materi didalam platform yang lebih menarik bagi mahasiswa.
4. Untuk Strategi pelaksanaan UAS dan pengaturan materi yang optimal diperlukan studi lebih lanjut.

DAFTAR PUSTAKA

- [1] K. Sunday, P. Ocheja, S. Hussain, S. S. Oyelere, B. O. Samson, and F. J. Agbo, "Analyzing Student Performance in Programming Education Using Classification Techniques," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 15, no. 02, p. 127, 2020, doi: 10.3991/ijet.v15i02.11527.
- [2] S. Sinche *et al.*, "Analysis of Student Academic Performance Using Human-in-the-Loop Cyber-Physical Systems," *Telecom 2020, Vol. 1, Pages 18-31*, vol. 1, no. 1, pp. 18–31, 2020, doi: 10.3390/TELECOM1010003.
- [3] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. van Erven, "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil," *J Bus Res*, vol. 94, no. August 2017, pp. 335–343, 2019, doi: 10.1016/j.jbusres.2018.02.012.

- [4] L. H. Son and H. Fujita, "Neural-fuzzy with representative sets for prediction of student performance," *Applied Intelligence*, vol. 49, no. 1, pp. 172–187, 2019, doi: 10.1007/s10489-018-1262-7.
- [5] M. Norouzi *et al.*, "Model of Learning Management System for Self-Directed Learning," *ISSN International Journal of Information Dissemination and Technology International Journal of Information Dissemination and Technology*, vol. 4, no. 4, pp. 2229–5984, 2014.
- [6] A. Rivas, J. M. Fraile, P. Chamoso, A. González-Briones, S. Rodríguez, and J. M. Corchado, "Students Performance Analysis Based on Machine Learning Techniques," in *Learning Technology for Education Challenges*, 2019, pp. 428–438.
- [7] U. Fayyad, "Knowledge Discovery In Databases : An Overview," in *Relational Data Mining*, Springer, 2001, pp. 28–47.
- [8] R. Lior and O. Maimon, *Data mining with decision trees: theory and applications*, 2nd ed., vol. 81. Singapore: World Scientific Publishing Co. Pte. Ltd, 2014.
- [9] A. J. Bowers and X. Zhou, "Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes," *J Educ Stud Placed Risk*, vol. 24, no. 1, pp. 20–46, 2019, doi: 10.1080/10824669.2018.1523734.
- [10] N. I. Jha, I. Ghergulescu, and A. N. Moldovan, "OULAD MOOC dropout and result prediction using ensemble, deep learning and regression techniques," *CSEdu 2019 - Proceedings of the 11th International Conference on Computer Supported Education*, vol. 2, no. Csedu, pp. 154–164, 2019, doi: 10.5220/0007767901540164.
- [11] M. Kubat, *An Introduction to Machine Learning Second Edition*. 2017.
- [12] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?," *Comput Stat*, no. 0123456789, 2020, doi: 10.1007/s00180-020-00999-9.
- [13] D. C. Duygu, N. Alkiş, and S. Ozkan-Yildirim, "A structural model for students' adoption of Learning Management Systems: An empirical investigation in the higher education context," *Educational Technology and Society*, vol. 21, no. 2, pp. 13–27, 2018.
- [14] M. A. López Garrido, E. Y. Morales Mateos, J. A. Hernández Aguilar, C. A. Ochoa Ortiz, C. González Constantino, and O. A. González González, "Patterns of Motivational Orientation and its Relationship with Academic Performance in University Students," *Research in Computing Science*, vol. 148, no. 6, pp. 197–203, 2019, doi: 10.13053/rics-148-6-14.