

TINJAUAN PUSTAKA SISTEMATIS: PREDIKSI PRESTASI BELAJAR PESERTA DIDIK DENGAN ALGORITMA PEMBELAJARAN MESIN

Sabar Aritonang Rajagukguk
Sekolah Cendekia Harapan
Jimbaran, Kuta Selatan
Bali
sabar@cendekiaharapan.sch.id

ABSTRAKSI

Penggunaan data profil dan capaian pembelajaran peserta didik dengan pembelajaran mesin untuk memprediksi prestasi belajar peserta didik akan memberi kontribusi penting bagi penyusunan strategi yang tepat dalam peningkatan prestasi belajar. Artikel ini menampilkan tinjauan pustaka sistematis (*systematic literature review*) terhadap 15 artikel penelitian yang mengimplementasikan algoritma pembelajaran mesin untuk memprediksi prestasi belajar peserta didik. Hasilnya adalah pembelajaran mesin relevan digunakan dalam memprediksi prestasi belajar peserta didik; sementara empat algoritma yang paling sering digunakan untuk memprediksi prestasi belajar peserta didik adalah *ANN*, *Naïve Bayes*, *Logistic Regression*, *SVM* dan *Decision Tree*. Temuan lain adalah bahwa pemangku kepentingan yang paling mendapat manfaat dari penelitian sejenis adalah lembaga pendidikan, dan jenis algoritma pembelajaran mesin yang paling banyak digunakan adalah jenis algoritma untuk klasifikasi.

Kata Kunci

prestasi belajar peserta didik; algoritma pembelajaran mesin; tinjauan pustaka sistematis.

1. PENDAHULUAN

Sebagaimana salah tujuan akhir pembelajaran adalah prestasi belajar, prediksi prestasi belajar peserta didik menjadi urgen untuk dilakukan. Prediksi prestasi belajar peserta didik dapat dijadikan sebagai dasar untuk melakukan intervensi dini terhadap potensi kegagalan peserta didik untuk mencapai tujuan pembelajaran; dan di saat yang sama dapat melakukan perubahan strategi pembelajaran agar dapat memfasilitasi keragaman peserta didik [1].

Prestasi belajar peserta didik yang dimaksud dalam artikel ini adalah *students' performance*. Namun secara kontekstual, *students' performance* tidak dapat direduksi hanya menjadi prestasi belajar, namun juga mencakup hasil belajar. Hasil belajar adalah wujud konkret dari potensi yang dimiliki peserta didik [2], sebuah luaran dari proses perubahan tingkah laku setelah mengikuti pembelajaran [3]. Wujud konkret hasil belajar dapat dilihat dari pemahamannya terhadap ilmu yang dipelajari,

kepiawaiannya mengolah informasi dan mengambil keputusan yang didasarkan atas pemikiran tertentu atau pun keterampilan motoriknya [4].

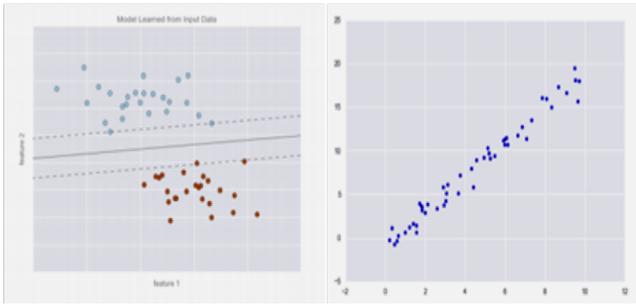
Prestasi belajar adalah ukuran pencapaian peserta didik setelah mengikuti serangkaian kegiatan pembelajaran berupa skala penilaian [5]. Pemberian bobot atau skala penilaian untuk prestasi belajar menuntut disusunnya indikator penilaian, serta perlu dipastikan validitas dan reliabilitasnya [6]. Nilai tersebutlah yang kemudian digunakan sebagai gambaran prestasi belajar dalam kurun waktu tertentu [7].

Data peserta didik pada lembaga-lembaga pendidikan yang dapat diolah untuk melakukan prediksi prestasi belajar saat ini semakin banyak. Pesatnya digitalisasi dan perubahan pola pembelajaran dan tata kelola lembaga pendidikan dari yang sebelumnya berbasis luring menjadi berbasis daring turut mendorong melimpahnya data peserta didik dan bersifat sangat dinamis seperti data perilaku, jenis dan frekuensi aktivitas pembelajaran peserta didik yang sangat berguna untuk memprediksi prestasi belajar. Pemrograman eksplisit (*explicit programming*) tidak lagi memadai pada beberapa kasus untuk melakukan prediksi prestasi belajar peserta didik dalam kondisi data yang sangat besar dan dinamis ini. Dalam hal ini, pembelajaran mesin mengisi kekosongan tersebut.

Pembelajaran mesin (*machine learning*) pada prinsipnya bertujuan membangun model matematika untuk membantu mesin memahami data. Model matematika ini terdiri dari berbagai variabel yang bisa berubah-ubah, sehingga program yang dibentuk dengan model matematika tersebut memungkinkan program beradaptasi dengan perubahan data [8]. Secara garis besar, algoritma pembelajaran mesin dapat dibagi menjadi dua tipe, yaitu *supervised learning* dan *unsupervised learning*.

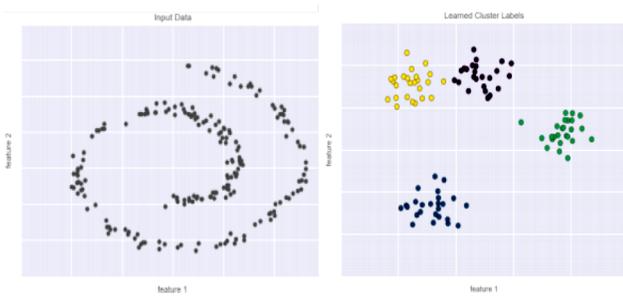
Supervised learning adalah kategori dalam pembelajaran mesin yang dilakukan dengan cara memberi label pada *input* dan *output* yang diharapkan. Setelahnya, algoritma yang digunakan dalam *supervised learning* akan menghasilkan *output* tanpa intervensi manusia semisal modifikasi pemrograman. [9]. *Supervised learning* dibagi menjadi dua jenis, yaitu *classification* dan *regression*. Tujuan dari

classification adalah untuk memprediksi kelas pada data yang dilabel, di mana sudah disediakan pilihan dari daftar kemungkinan yang ada. Sedangkan *regression* bertujuan untuk memprediksi gambaran hubungan regresi dari data, yakni sebuah *floating-point number* (angka dalam rentang yang sudah ditentukan sebelumnya), misalnya prediksi penghasilan tahunan seseorang berdasarkan jenjang pendidikan, umur, dan lokasi tempat tinggalnya [9] (lihat Gambar 1 sebagai contoh visualisasi kasus klasifikasi dan regresi linear).



Gambar 1. Visualisasi Data untuk Kasus Klasifikasi dan Regresi Linear.

Tipe kedua dalam pembelajaran mesin adalah *unsupervised learning*. Pada *unsupervised learning*, tidak ada pemberian label (seperti yang dilakukan pada *supervised learning*) pada semua input data. Algoritma dalam *unsupervised learning* akan membaca data apa adanya tanpa ada kaitan dengan label yang diberikan sehingga program akan menarik informasi langsung dari data yang ada [8]. Secara garis besar, ada dua hal yang dapat dilakukan dengan *unsupervised learning* yakni dengan melakukan transformasi *dataset* dan *clustering*. Transformasi *dataset* dilakukan dengan algoritma yang memungkinkan program untuk mengubah *dataset* yang ada menjadi *dataset* yang lebih mudah dipahami. Sedangkan dalam *clustering*, algoritma akan melakukan pemisahan data menjadi beberapa klaster berdasarkan kesamaan jenis data (lihat Gambar 2 sebagai contoh visualisasi data transformasi *dataset* dan *clustering*) (lihat Gambar 2 sebagai contoh visualisasi kasus transformasi *dataset* dan *clustering*) [9].



Gambar 2. Visualisasi Data untuk Kasus Transformasi Dataset dan Clustering

Penelitian ini bertujuan untuk mengkaji secara sistematis sejauh mana implementasi algoritma pembelajaran mesin

dan pemodelan telah dilakukan dalam memprediksi prestasi belajar peserta didik. Hingga Mei 2021, penelitian serupa belum ditemukan pernah dilakukan berdasarkan pencarian pada *database* IEEE Access dan Science Direct. Sehingga artikel ini diharapkan dapat memberi kontribusi ilmiah bagi para peneliti dalam area kajian prediksi prestasi belajar peserta didik, bidang ilmu komputer, dan statistika (atau bidang sains data).

Diharapkan hasil dari penelitian tinjauan pustaka sistematis ini dapat memberikan *insight* algoritma pembelajaran mesin yang paling efektif untuk memprediksi prestasi belajar siswa dan meningkatkan pemahaman para pemangku kepentingan tentang implementasi pembelajaran mesin untuk memprediksi prestasi belajar siswa.

2. METODE

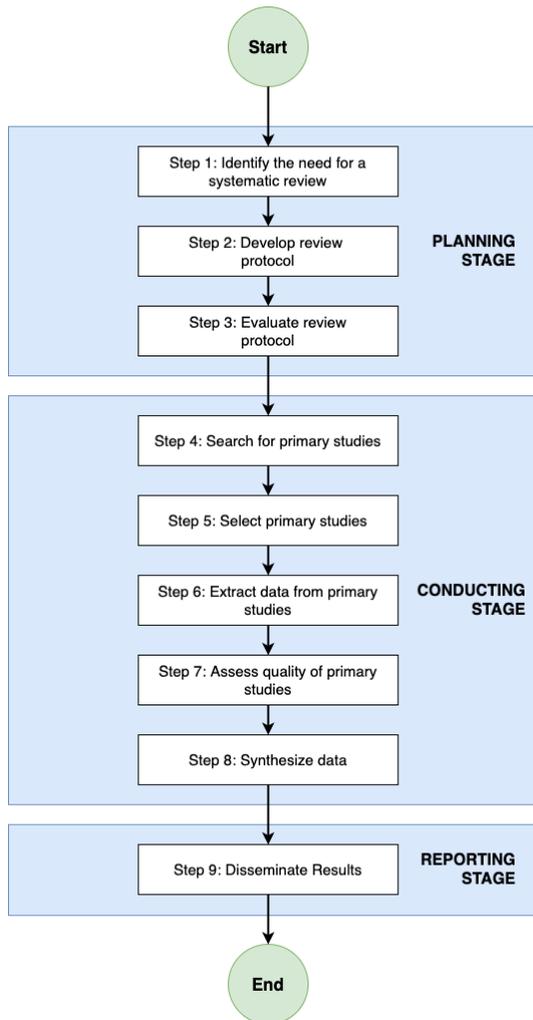
Penelitian ini menggunakan metode tinjauan pustaka sistematis. Metode ini dipilih berdasarkan pertimbangan bahwa belum tersedia tinjauan pustaka sistematis tentang implementasi algoritma pembelajaran mesin untuk memprediksi prestasi belajar peserta didik hingga Mei 2021 pada *database* IEEE Access dan Science Direct. Tinjauan pustaka sistematis ini juga diharapkan dapat menjadi stimulus untuk penelitian lanjutan dalam bidang sains data untuk berkontribusi pada bidang pendidikan (sebagaimana data pendidikan juga dicakup dalam peta jalan penelitian UII) [10]. Pertanyaan penelitian disusun sebelum penelitian dilakukan dan didasarkan pada kriteria *Population, Intervention, Comparison, Outcomes* dan *Context* (PICOC) [11] diikuti dengan protokol tinjauan pustaka sistematis untuk dapat diuji reliabilitasnya.

2.1 Tinjauan Pustaka Sistematis

Tinjauan pustaka sistematis adalah sebuah metode kajian pustaka yang bertujuan menjawab pertanyaan penelitian dengan melakukan identifikasi, pengkajian, evaluasi dan interpretasi seluruh temuan terkait topik penelitian. Tinjauan pustaka sistematis telah terbukti merupakan metode penelitian yang efektif untuk memberikan gambaran mengenai tren topik penelitian tertentu baik hasil, metodologi maupun cakupan bidang penelitian sebelumnya [11].

Metode tinjauan pustaka sistematis dilakukan dengan mengikuti protokol tertentu. Adapun tahapan pelaksanaan kajian pustaka sistematis dimulai dengan memformulasikan pertanyaan penelitian dan membuat protokol tinjauan pustaka sistematis. Protokol tinjauan pustaka yang dimaksud adalah menentukan kata kunci apa yang hendak digunakan dalam pencarian artikel pada *database* dan apa saja kriteria yang hendak dicakup (kriteria inklusi) dan apa saja kriteria artikel yang tidak digunakan (kriteria eksklusi). Setelah artikel yang melalui protokol dikumpulkan, maka tahap berikutnya adalah mengidentifikasi literatur yang sesuai, melakukan pemilihan dari studi primer, melakukan ekstraksi data dan melakukan penilaian terhadap kualitas hasil-hasil penelitian yang didapatkan. Pada tahap akhir, dilakukan sintesis dari

berbagai hasil penelitian dan tahap penulisan artikel tinjauan pustaka sistematis dapat dilakukan (lihat Gambar 3)



Gambar 3. Tahapan Tinjauan Pustaka Sistematis

2.2 Pertanyaan Penelitian

Tujuan utama dari tinjauan pustaka sistematis ini adalah untuk mengkaji secara sistematis sejauh mana implementasi algoritma pembelajaran mesin dan pemodelan telah dilakukan dalam memprediksi prestasi belajar peserta didik. Untuk mencapai tujuan tersebut, pertanyaan-pertanyaan penelitian disusun dalam Tabel 1.

Empat pertanyaan penelitian yang disusun berdasarkan kriteria *Population*, *Intervention*, *Comparison*, *Outcomes* dan *Context* (PICOC). Tabel 2 di bawah ini menunjukkan struktur PICOC [11].

Tabel 1. Pertanyaan Penelitian

Pertanyaan	Motivasi
Apa relevansi pembelajaran mesin	Mengidentifikasi relevansi implementasi algoritma

Pertanyaan	Motivasi
dalam prediksi prestasi belajar peserta didik?	pembelajaran mesin dalam memprediksi prestasi belajar peserta didik.
Apa saja algoritma atau teknik pembelajaran mesin yang diimplementasikan dalam memprediksi prestasi belajar peserta didik?	Mengidentifikasi algoritma-algoritma pembelajaran mesin yang diimplementasikan dalam memprediksi prestasi belajar peserta didik.
Apa hasil dan simpulan yang didapatkan dari implementasi ini?	Mengidentifikasi luaran dari implementasi algoritma pembelajaran mesin dalam memprediksi prestasi belajar peserta didik.
Siapa pemangku kepentingan yang mendapatkan manfaat dari penelitian yang dilakukan?	Mengidentifikasi pemangku kepentingan yang paling mendapat manfaat dari penelitian yang dilakukan.

Tabel 2. Kriteria PICOC

Population	Prestasi belajar peserta didik, hasil belajar peserta didik, <i>students' performance</i>
Intervention	Prediksi prestasi belajar peserta didik menggunakan algoritma pembelajaran mesin
Comparison	-
Outcomes	Algoritma-algoritma yang digunakan dalam memprediksi prestasi belajar peserta didik.
Context	Penelitian yang dilakukan untuk prestasi belajar peserta didik di lembaga pendidikan formal dan pendidikan non-formal baik jenjang pendidikan dasar hingga pendidikan tinggi.

2.3 Pengumpulan Pustaka

Penelitian dilakukan awal hingga pertengahan Mei 2021. Aturan baku dalam pencarian *database* dilakukan sesuai dengan kriteria eksklusi dan inklusi, dan hasil pencarian dikumpulkan dan disusun dalam perangkat lunak *Mendeley*.

Pencarian Database. Artikel yang dikumpulkan bersumber dari database IEEE Access dan Science Direct.

Kata Kunci Pencarian. Kata kunci yang digunakan dalam pencarian adalah "*student's performance OR students' performance*" AND "*machine learning algorithm*" yang diisi pada kolom *advanced search*.

Kriteria Inklusi dan Eksklusi. Pencarian tersebut kemudian diseleksi berdasarkan tahun terbit, jenis artikel dan abstraknya. Artikel yang digunakan sebagai referensi adalah artikel penelitian yang terbit pada rentang 2018-2021 dan ditulis dalam bahasa Inggris. Kriteria eksklusinya adalah publikasi yang tidak tersedia dalam full text, tidak dalam Bahasa Inggris dan tidak membahas prediksi students' performance berdasarkan algoritma pembelajaran mesin. Student's performance yang dicakup adalah students' performance untuk semua jenjang pendidikan dari pendidikan dasar hingga pendidikan tinggi.

2.4 Analisis Pustaka

Pada tahap ini setiap artikel penelitian yang dipilih menjadi sampel dalam tinjauan pustaka sistematis dikelompokkan berdasarkan pertanyaan penelitian. Selain menjawab pertanyaan penelitian, artikel ilmiah juga diklasifikasikan berdasarkan beberapa kriteria tambahan. Maka, konten pustaka yang dianalisis adalah relevansi algoritma pembelajaran mesin dalam memprediksi prestasi belajar peserta didik, algoritma-algoritma yang diimplementasikan dalam memprediksi prestasi belajar peserta didik, hasil dari implementasi algoritma pembelajaran mesin terhadap prediksi prestasi belajar peserta didik, pemangku kepentingan yang mendapat manfaat dari implementasi algoritma pembelajaran mesin terhadap prediksi prestasi belajar peserta didik, tujuan implementasi pembelajaran mesin pada penelitian yang dilakukan dan jenis data yang dianalisis.

Pustaka tersebut perlu dianalisis untuk dapat memberikan pemahaman bagi para pemangku kepentingan akan bagaimana implementasi pembelajaran mesin untuk memprediksi prestasi belajar siswa dan menemukan algoritma yang paling efektif digunakan untuk memprediksi prestasi belajar peserta didik.

3. HASIL

3.1 Artikel yang Ditemukan

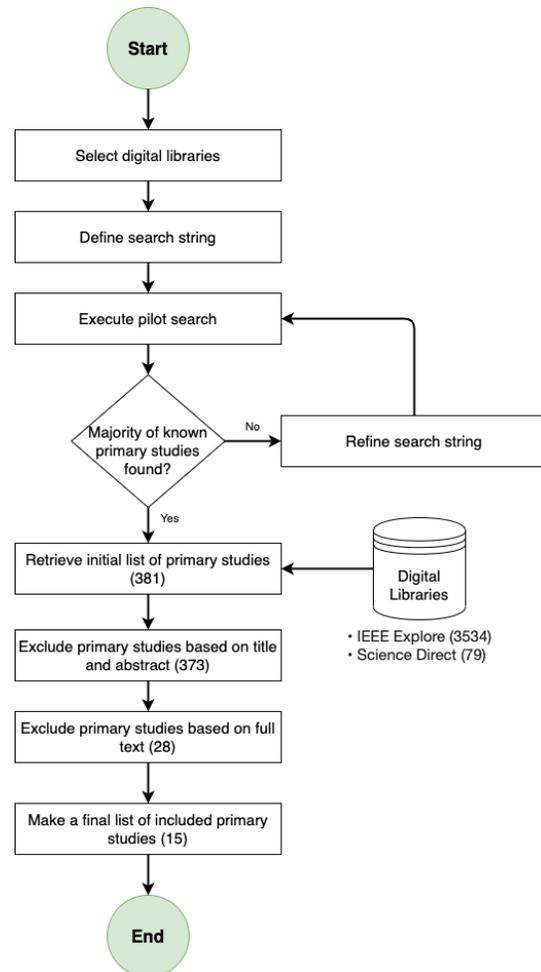
Penelitian ini dilakukan pada awal Mei 2021. Pada pencarian menggunakan kata kunci "*student's performance OR students' performance*" AND "*machine learning algorithm*" yang terbit pada 2018-2021, ditemukan sebanyak 3.613 artikel yang mencakup artikel jurnal ilmiah, buku, majalah, *proceeding* (3.534 artikel pada *database* IEEE Access dan 79 artikel pada *database* Science Direct). Setelah dilakukan seleksi berdasarkan artikel jurnal ilmiah saja, dengan melakukan *exclusion* publikasi artikel pada buku, majalah dan *proceeding* serta memilih hanya artikel pada jurnal yang berkaitan langsung dengan pendidikan dan *machine learning* saja serta memilih artikel dengan akses terbuka untuk *full text*, didapatkan 381 artikel (302 artikel pada *database* IEEE Access dan 71 artikel pada *database* Science Direct).

3.2 Seleksi Artikel dengan Kriteria Inklusi

Proses seleksi dimulai dari 381 artikel penelitian. Judul dan abstrak dibaca dengan seksama untuk dapat menentukan artikel mana yang sesuai dan tidak sesuai dengan kriteria

inklusi. Setelah melakukan pembacaan atas judul dan abstrak, terdapat sebanyak 373 artikel penelitian yang memenuhi kriteria.

Selain membaca judul dan abstrak, dilakukan pembacaan seluruh teks artikel untuk memastikan kesesuaiannya. Pada tahap tersebut, hanya 15 artikel ilmiah yang memenuhi kriteria inklusi; sementara selebihnya tidak digunakan karena tidak memenuhi kriteria inklusi semisal tidak memprediksi prestasi belajar peserta didik, atau tidak menggunakan algoritma pembelajaran mesin. Sampel sebanyak 15 artikel inilah yang membahas implementasi algoritma pembelajaran mesin dalam memprediksi prestasi belajar peserta didik (6 artikel pada *database* IEEE Access dan 9 artikel pada *database* Science Direct), sementara selebihnya masuk dalam kriteria eksklusi. 15 artikel ilmiah inilah yang dijadikan acuan dalam penyusunan kajian literatur sistematis (lihat Gambar 4).



Gambar 4. Pencarian dan Seleksi Artikel Penelitian

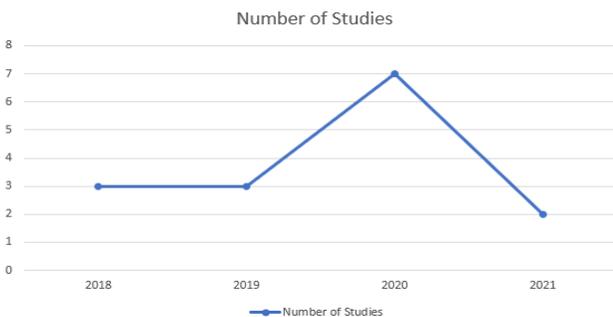
Setelah melakukan tinjauan pustaka sistematis terhadap 15 artikel penelitian, kemudian disusun jawaban atas pertanyaan-pertanyaan penelitian yang telah diajukan sebelumnya.

3.3 Artikel Penelitian yang Digunakan

Tinjauan pustaka sistematis atas 15 sampel artikel penelitian diperoleh dari 7 jurnal. Jurnal yang paling banyak menjadi sampel adalah jurnal IEEE Access dan Jurnal Computer & Education. Dari 7 jurnal yang digunakan, jurnal yang artikel ilmiahnya paling banyak digunakan adalah jurnal IEEE Access dan jurnal Computer & Education. Sebagian besar berada pada kategori Q1 dalam pemeringkatan *Scimago Journal Rank* [12] (lihat Tabel 3).

Tabel 3. Jumlah Artikel dan Scimago Journal Rank dari Jurnal yang Digunakan

No	Jurnal	Jumlah Artikel	SJR	Quartile
1.	IEEE Access	6	0.78	Q1 in Computer Science
2.	Computers in Human Behavior	2	2.17	Q1 in Arts & Humanities
3.	Computers & Education	3	3.05	Q1 in Computer Science
4.	Computers & Electrical Engineering	1	0.58	Q1 in Computer Science
5.	Knowledge-Based Systems	1	1.75	Q1 in Artificial Intelligence
6.	Socio-economic Planning Sciences	1	1.32	Q1 in Economics & Econometrics
7.	Materials Today: Proceedings	1	0.3	Not Yet Assigned



Gambar 5. Jumlah Penelitian Berdasarkan Tahun

Penggunaan algoritma pembelajaran mesin untuk memprediksi prestasi belajar peserta didik relatif berkembang. Hal ini diindikasikan dari peningkatan jumlah penelitian yang dilakukan dalam area kajian tersebut. Gambar 5 menunjukkan ada penurunan pada tahun 2021, namun hal tersebut masih berpotensi bertambah karena tahun 2021 belum berakhir. Peningkatan yang signifikan

terdapat pada 2020, yang melampaui jumlah penelitian pada 2018 dan 2019. Penelitian ini tidak mencari tahu apakah dampak pandemi dan perubahan mode belajar secara masif pada 2020 mendorong peningkatan jumlah penelitian pada 2020, sehingga masih menjadi ruang penelitian yang baik untuk dilanjutkan pada penelitian berikutnya.

3.4 Relevansi Pembelajaran Mesin

Implementasi algoritma pembelajaran mesin memiliki relevansi dalam upaya memprediksi prestasi belajar peserta didik. Penggunaan data dan algoritma pembelajaran mesin yang tepat diharapkan dapat membantu guru untuk memprediksi pencapaian siswa dan memberikan dukungan untuk pengambilan keputusan guna meningkatkan prestasi belajar peserta didik [1] [13]. Algoritma pembelajaran mesin juga relevan digunakan untuk mengekstrak kriteria prestasi belajar siswa di berbagai *domain* pendidikan untuk dapat membuat taksonomi yang koheren tentang hasil belajar siswa [14]. Selain itu, algoritma pembelajaran mesin juga dapat digunakan untuk memprediksi peserta didik yang berpeluang besar berprestasi secara akademis [15] dan juga mendeteksi yang memiliki resiko gagal [16], [17]; sehingga peserta didik yang teridentifikasi memiliki resiko gagal dapat diberikan tambahan dukungan lebih awal dan tepat waktu [1].

Proses pembelajaran yang semakin masif menggunakan jaringan internet di satu sisi juga menjadikan penggunaan algoritma pembelajaran mesin semakin relevan. Hubungan antara perilaku penggunaan internet dan kinerja akademis dapat digunakan untuk memprediksi kinerja akademis siswa [18]. Dalam konteks penggunaan *blended learning platform*, algoritma pembelajaran mesin relevan untuk digunakan dalam memodelkan prediksi kinerja akademik dari mahasiswa dengan menggunakan data *multiple-source* dan *multimodal data* [19] serta memprediksi hasil belajar siswa dengan menggunakan data perilaku siswa dan fitur latihan yang digunakannya dalam *platform online* [20].

Prestasi belajar peserta didik juga dapat dipengaruhi oleh berbagai faktor seperti lingkungan pembelajaran atau faktor lainnya baik di dalam dirinya atau di luar dirinya. Algoritma pembelajaran mesin didapati relevan dalam menggali perilaku belajar siswa, mengatasi masalah akademik siswa, mengoptimalkan lingkungan pendidikan, dan memungkinkan pengambilan keputusan berdasarkan data [21]. Lebih dari itu, algoritma pembelajaran mesin relevan dalam mengidentifikasi faktor-faktor kunci yang mempengaruhi keberhasilan akademik peserta didik di sekolah dan mengeksplorasi hubungan antar faktor-faktor kunci tersebut [22] [23]. Misalnya prediksi hasil belajar siswa dengan menggabungkan berbagai aspek kehidupan siswa yaitu kepribadian siswa; perilaku, gaya belajar serta gaya hidup seperti pola tidur, pola olahraga dan lain-lain [24]. Algoritma pembelajaran mesin juga relevan untuk mencari metode yang terbaik untuk *resampling* dan klasifikasi *dataset* hasil belajar siswa [25] dan memprediksi

hasil belajar siswa dari pencapaian bantuan *personalized* dari keluarga dan bimbingan sekolah [26].

3.5 Algoritma yang Diimplementasikan

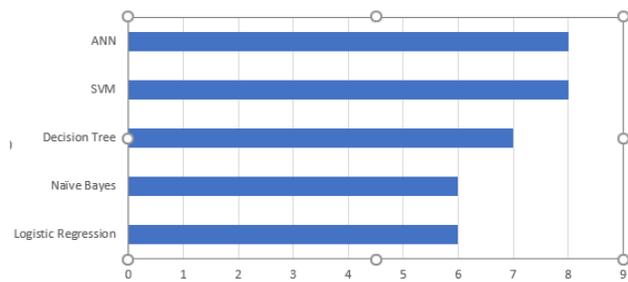
Tabel 4 menyajikan algoritma pembelajaran mesin yang diimplementasikan dalam prediksi prestasi belajar peserta didik.

Tabel 4. Algoritma Pembelajaran Mesin dalam prediksi Prestasi Belajar Peserta Didik

Algoritma	Referensi
JRIP	[19], [1]
Nnge	[19]
Naïve Bayes	[1], [17], [13], [15], [15], [24]
<i>K-Nearest Neighbor (KNN)</i>	[27], [17], [25]
<i>Regression tree</i>	[22]
<i>Random forest</i>	[22], [25]
<i>Decision tree</i> (termasuk di dalamnya J48, RepTree, RandomTree, CART Decision Tree)	[15], [16], [17], [18], [19], [1], [25]
<i>PART</i>	[19]
<i>Long Short-Term Memory (LSTM)</i>	[20], [23]
<i>Logistic regression</i>	[22], [17], [21], [13], [25], [15]
<i>Support Vector Machine (SVM)</i>	[15], [16], [17], [21], [18], [13], [25], [1]
<i>Multilayer perceptron (MLP) neural network</i>	[13], [15], [22]
<i>Layer Supervised MLP</i>	[13]
<i>Artificial Neural Network (ANN)</i>	[28][17], [21], [18], [16], [23], [15], [25], [1]
<i>Principal Component Analysis (PCA)</i>	[15]
<i>Feature Agglomeration (FA)</i>	[15]
<i>Data Envelopment Analysis (DEA)</i>	[21]
<i>Directional Distance Function (DDF)</i>	[21]
<i>k-Means</i>	[15]
<i>Singular Value Decomposition (SVD)</i>	[20]
<i>XG-boost</i>	[24]
<i>ICGAN-DVSM</i>	[25]

Algoritma	Referensi
<i>Recurrent neural network (RNN)</i>	[20]
<i>Bayesian knowledge tracing (BKT)</i>	[20]
<i>Deep knowledge tracing (DKT)</i>	[20]

Beragam algoritma pembelajaran mesin yang diimplementasi dalam memprediksi prestasi belajar peserta didik diuji akurasinya sesuai dengan tahapan analisis data yang disyaratkan dalam algoritma pembelajaran mesin. Artikel ini tidak memberikan kajian alat ukur akurasi yang diimplementasikan untuk menguji model matematika pembelajaran mesin, melainkan berfokus pada algoritma apa saja yang diimplementasikan. Daftar algoritma tersebut tersaji pada Gambar 6.



Gambar 6. Algoritma Pembelajaran Mesin yang paling banyak digunakan.

3.6 Hasil Implementasi Algoritma

Implementasi algoritma pembelajaran mesin pada 15 pustaka yang dikaji menunjukkan bahwa pemilihan algoritma untuk memprediksi prestasi belajar peserta didik disesuaikan dengan jenis data yang dimiliki serta tujuan yang hendak dicapai (baik dalam bentuk klasifikasi, regresi, *clustering* maupun kombinasi ketiganya). Tabel 5 menguraikan hasil penelitian beserta algoritma yang digunakan.

Tabel 5. Algoritma Pembelajaran Mesin Prediksi Prestasi Belajar Peserta Didik

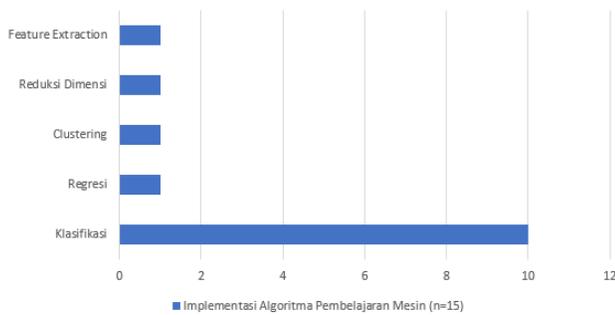
Artikel	Data dan Algoritma yang Digunakan serta Hasil Penelitian
[1]	Data aktifitas peserta didik di LMS (<i>choice view, course visit, forum, quiz activities, lesson view, viewing resources file</i>) digunakan untuk memprediksi prestasi belajar peserta didik (dengan klasifikasi <i>pass/fail</i>). Algoritma yang digunakan adalah <i>Naïve Bayes, J48 (Decision Tree), SMO (SVM)</i> dan JRip (<i>Rule-induction algorithm</i>). Tidak ada algoritma tunggal yang unggul di semua aspek untuk memprediksi prestasi belajar peserta didik. Tetapi, algoritma J48 (<i>Decision Tree</i>) dan JRip (<i>Rule-induction algorithm</i>) memberi prediksi

Artikel	Data dan Algoritma yang Digunakan serta Hasil Penelitian	Artikel	Data dan Algoritma yang Digunakan serta Hasil Penelitian
	yang lebih mudah dimengerti dalam bentuk <i>Tree</i> dan <i>Rule</i> .		dimensi adalah PCA dan FA. Penelitian ini juga menggunakan algoritma untuk <i>k-Means</i> untuk <i>clustering</i> . Ditemukan bahwa algoritma MLP lebih akurat dibanding algoritma lain untuk melakukan klasifikasi. Transformasi <i>dataset</i> dengan FA didapati lebih baik reduksi dimensi dibandingkan dengan PCA untuk data yang digunakan, dan algoritma <i>k-Means</i> dapat menghasilkan kluster peserta didik yang akan berhasil jika bekerja dalam kelompok belajar.
[13]	Data durasi penggunaan komputer oleh peserta didik, jumlah <i>logging</i> ke internet saat kelas berlangsung, berapa kali makan saat jam kelas selama periode satu semester untuk memprediksi prestasi belajar peserta didik dalam klasifikasi sebagai berikut. <ul style="list-style-type: none"> - Peserta didik dengan nilai >70 dan <70. - Peserta didik dengan nilai >80, <60, dan $60 \leq x \leq 80$. Algoritma yang digunakan: <i>Naïve Bayes</i> , <i>Logistic Regression</i> , SVM, MLP dan <i>Layer Supervised MLP</i> . MLP lebih presisi dibanding dengan <i>Naïve Bayes</i> , <i>Logistic Regression</i> dan SVM. Namun jika MLP dibandingkan dengan <i>Layer Supervised MLP</i> , <i>Layer Supervised MLP</i> lebih akurat.	[17]	Data prestasi belajar lampau peserta didik (nilai ujian sebelumnya), <i>engagement</i> (jumlah kunjungan peserta didik terhadap material pembelajaran, aktifitas pencarian, partisipasi dalam diskusi), demografis peserta didik (umur, keahlian, tanggal dan waktu, lokasi) dipergunakan untuk memprediksi nilai ujian di masa depan dalam bentuk klasifikasi (<i>pass/fail</i>) dan regresi (rentang nilai 0-100). Algoritma yang digunakan: <i>KNN</i> , <i>SVM</i> , <i>ANN</i> , <i>DT</i> , <i>Naïve Bayes</i> , dan <i>Logistic Regression</i> . Ditemukan bahwa algoritma ANN memiliki tingkat akurasi tertinggi ketika menggunakan <i>dataset engagement</i> dan prestasi belajar peserta didik di masa lampau. <i>Dataset demographic</i> tidak menunjukkan signifikansi apapun dalam proses prediksi.
[15]	Data peserta didik saat pendaftaran (<i>HSGA</i> , <i>SAT</i> dan <i>GAT</i>) digunakan untuk memprediksi prestasi belajar peserta didik dalam bentuk klasifikasi CGPA peserta didik sebagai berikut. <ul style="list-style-type: none"> - <i>Excellent</i> (≥ 4.5) - <i>Very good</i> (3.75 to < 4.5) - <i>Good</i> (2.75 to < 3.75) - <i>Average</i> (2.0 to < 2.75) - <i>Poor</i> (< 2.0) Algoritma yang digunakan adalah ANN, <i>Decision Tree</i> , SVM dan <i>Naïve Bayes</i> . Algoritma yang paling tinggi akurasi dan presisinya adalah ANN; sementara yang paling tinggi <i>recall</i> dan <i>F-1 measure</i> nya adalah algoritma DT.	[18]	Data aktifitas peserta didik di internet (durasi <i>online</i> , volume <i>traffic</i> , frekuensi koneksi) 4.000 peserta didik digunakan untuk memprediksi prestasi belajar peserta didik dalam bentuk klasifikasi <i>passed / failed</i> atau <i>high score / non-high score</i> . Algoritma yang digunakan adalah <i>Decision Tree</i> , ANN, dan SVM. Algoritma yang paling akurat dalam memprediksi adalah ANN dan SVM. Hasil menunjukkan bahwa volume <i>traffic</i> yang tinggi dalam menggunakan internet menunjukkan peserta didik yang <i>lower performance group</i> . Peserta didik dengan kategori <i>failed</i> memiliki volume <i>download</i> yang lebih tinggi sedangkan peserta didik yang <i>passed</i> dan <i>high score</i> lebih tinggi di frekuensi koneksi dan lama waktu di internet setiap harinya.
[16]	<i>Dataset LMS log files</i> seperti perbandingan jumlah <i>courses</i> dan peserta didik yang <i>ter-enroll</i> , durasi penyelesaian tiap <i>course</i> digunakan untuk membuat prediksi prestasi belajar peserta didik. Prediksi yang dilakukan ada dua, yaitu dalam bentuk klasifikasi (<i>at-risk/ balanced/ excellent</i>) dan <i>clustering</i> untuk memprediksi pengaruh kelompok belajar terhadap prestasi belajar peserta didik. Selain itu, penelitian ini juga melakukan reduksi dimensi terlebih dulu sebelum melakukan <i>clustering</i> , sebab data yang digunakan dimensinya terlalu besar, sementara algoritma <i>clustering</i> tidak akurat untuk <i>high dimensional spaces</i> . Algoritma yang digunakan untuk klasifikasi adalah <i>CART Decision Trees</i> , <i>Naïve Bayes</i> , <i>Logistic Regression</i> , MLP dan SVM sedangkan algoritma yang digunakan untuk reduksi	[19]	Data video rekaman dari kelas tatap muka (berapa kali peserta didik melihat komputer, berapa kali mencatat di buku, berapa kali memperhatikan guru), nilai dan jumlah partisipasi pada kelas praktek, kehadiran dan nilai pada kelas daring lewat <i>platform Moodle</i> , dan nilai ujian akhir untuk memprediksi prestasi belajar peserta didik dalam bentuk klasifikasi

Artikel	Data dan Algoritma yang Digunakan serta Hasil Penelitian	Artikel	Data dan Algoritma yang Digunakan serta Hasil Penelitian
	<p><i>fail/ pass/ dropout</i>. Algoritma yang digunakan adalah <i>J48, REPTREE, RandomTree, JRIP, Nnge, PART</i>. Algoritma <i>Nnge</i> dan <i>PART</i> memiliki nilai akurasi paling tinggi dalam memproses <i>numerical</i> dan <i>discretized data</i>. Ditemukan bahwa tingkat perhatian dalam kelas teori, skor kuis di Moodle, dan aktifitas di forum diskusi Moodle adalah atribut yang paling berpengaruh dalam prediksi kinerja akhir.</p>		<p>perempuan, <i>parental pressure</i>, dan kompetisi dipergunakan untuk mengidentifikasi faktor-faktor yang berdampak pada prestasi belajar peserta didik. Algoritma yang digunakan dalam <i>feature extraction</i> adalah <i>DDF, DEA, regression tree, random forest</i>. Algoritma <i>Random Forest</i> dan <i>Regression Tree</i> dapat menunjukkan dengan baik faktor-faktor yang mempengaruhi prestasi belajar peserta didik.</p>
[20]	<p>Data aktifitas pembelajaran (berapa kali peserta didik mengerjakan latihan soal, nilai yang diperoleh dari latihan soal, tipe latihan yang dikerjakan dan durasi pengerjaan latihan soal) digunakan untuk memprediksi prestasi belajar peserta didik (rentang 0-100), menggunakan algoritma dan <i>framework</i> tertentu. Algoritma yang digunakan adalah <i>DKT, RNN, BKT, dan LSTM</i>. Namun penelitian ini tidak berfokus pada pengujian algoritma, melainkan pada pengujian <i>framework</i>. <i>Framework</i> yang diuji adalah <i>MFA-DKT, BKT, DKT, DKT+, IRF-DKT dan DKVMN</i>. Dibandingkan <i>framework</i> lain, <i>MFA-DKT</i> lebih signifikan dalam meningkatkan akurasi prediksi prestasi belajar peserta didik serta mampu merekomendasikan program belajar yang sesuai. Hal ini dikarenakan <i>MFA-DKT</i> mampu mengkombinasikan data aktifitas pembelajaran dengan menggunakan <i>Neural Network</i> untuk membedakan prestasi belajar satu peserta didik dengan yang lainnya.</p> <p>Selain itu, data <i>clickstream behaviour</i> dan nilai asesmen digunakan untuk memprediksi prestasi belajar peserta didik yang diklasifikasikan menjadi <i>pass-fail/ withdrawn-pass/ distinction-fail/ distinction-pass</i>.</p> <p>Algoritma yang digunakan adalah <i>SVD, ANN, dan logistic regression</i>. Algoritma <i>ANN</i> lebih akurat dalam memprediksi dibanding dengan algoritma lainnya yang digunakan. Model yang dibentuk dapat memprediksi dengan tingkat akurasi 88%-95%. Pola belajar peserta didik seperti seberapa sering membuka modul pembelajaran dan melihat konten pembelajaran minggu sebelumnya mempunyai efek signifikan dalam pengklasifikasian.</p>	[22]	<p>Data hasil tes SABER PRO tahun 2016 dari 200,000 peserta yang berisi informasi sosioekonomi, karakteristik sekolah menengah atas dan status kerja; termasuk jenis kelamin, umur, dan program akademis yang dijalani. digunakan untuk memprediksi prestasi belajar peserta didik di perguruan tinggi dalam bentuk klasifikasi. Klasifikasi tingkat prestasi belajar peserta didik: <i>level 1 (0-125)/ level 2 (126-155)/ level 3 (156-200)/ level 4 (201-300)</i>. Algoritma yang digunakan adalah <i>ANN</i> dan <i>MLP</i>. Algoritma <i>ANN</i> memberikan hasil prediksi dengan tingkat akurasi terbaik.</p>
[21]	<p>Data rata-rata waktu peserta didik mengikuti pembelajaran dalam satu minggu, total waktu peserta didik mengikuti kelas tambahan, waktu pembelajaran dalam subjek tertentu semisal sains maupun bahasa dan matematika, ukuran kelas, ukuran sekolah, lokasi sekolah, proporsi</p>	[23]	<p>Data yang digunakan adalah <i>multi-source behavioral data</i> di dalam dan luar kelas. <i>Data personality (neuroticism, extraversion, dan agreeableness)</i>, profil individu (umur, jenis kelamin, tinggi badan, berat badan, kesehatan fisik, kesehatan kardiorespiratori, kesehatan aerobik, tingkat stress), data gaya hidup (pola makan, aktifitas fisik, manajemen waktu) dipergunakan untuk memprediksi prestasi belajar peserta didik dalam bentuk klasifikasi: <i>low/ medium/ high performance</i>. Algoritma yang digunakan adalah <i>LSTM</i>. Model <i>AugmentED</i> menggunakan algoritma <i>LSTM</i> dapat memprediksi prestasi belajar peserta didik.</p>
		[24]	<p>Data jenis kelamin, umur, kesehatan, pendidikan pekerjaan orang tua, akses internet, lokasi tempat tinggal, nilai sebelumnya digunakan untuk memprediksi prestasi belajar dalam 4 klasifikasi yaitu <i>poor, medium, good, excellent</i>. Algoritma yang digunakan untuk melakukan perbandingan hasil adalah <i>Random Forest, KNN, ANN, XG-boost, SVM, Decision Tree, Logistic Regression</i> dan <i>Naïve Bayes</i>. Algoritma <i>Random Forest</i> memberikan hasil prediksi yang lebih baik daripada algoritma lain yang digunakan.</p>
		[25]	<p>Data yang terdiri dari status tinggal (apakah peserta didik tinggal dengan orangtua atau tidak), pendidikan Ibu, pekerjaan Ibu, pendidikan Ayah,</p>

Artikel	Data dan Algoritma yang Digunakan serta Hasil Penelitian
	pendidikan Ayah, wali peserta didik, kualitas hubungan dalam keluarga, dukungan sekolah untuk pendidikan, dan dukungan keluarga untuk pendidikan digunakan untuk memprediksi prestasi belajar peserta didik dalam bentuk klasifikasi (<i>pass/ fail</i>). Algoritma yang digunakan adalah ICGAN-DSVM. Algoritma ICGAN-DSVM menunjukkan akurasi tinggi dalam memprediksi prestasi belajar peserta didik.
[26]	Data ketertarikan peserta didik, penghasilan orang tua, latar belakang orang tua, keterlibatan guru, keterlibatan teman, tingkah laku sosial dan <i>life style</i> , kesempatan kerja, <i>awareness</i> terhadap pendidikan dipergunakan untuk memprediksi prestasi belajar peserta didik apakah peserta didik akan berhasil atau gagal (dalam klasifikasi <i>fail/ pass</i>). Algoritma yang digunakan adalah KNN yang terbukti mampu memprediksi dengan tingkat akurasi yang tinggi.

Hasil kajian pustaka menunjukkan bahwa algoritma yang digunakan dalam memprediksi prestasi belajar peserta didik berbeda-beda sesuai dengan tipe implementasi pembelajaran mesin yang digunakan (klasifikasi, *clustering*, regresi, transformasi *dataset* dan *feature extraction*) (lihat Gambar 7).



Gambar 7. Implementasi Pembelajaran Mesin untuk Memprediksi Prestasi Belajar Peserta Didik

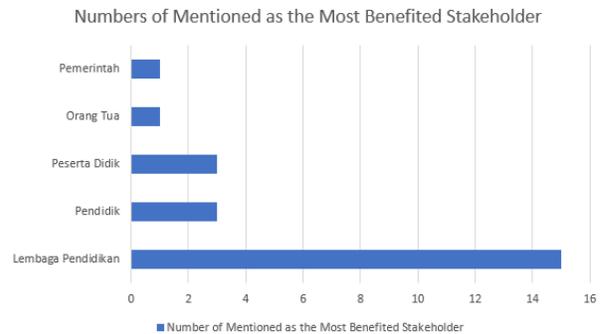
Kajian juga menemukan bahwa 15 artikel penelitian menggunakan beragam data yang ada pada lembaga pendidikan. Penggunaan data yang diambil dari LMS pun semakin banyak, seiring semakin pesatnya penggunaan LMS dalam pembelajaran daring. Penggunaan data video rekaman kegiatan pembelajaran dan aspek-aspek laten peserta didik hingga data-data yang dikumpulkan berdasarkan hasil tes psikologis dan tes masuk lembaga pendidikan pun digunakan untuk memberikan prediksi prestasi belajar peserta didik.

Secara dominan, prediksi prestasi belajar peserta didik diberikan dalam bentuk dua atau tiga klasifikasi (semisal

pass/ fail; atau *fail/ pass/ excellent*). Sementara *clustering* dilakukan untuk mengetahui komposisi kelompok belajar peserta didik yang bagaimana yang akan memberi prestasi belajar terbaik.

3.7 Pemangku Kepentingan (*Stakeholders*)

Berdasarkan 15 artikel penelitian yang ditinjau, ditemukan bahwa pemangku kepentingan yang mendapat manfaat dari penggunaan algoritma pembelajaran mesin adalah lembaga pendidikan, guru, peserta didik, keluarga, analis data, para peneliti dan pemerintah (lihat Gambar 8). Pemangku kepentingan terbesar yang mendapatkan manfaat adalah lembaga pendidikan. Hal ini menunjukkan urgensi lembaga pendidikan untuk menggunakan data melimpah peserta didik dengan lebih optimal untuk memprediksi prestasi peserta didik. Dengan prediksi yang dilakukan lebih awal, berbagai intervensi dapat disusun dan dilakukan oleh lembaga pendidikan untuk memastikan keberhasilan peserta didik. Hal ini juga mengindikasikan bahwa pihak pemangku kepentingan lembaga pendidikan akan mendapat keuntungan terbesar dengan melakukan implementasi algoritma pembelajaran mesin untuk memprediksi prestasi belajar peserta didik.



Gambar 8. Pemangku Kepentingan yang paling mendapat manfaat

4. SIMPULAN DAN USULAN

Tinjauan pustaka ini bertujuan untuk mengkaji secara sistematis sejauh mana implementasi algoritma pembelajaran mesin dan pemodelan telah dilakukan dalam memprediksi prestasi belajar peserta didik pada rentang 2018-2021. Berdasarkan kriteria eksklusi dan inklusi, 15 artikel penelitian digunakan untuk dikaji. Tinjauan pustaka ini dilakukan dengan tinjauan pustaka sistematis, yang merupakan sebuah pendekatan untuk mengidentifikasi, mengumpulkan dan memperoleh informasi berdasarkan pertanyaan penelitian yang telah disusun sebelumnya.

Empat pertanyaan dalam penelitian ini telah terjawab. Algoritma pembelajaran mesin relevan penggunaannya dalam memprediksi prestasi belajar peserta didik. Sampel artikel ilmiah yang terpilih untuk dianalisis dalam penelitian ini menunjukkan bahwa pembelajaran mesin untuk memprediksi prestasi belajar peserta didik diimplementasikan dengan cara terbanyak menggunakan

klasifikasi. Empat algoritma pembelajaran mesin yang paling lazim digunakan adalah *Artificial Neural Network* (ANN), *Support Vector Machine* (SVM), *Logistic Regression*, dan *Decision Tree*. Sedangkan pemangku kepentingan yang paling mendapat manfaat dari penelitian yang berupaya untuk memprediksi prestasi belajar peserta didik dengan mengimplementasikan algoritma pembelajaran mesin ini adalah lembaga pendidikan.

Berdasarkan tinjauan pustaka sistematis terhadap 15 artikel penelitian, ditemukan bahwa penelitian serupa relatif belum banyak dilakukan dan masih terus bertambah. Penggunaan LMS dan jaringan internet dalam proses pembelajaran membuka ruang yang sangat besar bagi penelitian yang berorientasi pada prediksi prestasi belajar peserta didik, sebab data yang tersedia pun semakin besar dan semakin representatif mewakili peserta didik. Di sisi lain, implementasi algoritma pembelajaran mesin dalam memprediksi prestasi belajar peserta didik memiliki tantangan terkait perbedaan indikator penilaian prestasi dan cara penilaiannya pada setiap kelas, mata pelajaran dan lembaga pendidikan.

Temuan di atas memberikan pemahaman bagi lembaga pendidikan untuk dapat mengimplementasikan pembelajaran mesin serta menggunakan algoritma pembelajaran mesin yang paling tepat berdasarkan kebutuhan dan *dataset* yang dimiliki untuk memastikan ketercapaian tujuan akhir pembelajaran yaitu prestasi atau keberhasilan belajar peserta didik. Prediksi prestasi belajar yang lebih dini dilakukan dapat membantu lembaga pendidikan menyusun strategi dan intervensi dini bagi ketercapaian tujuan akhir tersebut.

Penelitian lanjutan mengenai keragaman *dataset* yang digunakan untuk memprediksi prestasi belajar peserta didik serta pengembangan algoritma dan model pembelajaran mesin yang mampu memfasilitasi keragaman *dataset* untuk memprediksi prestasi belajar peserta didik akan sangat berguna untuk dikembangkan lebih lanjut. Urgensi penelitian serupa dan lanjutan yang diusulkan di atas tidak hanya berguna bagi lembaga pendidikan sebagai pemangku kepentingan yang paling mendapatkan manfaat dari penelitian serupa tetapi juga berguna bagi perumusan strategi pendidikan negara.

5. PENGHARGAAN

Artikel dan penelitian yang dilakukan untuk menghasilkan tinjauan pustaka sistematis ini merupakan produk akademis dari kemapanan ekosistem pembelajaran pada Program Studi Informatika Program Magister Fakultas Teknik Informatika Universitas Islam Indonesia serta bimbingan Prof. Fathul Wahid, S.T., M.Sc., Ph.D. dan Zainudin, S.T., M.I.T. dalam mata kuliah Metodologi Penelitian.

6. REFERENSI

- [1] Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D.J., dan Long, Q. 2017. Predicting academic performance by considering student heterogeneity. *Knowledge-Based Syst.*, 161 (Dec. 2017), 134–146. DOI= <http://doi.acm.org/10.1016/j.knosys.2018.07.042>.
- [2] Sukmadinata, N. S. 2005. *Landasan Psikologi Proses Pendidikan*. Remaja Rosda Karya, Bandung.
- [3] Jihad, A. 2009. *Evaluasi Pembelajaran*. Multi Pressindo, Yogyakarta.
- [4] Hamalik, O. 2007. *Proses Belajar Mengajar*. Bumi Aksara, Jakarta.
- [5] Dimiyati dan Mudjiyono. 2009. *Belajar dan Pembelajaran*. Rineka Cipta, Jakarta.
- [6] Azwar, S. 2006. *Pengantar Psikologi Integrasi*. Raja Grafindo Persada, Jakarta.
- [7] Suryabrata, S. 2006. *Psikologi Pendidikan*. Raja Grafindo Perkasa, Jakarta.
- [8] Vanderplas, J. 2017. *Python Data Science Handbook*. Sebastopol, CA 95472: O'Reilly Media, Inc., 2017.
- [9] Guido, S. 2017. *Introduction to Machine Learning with Python*. O'Reilly Media, Inc., Sebastopol, CA 95472.
- [10] Jurusan Informatika UII. 2021. *Penelitian Sains Data*. Diakses dari: <https://informatics.uii.ac.id/riset-pengabdian/sains-data/penelitian-sains-data/>.
- [11] Wahono, R. S. 2015. A systematic literature review of software defect prediction: research trends, datasets, methods, and frameworks. *Journal of Software Engineering*, 1, 1 (April 2015), 1-16.
- [12] Scimago. 2019. *Scimago Journal and Country Rank*. Diakses dari: <https://www.scimagojr.com/journalrank.php>.
- [13] Qu, S., Li, K., Zhang, S., dan Wang, Y. 2018. Predicting achievement of students in smart campus. *IEEE Access*. 6 (Oct. 2018), 60264–60273. DOI= <http://doi.acm.org/10.1109/ACCESS.2018.2875742>.
- [14] Zughoul, O., Momani, F., Almasri, O. H., Zaidan, A. A., Zaidan, B. B., Alsalem, M. A., Albahri, O. S., dan Hashim, M. 2018. Comprehensive insights into the criteria of student performance in various educational domains. *IEEE Access*. 6 (Nov. 2018), 73245–73264. DOI= <http://doi.acm.org/10.1109/ACCESS.2018.2881282>.
- [15] Mengash, H. A. 2020. Using data mining techniques to predict student performance to support decision making in university admission systems. *IEEE Access*. 8 (March 2020), 55462–55470. DOI= <http://doi.acm.org/10.1109/ACCESS.2020.2981905>.
- [16] Riestra-González, M., Paule-Ruíz, M. del. P., dan Ortin, F. 2020. Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers and Education*. 163 (Dec. 2020) DOI= <http://doi.acm.org/10.1016/j.compedu.2020.104108>.

- [17] Tomasevic, N., Gvozdenovic, N., dan Vranes, S. 2019. An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers and Education*. 143, 103676 (August 2019), 1-18. DOI= <http://doi.acm.org/10.1016/j.compedu.2019.103676>.
- [18] Xu, X., Wang, J., Peng, H., dan Wu, R. 2019. Prediction of academic performance associated with internet usage behaviors using machine learning algorithms. *Computers in Human Behavior*. 98 (Jan. 2019), 166–173. DOI= <http://doi.acm.org/10.1016/j.chb.2019.04.015>.
- [19] Chango, W., Cerezo, R., dan Romero, C. 2020. Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses. *Comput. Electr. Eng.* 89, 106908 (Oct. 2020), 1-13. DOI= <http://doi.acm.org/10.1016/j.compeleceng.2020.106908>.
- [20] Liu, D., Zhang, Y., Zhang, J. U. N., Li, Q., Zhang, C., dan Yin, Y. U. 2020. Multiple features fusion attention mechanism enhanced deep knowledge tracing for student performance prediction. *IEEE Access*. 8, 194896 (Oct. 2020), 1-10. DOI= <http://doi.acm.org/10.1109/ACCESS.2020.3033200>.
- [21] Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., dan Nawaz, R. 2019. Predicting academic performance of students from VLE big data using deep learning models. *Computer in Human Behavior*. 104, 106189 (Nov. 2019), 1-13. DOI= <http://doi.acm.org/10.1016/j.chb.2019.106189>.
- [22] Rebai, S., Yahia, F. B., dan Essid, H. 2018. A graphically based machine learning approach to predict secondary schools' performance in Tunisia. *Socio-Economic Planning Science*. 70, 100724 (Aug. 2018), 1-14. DOI= <http://doi.acm.org/10.1016/j.seps.2019.06.009>.
- [23] Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., dan Cascallar, E. 2021. Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*. 2, 100018 (March 2021). DOI= <http://doi.acm.org/10.1016/j.caeai.2021.100018>.
- [24] Zhao, L., Chen, K., Song, J., Xiaoliang, Z., Sun, J., Caulfield, B., Namee, B. M. 2021. Academic performance prediction based on multisource, multifeature behavioral data. *IEEE Access*. 9, 5454 (Jan. 2021), 1-13. DOI= <http://doi.acm.org/10.1109/ACCESS.2020.3002791>.
- [25] Ghorbani, R., dan Ghousi, R. 2020. Comparing different resampling methods in predicting students' performance using machine learning techniques. *IEEE Access*. 8 (April 2020), 67899–67911. DOI= <http://doi.acm.org/10.1109/ACCESS.2020.2986809>.
- [26] Zhao, M. 2020. Predicting students' performance with school and family tutoring using generative adversarial network-based deep support vector machine. *IEEE Access*. 8 (May 2020), 86745–86752. DOI= <http://doi.acm.org/10.1109/ACCESS.2020.2992869>.
- [27] Maheswari, K., Priya, A., Balamurugan, A., dan Ramkumar, S. 2021. Analyzing student performance factors using KNN algorithm. *Materials Today: Proceedings*. DOI= <http://doi.acm.org/10.1016/j.matpr.2020.12.1024>.
- [28] Wilson, J., dan Rodrigues, J. 2020. Classification accuracy and efficiency of writing screening using automated essay scoring. *Journal of School Psychology*. 82 (August, 2020), 123–140. DOI= <http://doi.acm.org/10.1016/j.jsp.2020.08.008>.