

Tinjauan Pustaka Sistematis: Implementasi Metode *Deep Learning* pada Prediksi Kinerja Murid

(Implementation of Deep Learning Methods in Predicting Student Performance: A Systematic Literature Review)

Muhammad Haris Diponegoro¹, Sri Suning Kusumawardani², Indriana Hidayah³

Abstract—The use of machine learning, which is one of the implementations in the field of artificial intelligence, has penetrated into various fields, including education. By using a combination of machine learning techniques, statistics, and databases, educational data mining can be carried out to find out the patterns that exist in a particular dataset. One use of educational data mining is to predict student performance. The results of student performance predictions can be used as an instrument for monitoring and evaluating the learning process so that it can help determine further steps in order to improve the learning process. This study aims to determine the state of the art implementation of deep learning which is part of machine learning in the context of educational data mining, especially regarding student performance predictions. In this study, a systematic literature review is presented to determine the variation of deep learning techniques or algorithms used and their performance. Twenty scientific publications were found and the average performance achieved in making predictions was 89.85%. The majority of the techniques used are Deep Neural Network (DNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM) with demographic, behavioral, and academic data features.

Intisari—Pemanfaatan *machine learning* yang merupakan salah satu implementasi dalam bidang *artificial intelligence* telah merambah ke berbagai bidang, salah satunya adalah bidang pendidikan. Dengan menggunakan kombinasi teknik *machine learning*, statistik, dan basis data, dapat dilakukan *educational data mining* untuk mengetahui pola yang ada dalam suatu *dataset* tertentu. Salah satu penggunaan *educational data mining* adalah untuk melakukan prediksi kinerja murid. Hasil dari prediksi kinerja murid dapat digunakan sebagai salah satu instrumen untuk melakukan *monitoring* dan evaluasi terhadap proses pembelajaran sehingga dapat membantu menentukan langkah-langkah lanjutan dalam rangka meningkatkan proses pembelajaran. Makalah ini bertujuan untuk mengetahui *state of the art* implementasi *deep learning* yang merupakan bagian dari *machine learning* pada konteks *educational data mining*, khususnya mengenai prediksi kinerja murid. Pada makalah ini disajikan *systematic literature review* untuk mengetahui variasi teknik atau algoritme *deep learning* yang digunakan beserta kinerja yang dicapai. Dari dua puluh publikasi ilmiah yang ditelusuri, rata-rata kinerja yang dicapai dalam melakukan prediksi adalah 89,85%. Mayoritas teknik yang digunakan adalah *Deep Neural Network* (DNN), *Recurrent Neural Network* (RNN),

dan *Long Short-Term Memory* (LSTM) dengan fitur data demografis, perilaku, dan akademis.

Kata Kunci—*Deep Learning*, *Educational Data Mining*, *Prediksi*, *Pendidikan*, *Kinerja*.

I. PENDAHULUAN

Salah satu pembelajaran yang dilakukan oleh manusia adalah melalui pengalaman. Dengan bertambahnya pengalaman pada seseorang, semakin banyak hal-hal yang dapat diketahui dan dipelajari. Sebuah mesin atau komputer yang dibekali dengan kemampuan *Artificial Intelligence* (AI) melalui perangkat keras (*hardware*) dan/atau perangkat lunak (*software*) juga dapat melakukan pembelajaran melalui pengalaman yang didapat dari pemrosesan data untuk dapat melakukan tugas-tugas tertentu. Kualitas dan kuantitas data yang diproses juga menentukan kualitas dan kuantitas pembelajaran yang diperoleh oleh mesin atau komputer tersebut.

Deep Learning (DL) yang merupakan sebuah teknik berbasis jaringan saraf tiruan telah banyak digunakan dalam beberapa tahun terakhir sebagai salah satu metode implementasi *Machine Learning* (ML) [1]. Pada beberapa artikel disebutkan bahwa DL tidak hanya spesifik untuk bidang tertentu, tetapi telah didefinisikan sebagai bentuk pembelajaran umum yang dapat menyelesaikan hampir berbagai macam masalah di berbagai bidang [2]. Publikasi penelitian implementasi DL dalam konteks *Educational Data Mining* (EDM) pertama kali muncul pada tahun 2015 dan tiap tahun terus meningkat secara signifikan [3].

Penggunaan pendekatan komputasi pada bidang pendidikan atau pembelajaran, khususnya terkait EDM, dilakukan untuk menganalisis data-data pendidikan [4]. EDM merupakan salah satu implementasi dari teknik *Data Mining* (DM) yang digunakan untuk melakukan analisis data dan menyelesaikan isu-isu dalam bidang pendidikan [5]. Penggunaan DM dalam bidang lain, seperti transportasi, kependudukan, hingga ke dunia ritel, telah dilakukan bertahun-tahun oleh pebisnis, peneliti, hingga pemerintah (*government*) [6]. Beberapa topik umum terkait implementasi EDM antara lain adalah deteksi kegagalan atau keberhasilan seorang murid pada suatu pelajaran tertentu, prediksi nilai akhir seorang murid, dan identifikasi seorang murid yang diperkirakan akan mengalami *dropout* [7]. Kemampuan untuk melakukan prediksi kinerja murid dan identifikasi kesulitan yang dihadapi merupakan kegiatan yang cukup signifikan dan bermanfaat, baik bagi murid maupun institusi pendidikan [8]. Berbekal kemampuan

^{1,2,3} Departemen Teknik Elektro dan Teknologi Informasi, Fakultas Teknik, Universitas Gadjah Mada, Jl. Grafika No.2 Kampus UGM Yogyakarta 55281 INDONESIA (e-mail: harisdiponegoro@mail.ugm.ac.id¹)

tersebut, dapat dilakukan langkah perbaikan, seperti melakukan intervensi melalui bimbingan atau dapat dijadikan sebagai indikator peringatan dini [9]. Jika dimanfaatkan secara efektif, hal tersebut juga dapat membantu institusi pendidikan untuk meningkatkan proses pembelajaran dan mengembangkan strategi yang tepat sehingga dapat menurunkan tingkat *dropout rates* [10]. Namun, pengukuran kinerja akademis seorang murid juga dihadapkan pada tantangan bahwa hal tersebut dipengaruhi oleh faktor-faktor yang berbeda-beda. Keterkaitan antara variabel dan faktor-faktor untuk melakukan prediksi kinerja murid bersifat nonlinier. Teknik DM dan ML yang tradisional mungkin tidak dapat diaplikasikan secara langsung pada jenis data dan permasalahan tersebut [11].

Referensi [12] menyajikan tinjauan sistematis mengenai penelitian terkait EDM dan tren yang berkembang. Dalam penelitian tersebut dinyatakan bahwa teknik ML bahkan DL telah digunakan secara luas pada konteks EDM dan prediksi kinerja murid merupakan salah satu topik penelitian utama [12]. Penelitian lain memaparkan tinjauan mengenai analisis dan prediksi kinerja murid melalui penggunaan ML [13]. Penelitian tersebut memaparkan beberapa teknik yang populer digunakan, yaitu ML, *Recommender System*, dan *Artificial Neural Network* [13].

Makalah ini menyajikan gambaran mengenai penggunaan DL secara spesifik dalam melakukan prediksi kinerja murid, seperti arsitektur/teknik/metode/algoritme yang digunakan beserta kinerja model DL yang dicapai, metrik yang digunakan untuk pengukuran kinerja arsitektur/teknik/metode/algoritme, tingkat/*level/grade* dari data murid yang digunakan, keluaran yang diprediksi, serta jenis data yang digunakan. Metodologi yang digunakan dalam makalah ini adalah tinjauan pustaka sistematis (*Systematic Literature Review/SLR*) yang kemudian hasilnya disajikan dalam lima bagian, yaitu pendahuluan, kajian pustaka, metodologi, hasil dan pembahasan, serta kesimpulan.

II. EDUCATIONAL DATA MINING DAN DEEP LEARNING

A. Educational Data Mining (EDM)

EDM merupakan proses analisis dari data-data yang diambil dari institusi pendidikan menggunakan teknik DM untuk menghasilkan informasi-informasi yang berguna dalam penelitian maupun praktik-praktik dalam bidang pendidikan [3], [4]. Proses ini menggunakan susunan langkah yang sama dengan implementasi DM pada bidang lainnya, seperti bisnis, genetika, dan kedokteran, yaitu *pre-processing*, *data mining*, dan *post-processing* [14].

EDM memiliki tujuan baik dari sudut pandang praktis maupun penelitian, seperti mengembangkan proses pembelajaran menjadi lebih baik hingga melakukan kajian lebih dalam mengenai fenomena-fenomena yang terjadi dalam dunia pendidikan. Tujuan-tujuan tersebut terkadang tidak mudah diukur, baik secara kualitatif maupun kuantitatif, karena masing-masing tujuan memiliki karakteristik teknik pengukuran [4].

Tujuan-tujuan tersebut dapat dikategorikan lagi menjadi beberapa hal yang lebih spesifik sebagai berikut [15].

- Mendukung komunikasi di antara para pemangku kepentingan.
- Mengelola dan mengembangkan pembelajaran.
- Memberikan rekomendasi.
- Melakukan prediksi atas kinerja murid.
- Melakukan pemodelan terhadap murid.
- Analisis terhadap struktur domain.

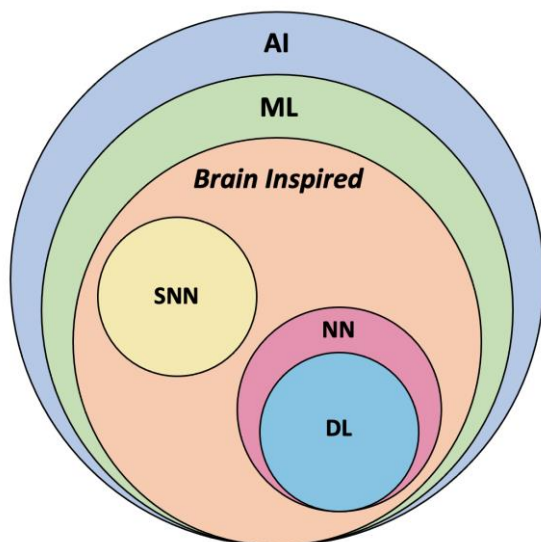
Dalam implementasi EDM, para peneliti menggunakan berbagai macam data yang bersumber antara lain dari *intelligence computer tutor*, *classic computer-based educational systems*, kelas *online*, data-data akademik yang ada di institusi pendidikan, serta data ujian yang terstandarkan [15].

EDM dapat dilakukan dengan menggunakan ide-ide yang terdapat pada *Organizational Data Mining* (ODM). Fokus dari ODM adalah membantu organisasi dalam berkompetisi dengan menjaga kemampuan yang unggul dalam organisasi serta mengacu pada teori-teori organisasi sebagai referensi. Organisasi yang dapat melakukan transformasi data yang dimiliki menjadi informasi yang berguna akan dapat memperoleh keuntungan, seperti kemampuan pengambilan keputusan yang lebih baik, peningkatan kemampuan berkompetisi, dan peningkatan finansial. EDM dapat menggunakan teori organisasi untuk menelaah fenomena-fenomena yang terjadi melalui berbagai tingkatan analisis, mulai dari lingkungan, organisasi, unit, hingga individu [16].

Tipe dari penelitian yang dilakukan dengan EDM umumnya memiliki fokus pada analisis kuantitatif. Hal ini diperlukan karena DM menggunakan metode statistik, ML, dan AI. Penelitian-penelitian yang ada biasanya berupa prediksi, klusterisasi atau klasifikasi dan asosiasi, serta bersifat spesifik terhadap institusi tertentu dan waktu tertentu [17].

Dalam EDM, pemodelan prediksi biasanya digunakan untuk melakukan prediksi terhadap kinerja murid. Untuk melakukan pemodelan prediksi, beberapa pendekatan yang sering digunakan adalah klasifikasi, regresi, dan kategorisasi. Pendekatan klasifikasi sering dipilih untuk melakukan prediksi terhadap kinerja murid dengan menggunakan algoritme atau arsitektur seperti *Decision Tree*, *Neural Network* (NN), *Naive Bayes*, *K-Nearest Neighbor*, dan *Support Vector Machine* (SVM) [18]. Referensi [18] menunjukkan bahwa *Neural Network* memiliki tingkat akurasi tertinggi dibandingkan algoritme atau arsitektur lainnya.

Referensi [3] menyebutkan bahwa sepanjang tahun 2015 hingga tahun 2018 terdapat peningkatan yang cukup signifikan pada jumlah publikasi terkait penerapan DL dalam konteks EDM. Arsitektur yang digunakan juga cukup bervariasi, antara lain *Multi Layer Perceptron* (MLP), *Long Short-Term Memory* (LSTM), *Word Embedding* (WE), *Convolutional Neural Network* (CNN) dan variannya (VGG16 dan AlexNet), *Feedforward Neural Network* (FNN), *Recurrent Neural Network* (RNN), *Autoencoder*, *Bidirectional LSTM* (BLSTM), dan *Memory Networks* (MN) [3]. Penelitian ini juga melakukan penggolongan terhadap tugas-tugas yang akan diselesaikan menggunakan EDM, yaitu prediksi kinerja murid, deteksi perilaku murid yang tidak dikehendaki, memberikan rekomendasi, serta evaluasi pembelajaran [3].



Ket.: AI : Artificial Intelligence DL : Deep Learning
 ML: Machine Learning SNN: Spiking Neural Network
 NN : Neural Network

Gbr. 1 Taksonomi AI.

B. Deep Learning (DL)

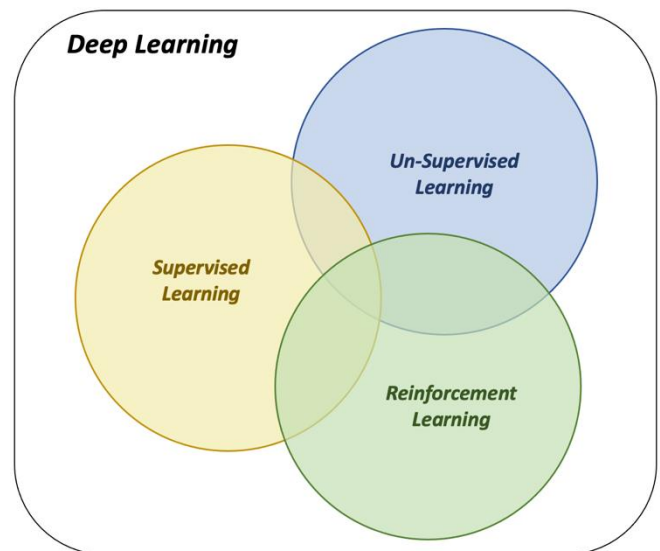
Semenjak tahun 1950-an, salah satu cabang dari AI yang disebut dengan ML telah berkembang cukup pesat dengan implementasi di beberapa bidang. NN adalah salah satu implementasi dari ML, sedangkan DL merupakan salah satu implementasi dari NN. Taksonomi ini digambarkan seperti pada Gbr. 1 [19].

DL, yang mulai populer digunakan sejak tahun 2006, menggunakan mekanisme *deep architecture of learning* atau pendekatan *hierarchical learning*. *Learning* atau pembelajaran dalam hal ini adalah sebuah prosedur yang berisi proses estimasi parameter-parameter suatu model sehingga model yang dikembangkan (algoritme) dapat menyelesaikan suatu tugas atau permasalahan tertentu [19].

DL menggunakan beberapa lapisan (*layers*) di antara lapisan masukan (*input layer*) dan lapisan keluaran (*output layer*). Arsitektur tersebut dapat digunakan untuk melakukan pemrosesan nonlinier dengan beberapa tahap yang hasilnya dapat digunakan untuk *feature learning* dan klasifikasi pola (*pattern classification*) [20], [21]. Jumlah lapisan dalam DL yang bervariasi dapat digunakan untuk melakukan abstraksi dengan tingkat yang berbeda-beda [22].

DL juga dapat dideskripsikan sebagai sebuah kelas dalam algoritme ML yang menggunakan beberapa lapisan pemrosesan nonlinier yang disusun secara *cascade* untuk ekstraksi fitur (*feature extraction*) dan *transformation*. Setiap lapisan menggunakan keluaran dari lapisan sebelumnya sebagai masukannya. Algoritme yang digunakan dapat bertipe *supervised* dan *unsupervised* serta implementasinya dapat digunakan sebagai *pattern analysis (unsupervised)* dan klasifikasi (*supervised*) [22].

Beberapa teknik dalam DL dapat dikategorikan menjadi *supervised*, *semi-supervised*, dan *unsupervised*. Kategori lain, seperti *Reinforcement Learning (RL)* atau *Deep RL (DRL)*,



Gbr. 2 Kategori dalam Deep Learning (DL).

seringkali dikategorikan menjadi *semi-supervised* atau *unsupervised*. Pengategorian tersebut diperlihatkan pada Gbr. 2 [19].

1) *Deep Supervised Learning*: Teknik *learning* yang digunakan dalam kategori ini menggunakan data yang telah diberi label sebelumnya (*labeled data*). Contoh yang populer dalam kategori ini adalah *Deep Neural Networks (DNN)*, *CNN*, (*RNN*), termasuk juga *LSTM*, dan *Gated Recurrent Unit (GRU)* [19].

2) *Deep Semi-Supervised Learning*: *Semi-supervised learning* menggunakan teknik *learning* yang menggunakan sebagian data yang telah diberi label sebelumnya (*partially labeled data*). Pada beberapa kasus, *DRL*, *Generative Adversarial Networks (GAN)*, serta *RNN*, termasuk *LSTM*, dan *GRU* juga menggunakan teknik *learning* ini [19].

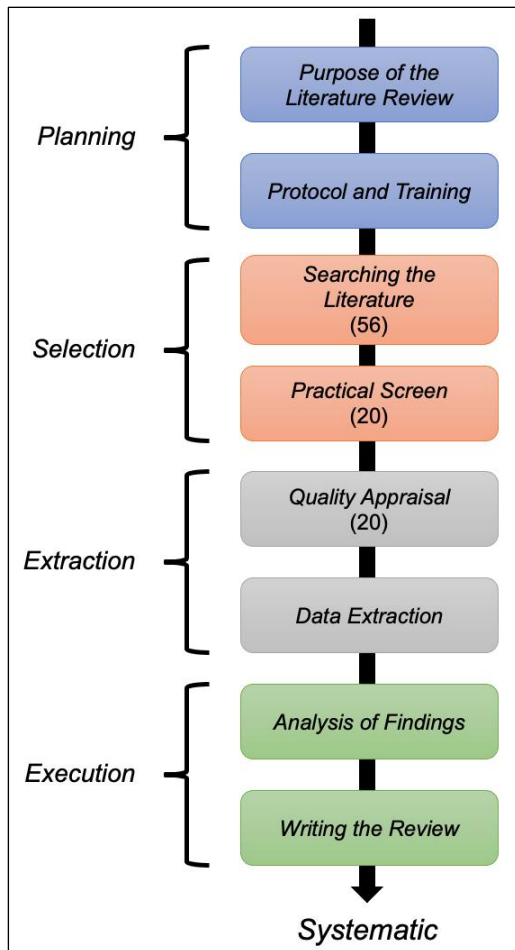
3) *Deep Unsupervised Learning*: Teknik *learning* ini menggunakan data yang tidak diberi label sebelumnya (*unlabeled data*). *Auto Encoders (AE)*, *Restricted Boltzmann Machines (RBM)*, dan generasi terbaru dari *GAN* menggunakan teknik *learning* ini pada implementasinya [19].

4) *Deep Reinforcement Learning*: Teknik *learning* ini digunakan pada lingkungan atau *environments* yang tidak diketahui (*unknown environments*). Pada tahun 2013, *DRL* dimulai dengan hadirnya *Google DeepMind* [23], [24].

III. METODOLOGI

A. Tinjauan Pustaka Sistematis (Systematic Literature Review)

Penelitian ini dilakukan dengan menggunakan metode tinjauan pustaka sistematis atau *SLR*. *SLR* adalah suatu proses untuk mengidentifikasi, menilai, dan menafsirkan fakta dan bukti dari penelitian yang tersedia dengan tujuan untuk mencari jawaban dari sebuah pertanyaan penelitian (*research question*) tertentu [25]. Gbr. 3 menunjukkan panduan sistematis untuk melakukan pengembangan tinjauan literatur [26] beserta



Gbr. 3 Panduan sistematis untuk melakukan pengembangan tinjauan literatur beserta jumlah literatur yang ditelusuri.

jumlah literatur yang ditelusuri. Beberapa tahapan yang dilakukan dalam SLR dijelaskan sebagai berikut [26].

1) *Purpose of the Literature Review*: Tahapan pertama dalam setiap kajian literatur adalah menentukan tujuan dan hasil yang diharapkan dari kajian literatur yang akan dilakukan. Definisi tujuan tersebut perlu disajikan secara eksplisit kepada pembaca.

2) *Protocol and Training*: Jika kajian literatur dikerjakan oleh lebih dari satu orang, perlu ditetapkan standar dan prosedur dalam melakukan kajian sehingga hasil kajian yang disajikan konsisten.

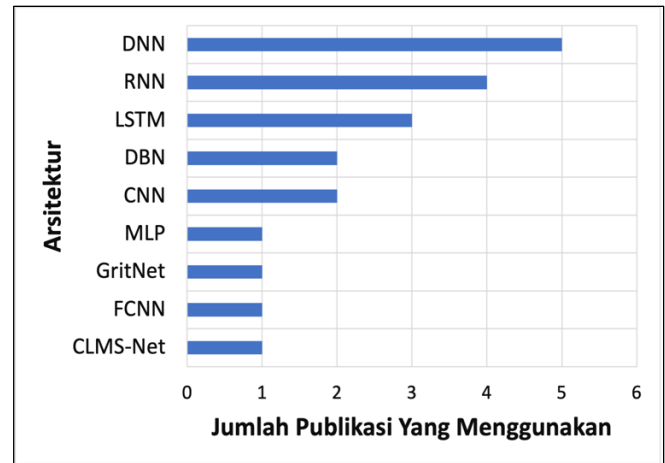
3) *Searching for the Literature*: Metode dan proses pencarian literatur yang akan dilakukan perlu dijelaskan secara eksplisit.

4) *Practical Screen*: Tahapan ini disebut juga *screening for inclusion*. Dalam tahapan ini dijelaskan faktor-faktor yang membuat sebuah literatur masuk dalam kategori literatur yang akan dikaji. Selain itu, faktor-faktor yang membuat sebuah literatur dikesampingkan juga perlu dijelaskan.

5) *Quality Appraisal*: Tahapan ini disebut juga *screening for exclusion*. Dalam tahapan ini dijelaskan faktor-faktor yang

TABEL I
KRITERIA PICOC YANG DIGUNAKAN

Population	Sistem pendidikan, murid, <i>data mining</i>
Intervention	Prediksi kinerja murid, <i>deep learning</i>
Comparison	<i>n/a</i>
Outcomes	Kinerja pemodelan, algoritme, metode, metrik pengukuran
Context	Penelitian akademis maupun industri



Gbr. 4 Grafik arsitektur yang digunakan.

menyebabkan sebuah literatur tidak masuk dalam kategori literatur yang akan dikaji.

6) *Data Extraction*: Setelah literatur-literatur yang akan dikaji teridentifikasi, perlu dilakukan ekstraksi secara sistematis mengenai informasi yang berkaitan dengan tujuan dilakukannya kajian terhadap setiap literatur tersebut.

7) *Synthesis of Studies*: Tahapan ini disebut juga tahapan analisis. Langkah ini mengombinasikan hasil dari ekstraksi informasi terhadap keseluruhan literatur menggunakan metode kualitatif, kuantitatif, atau keduanya.

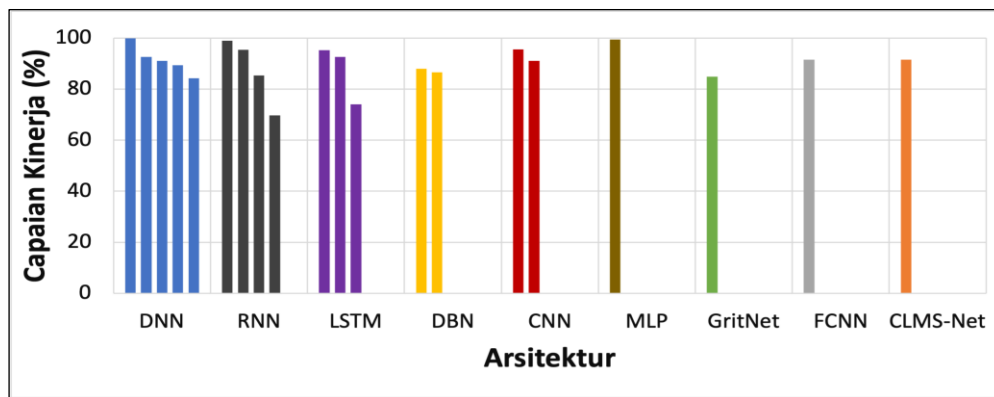
8) *Writing the Review*: Untuk memenuhi standar penulisan artikel penelitian, proses dan hasil SLR tersebut perlu dicantumkan.

Proses penelusuran literatur dilakukan pada situs pengindeks publikasi ilmiah Google Scholar dengan menggunakan strategi pencarian sebagai berikut.

- Tahun publikasi: 2016 – 2020.
- Jenis publikasi: jurnal atau prosiding.
- Kata kunci: kombinasi “*students performance prediction*,” “*deep learning*,” dan “*deep neural network*”.

Kriteria inklusi literatur yang digunakan dalam penelitian ini adalah:

- ditulis dalam bahasa Indonesia atau bahasa Inggris,
- mencantumkan arsitektur/teknik/metode/algoritme DL yang digunakan,
- mencantumkan kinerja model DL yang dicapai, dan
- mencantumkan metrik yang digunakan dalam mengukur kinerja prediksi.



Gbr. 5 Capaian kinerja masing-masing arsitektur/teknik/metode/algortime.

Sementara itu, kriteria eksklusi literatur yang digunakan adalah:

- literatur diterbitkan sebelum tahun 2016,
- tidak spesifik membahas implementasi DL untuk melakukan prediksi kinerja murid.

B. Pertanyaan Penelitian (Research Question)

Pertanyaan penelitian yang digunakan dalam penelitian ini disusun menggunakan bantuan kriteria PICOC (*Population, Intervention, Comparison, Outcomes, Context*) [25] dan dijabarkan pada Tabel I.

Untuk mengetahui gambaran mengenai penggunaan DL dalam melakukan prediksi kinerja murid, disusun tiga pertanyaan penelitian sebagai berikut.

1. Pertanyaan Penelitian 1 (P1): Jenis arsitektur/teknik/metode/algortime apa yang digunakan untuk melakukan prediksi dan kinerja yang dicapai serta *venue* dari publikasi tersebut?
2. Pertanyaan Penelitian 2 (P2): Metrik apa saja yang digunakan untuk mengukur kinerja prediksi yang dilakukan?
3. Pertanyaan Penelitian 3 (P3): Keluaran apa yang diprediksi sebagai acuan untuk mengukur kinerja murid serta bagaimana jenis data dan tingkat/grade/level pendidikan murid yang digunakan untuk melakukan prediksi?

IV. HASIL DAN PEMBAHASAN

Dari hasil penelusuran publikasi ilmiah pada situs pengindeks yang disebutkan pada bagian III, dipilih dua puluh publikasi ilmiah yang dapat digunakan sebagai referensi untuk menjawab pertanyaan penelitian yang telah disusun. Terdapat beberapa publikasi ilmiah yang tidak dapat digunakan sebagai referensi karena tidak membahas secara spesifik mengenai implementasi DL untuk melakukan prediksi kinerja murid.

A. P1: Jenis Arsitektur/Teknik/Metode/Algortime Apa yang Digunakan untuk Melakukan Prediksi dan Kinerja yang Dicapai serta Venue dari Publikasi Tersebut?

Dari hasil penelusuran yang dilakukan terhadap publikasi ilmiah dengan topik implementasi DL pada prediksi kinerja murid, diketahui bahwa pendekatan yang dilakukan adalah dengan menggunakan *deep supervised learning* dengan

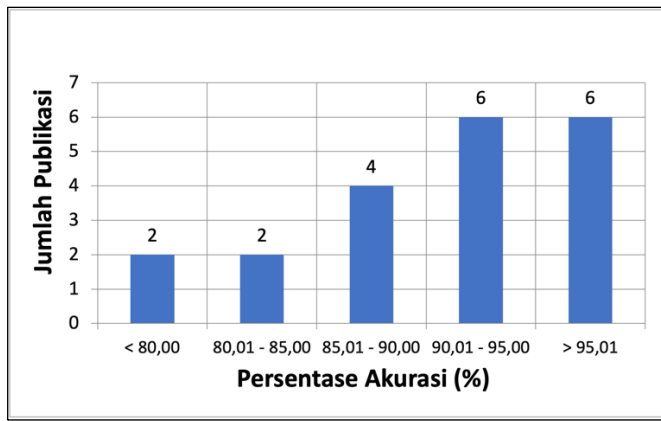
arsitektur yang banyak digunakan adalah DNN, RNN, dan LSTM. Gbr. 4 menunjukkan frekuensi penggunaan arsitektur/teknik/metode/algortime, sedangkan pada Gbr. 5 ditunjukkan grafik capaian kinerja masing-masing arsitektur/teknik/metode/algortime.

Referensi [27] menggunakan kombinasi antara CNN, LSTM, dan SVM, yang diberi nama CLMS-Net. CNN digunakan untuk ekstraksi fitur, LSTM berfungsi untuk menangani karakteristik yang berkaitan dengan *timing*, serta SVM digunakan untuk meningkatkan kinerja dalam melakukan prediksi [27]. Pada penelitian lain digunakan metode *hybrid*, yaitu *Fuzzy Clustering Neural Network* (FCNN), yang di dalamnya terdiri atas *unsupervised fuzzy c-means* dan *backpropagation neural network* [28].

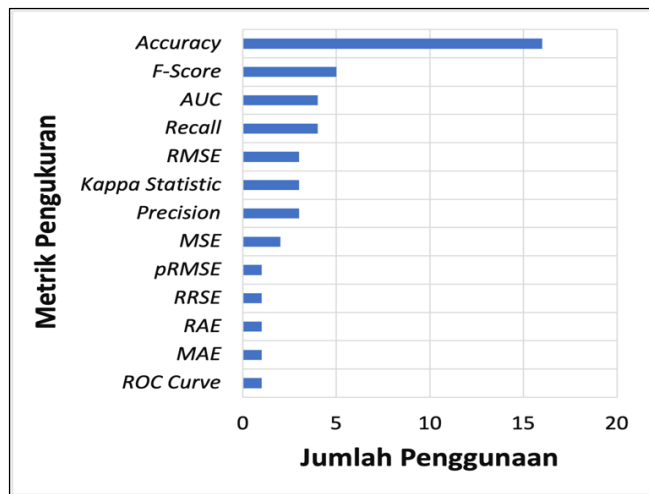
Referensi [29] memperkenalkan arsitektur baru yang diberi nama GritNet. Arsitektur tersebut diterapkan pada implementasi prediksi kinerja murid dalam *Massive Open Online Course* (MOOCs). Pada penelitian tersebut juga dikembangkan metode *unsupervised domain adaptation* untuk mentransfer model yang dihasilkan dari pembelajaran yang telah dilakukan GritNet sebelumnya ke data baru yang belum dibubuhi label. Strategi ini diterapkan untuk melakukan prediksi secara *real-time* terhadap data baru, terutama pada data minggu-minggu awal suatu *course* [29].

Hasil penelusuran juga mendapatkan beberapa teknik optimasi yang digunakan pada beberapa arsitektur tersebut guna mendapatkan hasil yang lebih baik. *ADAM optimizer* digunakan beberapa penelitian [30]-[32]. Pada penelitian lain digunakan *stochastic gradient descent* sebagai teknik optimasi [33] dan teknik *bidirectional* pada arsitektur LSTM untuk meningkatkan akurasi [34].

Referensi [35] mencapai tingkat akurasi yang paling tinggi, yaitu sebesar 99,80%. Penelitian tersebut menggunakan arsitektur DNN dengan lima belas variasi yang berbeda, dengan cara melakukan kombinasi antara tiga lapisan tersembunyi (*hidden layers*), yaitu sebanyak tiga, lima, dan tujuh, dengan lima jumlah *neuron*, yaitu 64, 128, 256, 512, dan 1.024. Penelitian tersebut juga menggunakan data yang terbanyak dibandingkan penelitian lain yang disajikan dalam makalah ini, yaitu sebanyak 641.138 data. Pada Gbr. 6 disajikan histogram distribusi jumlah publikasi dengan persentase akurasi yang dicapai. Dari histogram tersebut dapat diketahui bahwa



Gbr. 6 Histogram distribusi jumlah publikasi dengan persentase akurasi yang dicapai.



Ket.: AUC : Area Under the Curve
 RMSE : Root Mean Squared Error
 MSE : Mean Squared Error
 pRMSE : Predictive Mean Squared Error
 RRSE : Root Relative Squared Error
 RAE : Relative Absolute Error
 MAE : Mean Absolute Error
 ROC Curve: Receiver Operating Characteristic Curve

Gbr. 7 Grafik metrik pengukuran yang digunakan.

mayoritas penelitian-penelitian hasil penelusuran memiliki persentase kinerja di atas 85%.

Hasil penelusuran juga mendapatkan hasil bahwa sebanyak 65% publikasi yang ditelusuri merupakan publikasi pada prosiding (tiga belas publikasi) dan 35% publikasi merupakan publikasi pada jurnal (tujuh publikasi).

B. P2: Metrik Apa Saja yang Digunakan untuk Mengukur Kinerja Prediksi yang Dilakukan?

Gbr. 7 menunjukkan bahwa metrik *accuracy* (*classification accuracy*), *f-score*, *Area Under the Curve* (AUC), dan *recall* merupakan metrik yang paling sering digunakan. Selain menggunakan metrik *accuracy*, juga dilakukan penilaian terhadap kualitas model prediksi yang dihasilkan dengan mengukur kedekatan hasil prediksi dengan nilai yang sesungguhnya [36]. Penelitian tersebut menggunakan *confusion matrix* berbentuk grafis untuk melakukan kalkulasi

TABEL II

KELUARAN YANG DIPREDIKSI DAN JENIS DATA YANG DIGUNAKAN

Penelitian	Keluaran yang Diprediksi	Jenis Data
[27]	<i>Dropout or Not Dropout</i>	Perilaku
[28]	<i>Performance Level</i>	Demografis, Perilaku
[29]	<i>Graduate or Not Graduate</i>	Akademis, Perilaku
[30]	<i>Pass or Fail</i>	Akademis
[31]	<i>Pass or Fail</i>	Perilaku
[32]	<i>Total Grade/Mark</i>	Demografis, Akademis Perilaku
[33]	<i>Pass or Fail</i>	Perilaku
[34]	<i>Grade Point Average (GPA)</i>	Akademis
[35]	<i>Dropout or Not Dropout</i>	Perilaku
[36]	<i>Grade</i>	Demografis, Akademis, Perilaku
[37]	<i>Final Grade</i>	Akademis, Perilaku
[38]	<i>Performance Level</i>	Demografis, Akademis, Perilaku
[39]	<i>Performance Level</i>	Demografis, Akademis, Perilaku
[40]	<i>Final Grade</i>	Akademis, Perilaku
[41]	<i>Total Grade/Mark</i>	Demografis, Akademis, Perilaku
[42]	<i>Score</i>	Akademis, Perilaku
[43]	<i>Grade</i>	Akademis
[44]	<i>Final Grade</i>	Demografis, Akademis, Perilaku
[45]	<i>Post Test Scores</i>	Akademis
[46]	<i>Final Grade</i>	Demografis, Akademis, Perilaku

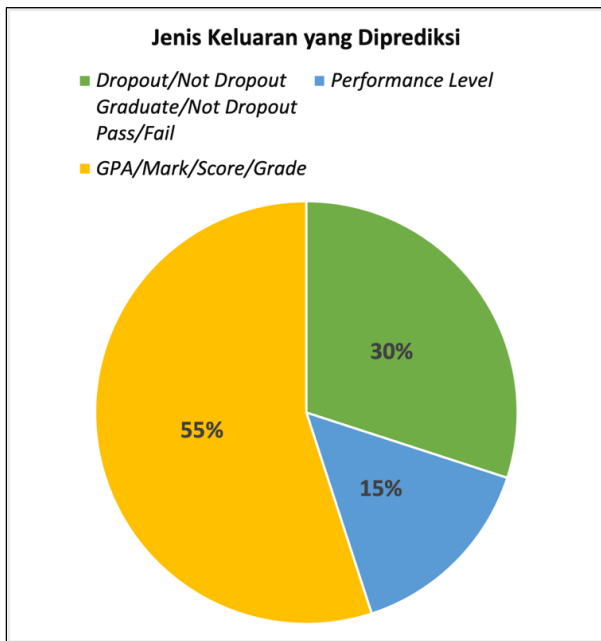
Predictive Mean Squared Error (pRMSE) [36]. Berbeda dengan metrik *accuracy*, semakin kecil nilai pRMSE, semakin baik model prediksi yang dihasilkan [36].

Referensi [27] melakukan prediksi *dropout* sehingga klasifikasi yang dilakukan merupakan klasifikasi biner (*binary*). Penelitian tersebut menyebutkan bahwa metrik *accuracy* tidak cocok digunakan karena adanya permasalahan *imbalanced* pada data target.

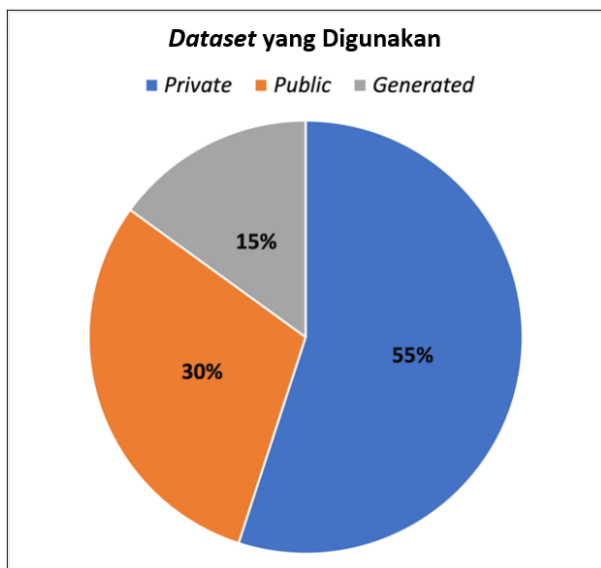
Sementara itu, telah dilakukan juga prediksi kinerja murid dengan target yang diprediksi adalah nilai akhir (*final grade/grade point average*) [34], [37], [38]. Penelitian tersebut menggunakan *Root Mean Squared Error* (RMSE) sebagai salah satu metrik pengukuran kinerja prediksi yang dilakukan. RMSE digunakan untuk menunjukkan perbedaan antara nilai aktual dengan nilai prediksi yang dilakukan.

C. P3: Keluaran Apa yang Diprediksi sebagai Acuan untuk Mengukur Kinerja Murid serta Bagaimana Jenis Data dan Tingkat/Grade/Level Pendidikan Murid yang Digunakan untuk Melakukan Prediksi?

Tabel II menunjukkan keluaran yang diprediksi dan jenis data yang digunakan oleh masing-masing penelitian. Gbr. 8 menunjukkan bahwa mayoritas penelitian melakukan prediksi nilai murid (*grade point average/GPA, mark, score, grade*)



Gbr. 8 Grafik jenis keluaran yang diprediksi.



Gbr. 9 Grafik dataset yang digunakan.

sebagai salah satu acuan untuk mengukur kinerja murid. Jenis data yang digunakan dalam penelitian-penelitian tersebut dapat dikategorikan menjadi tiga jenis, yaitu data demografis, akademis, dan perilaku.

Data yang termasuk jenis data demografis di antaranya adalah jenis kelamin, usia, profil keluarga dan orang tua murid, hingga fasilitas yang dimiliki, seperti komputer/laptop dan koneksi internet [28], [38], [41], [44], [46]. Kategori jenis data perilaku di antaranya adalah aktivitas dalam pembelajaran, seperti kehadiran, *raised hand*, diskusi, membaca materi dan pengumuman, jumlah jam belajar mandiri yang dilakukan di luar jam kelas, hingga frekuensi murid pergi ke luar rumah bersama teman [27], [35], [38], [40], [44], [46]. Selanjutnya, yang termasuk data akademis di antaranya adalah strata

pendidikan, kelas/tingkat/semester, kuis, tugas, dan GPA [29], [30], [34], [40], [41]. Hasil penelusuran juga menunjukkan bahwa 45% publikasi melakukan prediksi kinerja murid di tingkat universitas, 30% di tingkat sekolah menengah (*high/middle school*), dan 25% publikasi tidak menyebutkan secara eksplisit tingkat/*grade/level* murid yang diprediksi.

Dataset yang digunakan dalam penelitian-penelitian tersebut pada umumnya dapat diklasifikasikan menjadi tiga, yaitu *dataset generated*, *private*, dan *public* (Gbr. 9). Referensi [34], [36], dan [42] menggunakan *dataset generated*, dengan jumlah data 2.000 [34], [42], serta 10.000 [36]. *Dataset public* seperti UCI Machine Learning, OULAD, dan Kaggle digunakan dalam beberapa penelitian [27], [28], [31], [32], [46].

V. KESIMPULAN

Berdasarkan penelusuran terhadap penelitian-penelitian yang telah dilakukan, rata-rata kinerja arsitektur/teknik/metode/algortme DL dalam melakukan prediksi kinerja murid mencapai 89,85%. Dengan rata-rata tingkat akurasi tersebut, penggunaan DL dengan arsitektur seperti DNN, CNN, dan LSTM untuk melakukan prediksi kinerja murid dapat digunakan sebagai salah satu instrumen deteksi dini terhadap murid yang memiliki risiko mengalami kesulitan dan membantu institusi pendidikan, pendidik/pengajar, dan murid untuk meningkatkan proses pembelajaran. Metrik pengukuran kinerja terhadap prediksi yang dilakukan mayoritas menggunakan kombinasi metrik *accuracy* (*classification accuracy*), *f-score*, AUC, dan *recall*. Selain itu, penggunaan data selain data akademis, seperti data perilaku dan demografis, juga banyak digunakan sebagai fitur dalam melakukan prediksi. Dengan demikian, peningkatan proses pembelajaran yang didasarkan pada prediksi kinerja murid juga dapat dilakukan dengan memberikan perhatian pada aspek-aspek nonakademis sehingga evaluasi yang dihasilkan menjadi lebih komprehensif.

Penelitian selanjutnya dapat difokuskan pada arsitektur DNN, CNN, dan LSTM agar dapat diketahui tingkat kecocokan penggunaannya terhadap konteks, jenis, dan fitur/atribut dari data yang akan dianalisis. Keterkaitan antar fitur/atribut dan jenis fitur/atribut yang memengaruhi juga merupakan hal penting untuk diungkap. Dengan adanya informasi tersebut, diharapkan intervensi yang dilakukan terhadap murid yang memiliki risiko kinerja yang kurang dapat difokuskan terhadap fitur/atribut penyebabnya, sehingga proses intervensi lebih efektif dan efisien. Berbagai macam teknik untuk meningkatkan kinerja model prediksi, seperti *feature engineering/selection*, optimasi algoritme, maupun *hyperparameter* dan metode *ensemble*, juga dapat dijadikan fokus bahan kajian untuk penelitian selanjutnya.

REFERENSI

- [1] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, dan G.Z. Yang, "Deep Learning for Health Informatics," *IEEE Journal of Biomedical and Health Informatics*, Vol. 21, No. 1, hal. 4-21, Jan. 2017.
- [2] Y. Bengio, "Learning Deep Architectures for AI," *Found. Trends Mach. Learn*, Vol. 2, No. 1, hal 1-127, 2009.
- [3] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, dan B. Navarro-Colorado, "A Systematic Review of Deep Learning Approaches to Educational Data Mining," *Complexity*, Vol. 1, hal. 1-22, 2019.

- [4] C. Romero dan S. Ventura, "Educational Data Mining: A Review of the State of the Art," *IEEE Trans. on Systems, Man, and Cybernetics, Part C (Appl. and Rev.)*, Vol. 40, No. 6, hal. 601-618, Nov. 2010.
- [5] A. Algami, "Data Mining in Education," *Int. J. of Adv. Com. Sci. and Appl. (IJACSA)*, Vol. 7, No. 6, hal. 1-6, 2016.
- [6] J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques Third Edition*, Massachusetts, USA: Morgan Kaufmann Publishers, 2006.
- [7] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, dan O. Ragos, "Implementing AutoML in Educational Data Mining for Prediction Tasks," *Applied Sciences*, Vol. 10, No. 1, hal. 1-27, 2020.
- [8] A. Wolff, Z. Zdrahal, D. Herrmannova, dan P. Knoth, "Predicting Student Performance from Combined Data Sources," dalam *Educational Data Mining. Studies in Computational Intelligence*, Vol. 524, A. Peña-Ayala, Eds., Cham, Switzerland: Springer, 2014, hal. 175-202.
- [9] J.P. Campbell, P.B. DeBlois, dan D.G. Oblinger, "Academic Analytics: A New Tool for a New Era," *EDUCAUSE Review*, Vol. 42, No. 4, hal. 40-57, 2007.
- [10] B. Daniel, "Big Data and Analytics in Higher Education: Opportunities and Challenges," *British J. of Educ. Technol.*, Vol. 46, No. 5, hal. 904-920, 2015.
- [11] B. Guo, R. Zhang, G. Xu, C. Shi, dan L. Yang, "Predicting Students Performance in Educational Data Mining," *2015 Int. Symp. on Educ. Technol. (ISET)*, 2015, hal. 125-128.
- [12] X. Du, J. Yang, J.-L. Hung, dan B. Shelton, "Educational Data Mining: A Systematic Review of Research and Emerging Trends," *Information Discovery and Delivery*, Vol. 48, No. 4, hal. 225-236, 2020.
- [13] J.L. Rastrollo-Guerrero, J.A. Gómez-Pulido, dan A. Durán-Domínguez, "Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review," *Applied Sciences*, Vol. 10, No. 3, hal. 1-9, 2020.
- [14] C. Romero, S. Ventura, dan P. De Bra, "Knowledge discovery with genetic programming for providing feedback to courseware author. User Modeling and User-Adapted Interaction," *The J. of Personal. Research*, Vol. 14, No. 5, hal. 425-464, 2004.
- [15] C. Romero, S. Ventura, M. Pechenizkiy, dan R.S. Baker, *Handbook of Educational Data Mining*, Boca Raton, USA: CRC Press, 2010.
- [16] H. Nemati dan C. Barko, "Organizational Data Mining (ODM): An Introduction," dalam *Organizational Data Mining: Leveraging Enterprise Data Resources for Optimal Performance*, Pennsylvania, USA: Idea Group Publishing, 2004, hal. 1-8.
- [17] R.A. Huebner, "A Survey of Educational Data-Mining Research," *Research in Higher Education Journal*, Vol. 19, hal. 1-13, Apr. 2013.
- [18] A.M. Shahiri, W. Husain, dan N.A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," *Procedia Computer Science*, Vol. 72, hal. 414-422, 2015.
- [19] M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, M. Hasan, B.C. Van Essen, A.A.S. Awwal, dan V.K. Asari, "A State-of-the-Art Survey on Deep Learning Theory and Architectures," *Electronics*, Vol. 8, No. 3, hal. 1-3, 2019.
- [20] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," *Neural Netw.*, Vol. 61, hal. 85-117, 2015.
- [21] Y. LeCun, Y. Bengio, dan G. Hinton, "Deep Learning," *Nature*, Vol. 521, hal. 436-444, 2015.
- [22] Y. Bengio, A. Courville, dan P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Trans. on Pattern Analysis and Machine Intel.*, Vol. 35, No. 8, hal. 1798-1828, 2013.
- [23] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, dan D. Hassabis, "Human-Level Control Through Deep Reinforcement Learning," *Nature*, Vol. 518, hal. 529-533, 2015.
- [24] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, dan M. Riedmiller, "Playing Atari with Deep Reinforcement Learning," *arXiv*, arXiv:1312.5602, hal. 1-9, 2013.
- [25] B. Kitchenham dan S. Charters, "Guidelines for Performing Systematic Literature Reviews in Software Engineering," Keele University dan Durham University, UK, EBSE Technical Report, hal. 1-44, 2007.
- [26] C. Okoli dan K. Schabram, "A Guide to Conducting a Systematic Literature Review of Information Systems Research," *SSRN Electronic Journal*, Vol. 10, hal. 6-9, 2010.
- [27] N. Wu, L. Zhang, Y. Gao, M. Zhang, X. Sun, dan J. Feng, "CLMS-Net: Dropout Prediction in MOOCs with Deep Learning," *ACM TURC '19: Proc. of the ACM Turing Celebration Conf. - China*, 2019, hal. 1-6.
- [28] M. Karlık dan B. Karlık, "Prediction of Student's Performance with Deep Neural Networks," *Int. J. of Artificial Intel. and Expert Sys. (IJAE)*, Vol. 9, No. 2, hal. 39-47, Jun. 2020.
- [29] B.-H. Kim, "Deep Learning to Predict Student Outcomes," *ICLR AI for Social Good Workshop*, 2019, hal. 1-5.
- [30] W.W.T. Fok, Y.S. He, H.H.A. Yeung, K.Y. Law, K.H. Cheung, Y.Y. Ai, dan P. Ho, "Prediction Model for Students' Future Development by Deep Learning and Tensorflow Artificial Intelligence Engine," *2018 4th Int. Conf. Inf. Manag. ICIM 2018*, 2018, hal. 103-106.
- [31] N.R. Aljohani, A. Fayoumi, dan S.U. Hassan, "Predicting At-Risk Students Using Clickstream Data in the Virtual Learning Environment," *Sustain.*, Vol. 11, No. 24, hal. 1-12, 2019.
- [32] Bendangnugsung dan D. Prabu, "Students' Performance Prediction Using Deep Neural Network," *Int. J. Appl. Eng. Res.*, Vol. 13, No. 2, hal. 1171-1176, 2018.
- [33] R. Raga dan J. Raga, "Early Prediction of Student Performance in Blended Learning Courses Using Deep Neural Networks," *Proc. - 2019 Int. Symp. Educ. Technol. ISET 2019*, 2019, hal. 39-43.
- [34] A.P. Patil, K. Ganesan, dan A. Kanavalli, "Effective Deep Learning Model to Predict Student Grade Point Averages," *2017 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2017*, 2018, hal. 1-6.
- [35] A.S. Imran, F. Dalipi, dan Z. Kastrati, "Predicting Student Dropout in a MOOC: An Evaluation of a Deep Neural Network Model," *Proc. of the 2019 5th Int. Conf. on Computing and Artificial Intel. (ICCAI '19)*, 2019, hal. 190-195.
- [36] P. Sockhey dan T. Okazaki, "Comparative Study of Prediction Models on High School Student Performance in Mathematics," *34th Int. Tech. Conf. Circuits/Systems, Comput. Commun. ITC-CSCC 2019*, 2019, hal. 2-5.
- [37] M. El Fouki, N. Akin, dan K.E. El Kadiri, "Multidimensional Approach Based on Deep Learning to Improve the Prediction Performance of DNN Models," *Int. J. Emerg. Technol. Learn.*, Vol. 14, No. 2, hal. 30-41, 2019.
- [38] P. Sockhey dan T. Okazaki, "Development and Optimization of Deep Belief Networks Applied for Academic Performance Prediction with Larger Datasets," *IEIE Trans. Smart Process. Comput.*, Vol. 9, No. 4, hal. 298-311, 2020.
- [39] J. Sultana, M. Usha Rani, dan M. A. H. Farquad, "Student's Performance Prediction Using Deep Learning and Data Mining Methods," *Int. J. Recent Technol. Eng.*, Vol. 8, No. 1S4, hal. 1018-1021, 2019.
- [40] F. Okubo, A. Shimada, T. Yamashita, dan H. Ogata, "A Neural Network Approach for Students' Performance Prediction," *Proc. of the Seventh Int. Learn. Analytics & Knowl. Conf.*, 2017, hal. 598-599.
- [41] M. Akour, H. Al Sghaier, dan O. Al Qasem, "The Effectiveness of Using Deep Learning Algorithms in Predicting Students Achievements," *Indones. J. Electr. Eng. Comput. Sci.*, Vol. 19, No. 1, hal. 387-393, 2020.
- [42] H. Hussain, Z.F. Muhsin, Y.K. Salal, P. Theodorou, F. Kurtoglu, dan G.C. Hazarika, "Prediction Model on Student Performance Based on Internal Assessment Using Deep Learning," *Int. J. Emerg. Technol. Learn.*, Vol. 14, No. 8, hal. 4-22, 2019.
- [43] A. Buranasang, "Efficiency Assessment of Undergraduate Students Based on Academic Record Using Deep Learning Methodology," *Int. J. Inf. Educ. Technol.*, Vol. 10, No. 7, hal. 511-515, 2020.
- [44] A. Mondal dan J. Mukherjee, "An Approach to Predict a Student's Academic Performance Using Recurrent Neural Network (RNN)," *Int. J. Comput. Appl.*, Vol. 181, No. 6, hal. 1-5, 2018.
- [45] Y. Mao, C. Lin, dan M. Chi, "Deep Learning vs. Bayesian Knowledge Tracing: Student Models for Interventions," *J. Educ. Data Min.*, Vol. 10, No. 2, hal. 28-54, 2018.
- [46] S. Jain, V. Todwal, dan S.C. Jat, "Student Performance Assessment and Prediction Based on Machine Learning," *J. of the Gujarat Research Soc.*, Vol. 21, No. 16, hal. 1-7, 2019.