

BAB I

Pendahuluan

A. Latar Belakang

Performa siswa merujuk pada kinerja atau prestasi akademik yang ditunjukkan oleh seorang siswa selama periode tertentu. Tolak ukur dari performa siswa berasal dari nilai akademik atau nilai sikap siswa. Nilai tersebut didapatkan siswa pada satu semester atau tengah semester. Performa siswa memiliki peran dalam keberlangsungan kehidupan siswa di sekolah. Pada sebuah penelitian diketahui bahwa performa siswa merupakan tolak ukur dari sukses atau tidaknya sebuah pendidikan (Brew, Nketiah, dan Koranteng 2021). Selain sebagai penentu sukses tidaknya sebuah pendidikan. Dengan diadakannya analisis performa siswa dapat membantu pihak sekolah untuk menempatkan tenaga pendidik atau guru yang tepat pada siswa yang memiliki karakteristik performa tertentu (Pallathadka et al. 2023). Dengan diadakannya analisis performa siswa, pihak sekolah akan mendapatkan suatu hasil analisis yang berharga untuk membuat sebuah keputusan. Keputusan tersebut dapat berupa kapan pemberian motivasi atau peningkatan pemberdayaan siswa dengan tujuan menjadikan siswa sukses dalam pendidikan sekolah (Kim, Vizitei, dan Ganapathi 2018). Analisis performa siswa dapat berbentuk sebagai klasifikasi atau prediksi performa siswa. Pengukuran performa siswa salah satunya berasal dari nilai akademik siswa. Nilai tersebut diperoleh dari nilai ujian akhir semester atau ujian tengah semester siswa. Penggunaan nilai tengah semester ditunjukkan untuk mengetahui bagaimana performa siswa dalam rentang waktu tengah semester. Prediksi performa siswa dalam bentuk nilai tengah semester memiliki beberapa keuntungan tersendiri, diantaranya adalah pengambilan keputusan yang berdampak pada siswa karena didasari dengan prediksi, selain itu prediksi performa siswa dalam bentuk nilai tengah semester juga membantu guru dalam mengetahui nilai siswa baru, berdasarkan hal tersebut guru dapat membuat lingkungan belajar yang cocok dengan siswa baru tersebut.

Berbagai faktor mempengaruhi performa siswa. Beberapa diantara faktor – faktor tersebut adalah lingkungan tempat belajar, tersedianya peralatan yang menunjang pembelajaran seperti buku, meja dan kursi baik itu di rumah ataupun di sekolah, dan kebutuhan finansial yang digunakan untuk menempuh jenjang sekolah di tingkat manapun, serta peran orang tua sebagai pendamping yang membantu perkembangan diri siswa (Chinyoka dan Naidu 2013). Selain peran orang terdekat dan lingkungan, guru juga menjadi peran penting yang memiliki pengaruh pada performa siswa. Terdapat penelitian yang mengatakan bahwa guru atau dosen yang pengalamannya lebih senior memiliki efek positif terhadap performa siswa (Tuckman 1975). Dari sini diketahui bahwa penempatan guru yang sesuai dengan performa siswa akan memiliki efek positif yang mampu meningkatkan performa siswa kedepan. Sehingga dengan melakukan analisis pada performa siswa pihak institusi pendidikan mampu untuk menempatkan guru atau tenaga pendidik lainnya yang tepat berdasarkan performa dari siswa.

Dengan berbagai faktor yang mempengaruhi performa siswa membuat proses analisisnya menjadi semakin rumit dan kemungkinan tidak bisa diselesaikan menggunakan sembarang metode. Selain berbagai faktor yang mempengaruhi, analisis performa siswa juga menggunakan data yang memiliki jumlah besar dikarenakan jumlah siswa dalam satu sekolah itu juga tergolong tinggi. Salah satu teknik yang dapat digunakan untuk data yang berjumlah banyak dan memiliki berbagai fitur serta memiliki kinerja yang tergolong optimal adalah *Machine Learning*.

Menurut Samael (1959), *machine learning* adalah suatu cabang ilmu yang memberi komputer kemampuan untuk belajar tanpa pemrograman secara eksplisit. Metode *machine learning* berbeda dengan pemrograman tradisional yang memiliki beberapa keterbatasan. Pemrograman tradisional menggunakan aturan dan data supaya memperoleh solusi. Sehingga ketika memperoleh *input* yang terlalu kompleks menjadi masalah.

Dengan menerapkan teknik ini pengguna atau pihak sekolah mampu menemukan struktur yang tersembunyi dari data kemudian menggunakan data

tersebut untuk melakukan prediksi (Zhou et al. 2017). Selain itu data yang digunakan untuk melakukan analisis pada performa siswa tergolong sebagai data yang selalu diperbarui dikarenakan setiap tahun siswa akan naik kelas dan, siswa baru masuk sekolah, sehingga dengan menerapkan teknik *machine learning* pengguna memiliki keuntungan tersendiri. Hal ini mengacu dari cara kerja *machine learning*. Dimana *machine learning* mampu untuk beradaptasi pada kondisi baru yang diberikan kepadanya (Misilmani dan Naous 2019). Dengan beberapa keuntungan tersebut. *Machine learning* dapat diterapkan pada dunia pendidikan sehingga pihak yang terlibat seperti guru pengampu mata pelajaran dapat melakukan prediksi ataupun klasifikasi pada performa siswa. Hal ini memiliki keuntungan tersendiri yaitu membantu institusi pendidikan ataupun guru untuk mengetahui permasalahan dari siswa ketika pembelajaran di sekolah dan membantu menangani masalah tersebut (Albreiki, Zaki, dan Alashwal 2021).

Sebelum adanya penerapan teknik *machine learning* dalam dunia pendidikan terdapat teknik – teknik lain yang digunakan untuk menganalisis performa dari siswa. Seperti teknik pada model statistika klasik terdapat teknik *predictive Structural Equation Modeling* (pSEM). Meskipun begitu teknik ini masih kurang efektif dibandingkan dengan *machine learning*. Hal tersebut dapat dilihat dari perbandingan antara lima metode dari *machine learning* dengan pSEM untuk memprediksi performa siswa. Diperoleh hasil akurasi metode *Rdnom Forest* yang merupakan salah satu metode *machine learning* mencapai hasil akurasi tertinggi yaitu 80.58% dibandingkan dengan pSEM yang mencapai 56.40% (Sokkhey dan Okazaki 2019).

Terdapat banyak algoritma dalam *machine learning* yang digunakan untuk tujuan prediksi ataupun klasifikasi, seperti RNN (*Recurrent Neural Network*), ANN (*Artificial Neural Network*), pohon keputusan, *Extreme Gradient Boost* berbasis pohon keputusan, dan masih banyak lagi.

RNN merupakan salah satu metode dari model *machine learning* yang diawasi (*Supervised Learning*) (Salehinejad et al. 2018). Beberapa contoh penggunaan RNN yaitu untuk *speech recognition*, *time series prediction*, atau

machine translation (Sherstinsky 2020). Namun ketika data yang digunakan pada RNN membutuhkan jarak yang jauh pada setiap memorinya akan terjadi fenomena yang dinamakan ledakan gradien. Kejadian tersebut akan sangat mempengaruhi hubungan antara ketergantungan jangka panjang pada data (Salehinejad et al. 2018).

Selain RNN terdapat LSTM yang sama menggunakan prinsip dari jaringan syaraf tiruan untuk bekerja. LSTM cocok digunakan untuk mengklasifikasi ataupun memprediksi data berjenis *time series* dengan jangka waktu yang tidak tentu (Bashaiwth, Binsalleeh, dan AsSadhan 2023). Adapun kelebihan dari LSTM adalah kemampuannya untuk mempelajari representasi hierarki dari data input. Setiap lapisan LSTM dapat menangkap tingkatan abstraksi yang berbeda, memungkinkan model untuk memahami pola dan ketergantungan kompleks dalam data. Beberapa lapisan LSTM meningkatkan kapasitas model untuk belajar dari data, sehingga bermanfaat ketika menangani tugas atau dataset yang kompleks. Lapisan LSTM dapat membantu menangkap dan mempertahankan ketergantungan jangka panjang (*Long Term Dependencies*) dalam data, yang sangat penting untuk tugas yang melibatkan data berurutan atau time-series (Siami-Namini, Tavakoli, dan Namin 2019).

Selain LSTM, metode lain dalam *machine learning* yang digunakan untuk kegiatan prediksi ataupun klasifikasi adalah *Extreme Gradient Boost*. Prinsip kerja dari *Extreme Gradient Boost* yaitu berbasis pohon keputusan atau penggabungan pohon keputusan. Desain dari *Extreme Gradient Boost* sendiri dibuat untuk lebih efisien, fleksibel dan portabel. Penggunaan dari *Extreme Gradient Boost* sendiri telah membantu menyelesaikan masalah dari bidang *data science* seperti regresi hingga klasifikasi (Chen dan Guestrin 2016).

Penelitian-penelitian mengenai penggunaan metode *machine learning* telah dilakukan di dunia pendidikan dan menggunakan banyak algoritma. Beberapa diantaranya adalah penelitian oleh Sugin, mengenai penggunaan teknik *data mining* untuk menganalisis performa akademik siswa (Sugin Lal dan Porkodi 2023), teknik *data mining* yang digunakan dalam penelitian tersebut

adalah LSTM dikombinasikan dengan *Boltzmann Sparse Probabilities - Conditional Random Field (BSP-CRF)*. Penelitian yang lain dilakukan oleh Waheed dan kawan-kawan mengenai performa yang timbul dari interaksi antara siswa dengan teknologi pendidikan yang berfokus pada performa siswa dengan menggunakan model *deep learning* (Waheed et al. 2023), model yang digunakan adalah *deep LSTM*. Kemudian terdapat penelitian oleh Subitha Sivakumar dan Rajalakshmi Selvaraj yang menggunakan model *supervised learning classifier* yaitu *Decision Tree, Support Vector Machine, K-Nearest Neighbours, Naïve Bayes, Neural Network* dan *Decision Tree* yang telah ditingkatkan untuk mengklasifikasikan performa dari siswa dalam kategori baik, rata-rata, sangat baik, dan buruk (Sivakumar dan Selvaraj 2018).

Namun dari banyaknya algoritma tersebut, algoritma hanya akan dapat berfungsi dengan efisien ketika data yang digunakan cocok dengannya. Contohnya adalah LSTM yang dapat bekerja serta memiliki performa yang maksimal ketika data yang digunakan berjenis *time series* atau urutan waktu.. Prediksi ini memiliki tujuan untuk membantu pihak institusi pendidikan atau guru untuk mengetahui performa dari siswa ke depannya. Selain itu pihak guru maupun institusi pendidikan atau sekolah dapat membantu siswa untuk berkembang. Untuk membantu menjalankan metode LSTM dan XGBoost peneliti menggunakan bahasa pemrograman python dengan bantuan *website google collaboratory*, penggunaan bahasa dan *website* tersebut bertujuan mempermudah peneliti dalam melakukan prediksi menggunakan metode LSTM dan XGBoost. Penelitian ini dilaksanakan di SMAN 1 Kediri dengan data yang diambil sebanyak 4 kelas.

B. Rumusan Masalah

Bagaimana implementasi *machine learning* dengan metode LSTM dan XGBoost dalam memprediksi nilai performa siswa

C. Tujuan Penelitian

Berdasarkan paparan masalah di atas, tujuan diadakannya penelitian ini adalah untuk mengetahui implementasi *machine learning* dengan metode LSTM dan XGBoost dalam memprediksi performa siswa

D. Manfaat Penelitian

Manfaat Teoritis :

Memberikan kontribusi pada pengetahuan mengenai penerapan *machine learning* khususnya dalam bidang pendidikan, baik itu untuk melakukan klasifikasi, prediksi, ataupun melakukan analisis sentimen

Manfaat Praktis :

1. Memberikan guru kesempatan untuk mengetahui keberhasilan dari pemberlajaran yang dilakukan
2. Memberikan pihak sekolah sebuah kemampuan untuk mengambil keputusan yang sesuai dengan keadaan siswa

E. Penelitian Terdahulu

:

Tabel 1 Penelitian Terdahulu

No	Nama Peneliti/ Tahun	Judul	Hasil	Persamaan	Perbedaan
1	Mostafa Zafari, Abolghasem Sadeghi-Niaraki , Soo-Mi Choi dan Ali Esmaeily	A Practical Model for the Evaluation of High School Student Performance Based on Machine Learning	Penelitian ini menggunakan teknik – teknik klasifikasi diantaranya adalah <i>Rdnom Forest, Support Vector Machine, Logistic Regression, Artificial Neural Network</i> , untuk digunakan pada performa siswa, diperoleh bahwa ketika fitur-fitur penting dihilangkan dari data <i>Support Vector Machine</i> mampu meraih akurasi lebih tinggi dari yang lainnya.	<ol style="list-style-type: none"> 1. Menggunakan teknik-teknik dari <i>machine learning</i> untuk menganalisis data 2. Objek yang digunakan masih berhubungan dengan siswa yaitu performa siswa 	<ol style="list-style-type: none"> 1. Pada penelitian ini digunakan prediksi serta analisis sentimen yang dilakukan 2. Penelitian ini menggunakan data dari siswa yang berada di lingkup kabupaten atau kota kediri
2	G.Sugin Lal, Dr. R. Porkodi, 2023	Assessment of Student Educational Performance Analysis for Feature Extraction dan Classification with LSTM-Based Deep Learning Model	Penelitian ini menggunakan LSTM sebagai metode untuk mengklasifikasikan performa siswa. Hasil dari penelitian ini diperoleh nilai akurasi yang lebih tinggi dibandingkan dengan model tradisional seperti VGG, ResNet, dan Inception, akurasi yang diperoleh model yaitu 97%, model ini akan sangat membantu guru untuk mengidentifikasi arena dimana siswa perlu bantuan.	<ol style="list-style-type: none"> 1. Menggunakan algoritma <i>deep learning LSTM</i> 2. Mengklasifikasikan performa siswa untuk membantu suru 	<ol style="list-style-type: none"> 1. Menggunakan Extreme Gradient Boost sebagai model klasifikasi dan LSTM sebagai model prediksi 2. Menggunakan <i>k-clustering</i> untuk mengklasifikasikan performa siswa
3	Hajra Waheeda, Saeed-Ul Hassanb, Raheel Nawazb, Naif R. Aljohanic, Guanliang Chend, Dragan Gasevic, 2022	Early prediction of learners at risk in self-paced education: A neural network approach	Penelitian ini membahas mengenai penggunaan LSTM untuk memprediksi risiko kegagalan siswa dalam pembelajaran daring secara mandiri, model ini mencapai akurasi sebesar 84.57%, presisi sebesar 82.24%. Berdasarkan hasil penelitian ini diketahui bahwa model LSTM mampu melampaui model alternative lainnya	<ol style="list-style-type: none"> 1. Menggunakan algoritma <i>deep learning LSTM</i> 2. Menggunakan LSTM sebagai model prediksi 	<ol style="list-style-type: none"> 1. Menggunakan teknik <i>natural language processing</i> 2. Dilakukannya analisis sentimen untuk mengetahui performa siswa
4	Subitha Sivakumar, Rajalakshmi Selvaraj, 2018	Predictive Modeling of Students Performance Through the	Penelitian ini menggunakan beberapa model klasifikasi <i>supervised</i> yaitu <i>Decision Tree, Support Vector Machine, K-Nearest Neighbours, Naïve Bayes, Neural Network</i>	<ol style="list-style-type: none"> 1. Menggunakan model klasifikasi <i>supervised learning</i> 2. Mengklasifikasikan 	<ol style="list-style-type: none"> 1. Menggunakan teknik <i>natural language processing</i> 2. Dilakukannya analisis

		Enhanced Decision Tree	dan <i>Decision Tree</i> yang telah ditingkatkan, diperoleh bahwa <i>Decision tree</i> yang kemampuannya telah ditingkatkan bekerja lebih optimal dibandingkan model lainnya	siswa dengan kategori tertentu	senetimen untuk mengetahui performa siswa
5	Selvaprabu Jeganathan, Arun Raj Lakshminarayanan, Ndanhakumar Ramachdanran, 2022	Predicting Academic Performance of Immigrant Students Using Extreme Gradient Boost Regressor	Penelitian ini menggunakan Extreme Gradient Boost sebagai teknik untuk memprediksi performa dari siswa khususnya siswa yang merupakan seorang <i>immigrant</i> . Hasil dari penelitian ini diketahui bahwa penggunaan Extreme Gradient Boost sebagai teknik prediksi bagus dengan menggunakan <i>hyper parameters</i>	<ol style="list-style-type: none"> 1. Menggunakan model prediksi dari <i>supervised learning</i> yaitu Extreme Gradient Boost 2. Menggunakan teknik <i>machine learning</i> untuk memprediksi performa dari siswa 	<ol style="list-style-type: none"> 1. Menggunakan LSTM untuk teknik memprediksi 2. Menerapkan NLP (<i>Natural Language Processing</i>) sebagai metode untuk melakukan analisis sentimen
6	Nur Izzati Mohd Talib, Nazatul Aini Abd Majid, Shahnorbanun Sahran, 2023	Identification of Student Behavioral Patterns in Higher Education Using K-Means Clustering dan Support Vector Machine	Penelitian ini menggunakan metode klasifikasi dan <i>clustering</i> untuk membuat <i>cluster</i> dari performa siswa serta mengklasifikasikannya, teknik yang digunakan yaitu <i>Support Vector Machine</i> untuk klasifikasi dan untuk <i>clustering</i> menggunakan <i>K-means clustering</i> . Diperoleh hasil yang optimal ketika menggunakan <i>cluster</i> yang diperoleh dari teknik tersebut.	<ol style="list-style-type: none"> 1. Menggunakan teknik dari <i>machine learning</i> 2. Siswa merupakan objek yang dikaji dalam penelitian 	<ol style="list-style-type: none"> 1. Menggunakan Extreme Gradient Boost sebagai teknik untuk mengklasifikasi 2. Menggunakan LSTM sebagai teknik untuk prediksi 3. Menggunakan teknik analisis sentimen
7	Abdulaziz Salamah Aljaloud, Diao Mohammed Uliyan1 , Adel Alkhalil, Magdy Abd Elrhman, Azizah Fhad Mohammed Alogali, Yaser Mohammed Altameemi , Mohammed Altamimi1 , Dan Paul Kwan, 2022	A Deep Learning Model to Predict Student Learning Outcomes in LMS Using CNN dan LSTM	Penelitian ini menggunakan kombinasi CNN-LSTM untuk memprediksi performa dari siswa yang mengikuti atau menggunakan <i>Learning Management System</i> (LMS) secara daring berdasarkan KPI (<i>Key Performance Indicator</i>). Diperoleh bahwa kombinasi CNN-LSTM mampu memperoleh skor presisi sebesar 94.2%	<ol style="list-style-type: none"> 1. Menggunakan LSTM sebagai teknik untuk memprediksi 2. Objek yang digunakan berhubungan dengan siswa atau pendidikan 	<ol style="list-style-type: none"> 1. Menggunakan Extreme Gradient Boost sebagai teknik untuk mengklasifikasi 2. Menggunakan teknik analisis sentimen

(Sumber : Dokumen Penulis)

F. Definisi Konsep

1. LSTM (*Long Short Terms Memory*)

LSTM merupakan salah satu teknik *machine learning* yang menggunakan prinsip dari jaringan syaraf manusia untuk bekerja. LSTM dirancang dari pendahulunya yaitu RNN (*Rcurrent Neural Network*) dimana LSTM memiliki peningkatan yaitu kemampuan dari *Long Terms Dependencies* yang memungkinkannya untuk memiliki ingatan mengenai memori terhadap data yang sudah diinputkan lebih awal. Penerapan model LSTM pada penelitian ini menggunakan bantuan dari bahasa pemrograman *python* dengan *library* dari *tensorflow*

2. XGBoost (*Extreme Gradient Boost*)

Berbeda dengan LSTM yang menggunakan prinsip kerja dari jaringan syaraf XGBoost menggunakan prinsip kerja dari pohon keputusan. XGBoost berasal dari teknik atau metode *Gradient Boost* yang lebih ditingkatkan. XGBoost sendiri memiliki kelebihan untuk mampu meningkatkan nilai error menjadi lebih kecil atau akurasi yang lebih tinggi, selain itu XGBoost memiliki tingkat fleksibilitas terhadap data yang membuat model dapat bekerja pada data bertipe apapun. Pada penelitian ini XGBoost akan diterapkan menggunakan bahasa pemrograman *python* dan *library* dari XGBoost sendiri

3. Performa Siswa

Performa siswa merujuk pada tingkat pencapaian dan kemajuan siswa dalam konteks pendidikan. Evaluasi performa siswa melibatkan penilaian berbagai aspek, termasuk pemahaman materi, keterampilan penerapan pengetahuan, partisipasi aktif, dan perkembangan kemampuan lainnya. Performa siswa dapat diukur dari nilai akademik yang dimiliki oleh siswa misalnya saja nilai praktik, nilai ulangan harian, nilai ujian tengah semester, dan nilai ujian akhir semester. Pada penelitian ini nilai ujian tengah semester menjadi objek yang akan diprediksi

G. Batasan Masalah

- a. Pada penelitian ini nilai performa siswa adalah nilai uts
- b. Data yang digunakan berjumlah 100
- c. Metode prediksi yang digunakan yaitu LSTM dan XGBoost