

Prediksi Keterserapan Siswa SMK Pada Dunia Industri Dengan Pendekatan Educational Data Mining

Kuri Asih*, Lili Ayu Wulandhari

Computer Science Department,
Universitas Bina Nusantara, Jakarta, Indonesia

*kuriasih@binus.ac.id

Abstrak—Tingginya tingkat penyerapan tenaga kerja siswa kejuruan sangat menentukan kualitas suatu Sekolah Menengah Kejuruan (SMK). Semakin banyak siswa yang terserap ke dunia kerja dan semakin cepat bekerja setelah lulus maka semakin baik bagi SMK tersebut. Data mining adalah solusi yang berguna untuk mengidentifikasi pola tersembunyi dan memberikan saran untuk meningkatkan kinerja siswa. Penelitian ini menggunakan data rapor dari 167 siswa jurusan Teknik Jaringan Komputer SMK Negeri 26 Jakarta selama enam semester, dari tiga angkatan siswa yang lulus tahun 2015 hingga 2017. Penelitian ini menggunakan model SVR dan ANN dan metode Mean Absolute Error (MAE). Hasil penelitian menunjukkan bahwa ANN dengan data seluruh fitur yang digunakan, dengan model normalisasi Standard Scaler, dan algoritma aktivasi Relu, jumlah Neuron sebanyak 128 dan Iter Max 150 menunjukkan performa terbaik, yaitu MAE sebesar 2,2 bulan. Heatmap korelasi Pearson mengungkapkan bahwa semua mata pelajaran yang sangat erat hubungannya dan mempengaruhi jumlah serapan mahasiswa di dunia kerja adalah mata pelajaran produktif (vokasi) pada semester 1 & 2 pada aspek penilaian keterampilan (praktik). Untuk meningkatkan angka penyerapan tenaga kerja, mahasiswa harus mempertajam dan memperdalam kompetensi mata pelajaran praktik vokasi pada awal semester. Hasil penelitian ini dapat dijadikan acuan untuk memprediksi penyerapan lulusan SMK di dunia kerja dan sebagai langkah antisipatif untuk meningkatkan nilai kompetensi sebelum memasuki dunia kerja.

Kata Kunci—Data mining, MAE, Prediksi keterserapan, Regresi

DOI: 10.22441/jte.2024.v15i1.011

I. PENDAHULUAN

Permasalahan yang dihadapi tentang keterserapan kerja siswa SMK Negeri 26 Jakarta adalah banyak sekali data-data yang dihasilkan oleh divisi tersebut di sekolah, namun belum ada satupun yang dapat diolah dan diambil insight-nya untuk kebermanfaatan kemajuan dan perkembangan sekolah. Penelitian ini mengambil topik tentang prediksi keterserapan siswa SMK di dunia industri dengan pendekatan Educational Data mining (EDM).

Penyelesaian permasalahan pertama yang dihadapi yaitu melakukan analisis terhadap nilai akademis dan non akademis siswa serta pengaruhnya terhadap keterserapan lulusan di dunia industri, maka cara yang dilakukan adalah dengan mengklasifikasi data nilai akademis siswa, antara nilai

pengetahuan dan nilai keterampilan. Evaluasi secara menyeluruh sejak awal siswa belajar dari semester 1 sampai semester 6. Penelitian ini berguna untuk memprediksi kinerja siswa, sehingga siswa yang beresiko gagal dapat terdeteksi lebih awal, dan dapat diberikan perhatian khusus oleh guru, serta diberikan motivasi dan pengayaan dalam hal akademis dan softskill sehingga dapat meningkat menjadi lebih baik lagi.

Selain itu, untuk memprediksi keterserapan siswa di dunia kerja, maka digunakan nilai dari semester 1 sampai semester 6 dengan model regresi waktu (siswa terserap di dunia kerja pada bulan ke-berapa). Data yang dibutuhkan adalah data nilai akademis siswa, nilai absensi serta data nilai keaktifan siswa dalam kegiatan ekstrakurikuler.

Data yang dibutuhkan berupa data penelusuran tamatan siswa dari jurusan Teknik Komputer Jaringan (TKJ), yang didapat dari kordinator PKL siswa. Data tersebut berisi tentang nama siswa (alumni), nama perusahaan dan keterangan siswa yang sudah langsung terserap di dunia kerja dalam jangka waktu 0 bulan sampai 18 bulan setelah kelulusan.

Algoritma yang digunakan adalah algoritma data mining dengan menggunakan metode Support Vector Regression (SVR) dan Artificial Neural Network (ANN), berdasarkan referensi dari paper -paper terkait. Hasilnya, diharapkan akan dapat memberikan keluaran (output) berupa model prediksi waktu keterserapan lulusan SMK di dunia kerja. Tujuan penelitian ini adalah untuk memprediksi rentang waktu keterserapan siswa, dihitung dari waktu kelulusan mereka.

Data mining adalah proses penambangan data pada kumpulan besar data dengan melibatkan integrasi teknik dari berbagai disiplin ilmu. Metode ini sangat berguna dalam meningkatkan kualitas pendidikan dengan menemukan model prediksi dan menganalisis data. Data mining membantu mengungkap pengetahuan tersembunyi dalam data, berlaku di berbagai area penelitian, termasuk pendidikan. Sistem pendidikan menjadi domain penting dalam bidang ini, fokusnya adalah evaluasi dan peningkatan kualitas organisasi pendidikan [1], [2]. Data mining telah menjadi alat penilaian cerdas yang efektif dalam pembelajaran pada era modern. Dengan kemampuannya untuk memantau kinerja siswa secara cermat, data mining memungkinkan identifikasi kemajuan siswa dan prediksi kinerja akademik mereka dengan akurat [3]. Khususnya dalam bidang pendidikan, data mining menunjukkan kinerja

yang luar biasa dalam pengolahan data dan populer karena kemampuannya untuk memprediksi berdasarkan data yang ada [4]. Meskipun umumnya digunakan untuk memprediksi kinerja akademik siswa, perhatian terhadap prediksi ketersediaan kerja siswa di SMK masih terbatas, meskipun SMK saat ini berfokus pada persiapan ketersediaan kerja [4].

Menurut Larose & Larose [5] dan Susanto dan Suryadi [6] fungsi data mining terdiri atas fungsi deskripsi (description) untuk memahami data yang diteliti dengan tujuan menemukan pola yang tersembunyi. Pola-pola yang repetitif dan memiliki nilai membantu dalam mengidentifikasi karakteristik data yang relevan. Kemudian, fungsi estimasi (estimation) adalah metode untuk memperkirakan nilai sebuah populasi dengan menggunakan nilai sampel, untuk mendapatkan keputusan terbaik untuk menentukan langkah berikutnya [7]. Ketiga adalah fungsi prediksi (prediction) untuk memproses data dan menemukan pola khusus dari sebuah data dan beberapa variabel pendukung data. Jika sebuah pola sudah ditemukan, maka pola tersebut dapat digunakan untuk memprediksi variabel lain yang nilainya belum diketahui. Keempat yaitu fungsi klasifikasi (classification) yang merupakan teknik untuk menemukan aturan yang membagi data ke dalam kelompok yang berbeda. Prosesnya adalah dengan membuat kumpulan pengamatan serupa dari sebuah database yang besar. Algoritma yang dipakai adalah classifier [8]. adalah fungsi klusterisasi (clusterisation) yang merupakan proses identifikasi kelompok objek tertentu yang memiliki kesamaan dan karakteristik khusus [8].

II. PENELITIAN TERKAIT

Python adalah bahasa pemrograman serbaguna yang ditujukan untuk analisis data statistik serta berbagai aplikasi lainnya seperti Internet, situs web, basis data, desktop GUI, perhitungan ilmiah, dan pengembangan perangkat lunak dan game [9]. Dikembangkan Dikembangkan oleh Guido van Rossum, Python adalah bahasa pemrograman berorientasi objek tingkat tinggi dengan pustaka standar yang besar, mendukung berbagai sistem operasi dan basis data, serta bersifat multi-paradigma [10]. Fleksibel, mudah dipahami, dan ramah pengguna, Python juga dikenal karena kinerja cepat, kekuatan, dan dukungannya terhadap teknologi lain [10].

Penelitian "Educational data mining: prediction of students' academic performance using machine learning algorithms" oleh Yağcı [11] menggunakan model data mining dengan berbagai algoritma seperti Random Forest, Neural Network, Logistic Regression, Support Vector Machine, Naïve Bayes, dan k Nearest Neighbour. Hasil penelitian menunjukkan bahwa Algoritma RF, NN, dan SVM memiliki akurasi klasifikasi tertinggi, yaitu 73%. Model ini diharapkan dapat menjadi prediktor yang akurat untuk prestasi akademik siswa di masa mendatang [11].

Penelitian oleh Qiu et al. [12] menggunakan Behavior Classification-based E-Learning Performance (BCEP) untuk memprediksi perilaku data pada E-Learning. Hasilnya menunjukkan bahwa BCEP memiliki efek prediksi yang baik terhadap kinerja pembelajaran. Selain itu, model Process-Behavior Classification (PBC) juga menunjukkan kinerja yang

lebih baik dalam memprediksi kinerja pembelajaran dibandingkan metode klasifikasi tradisional [12].

Penelitian oleh Amjad et al. [13] bertujuan untuk memprediksi kinerja siswa dengan menggunakan teknologi dan media sosial. Hasilnya menunjukkan bahwa siswa yang menggunakan media sosial hanya di akhir pekan memiliki kinerja yang lebih baik dibandingkan dengan yang menggunakan media sosial setiap hari kerja. Decision Tree memiliki hasil terbaik dengan akurasi 90%, mengungguli berbagai algoritma lain seperti Random Forest, Support Vector Machine, Stochastic Gradient Descent (SGD) classifier, Ada Boost classifier, dan Logistic Regression (LR) classifier.

Penelitian Nahar et al. [14] menggunakan Educational Data mining (EDM) untuk memprediksi performa mahasiswa. Hasilnya menunjukkan bahwa performa akhir mahasiswa dari mata kuliah tertentu terkait erat dengan kinerja mata kuliah prasyaratnya, serta kinerja mahasiswa hingga Ujian Tengah Semester (UTS). Mereka menggunakan Decision Tree dan Naïve Bayes, dan Naïve Bayes menunjukkan performa dengan akurasi 75% [14].

Harris Hawks Optimization (HHO) dengan algoritma KNN digunakan oleh peneliti Turabieh et al. (2021) sebagai feature selection dalam memprediksi kinerja siswa, dan membandingkan hasil antara HHO asli dan HHO modifikasi sebagai fitur yang berpengaruh terhadap kinerja siswa [15].

Dalam penelitiannya Rivas et al., (2020), mengusulkan ANN dan Decision Tree untuk mengolah variabel yang mempengaruhi performa akademik siswa dalam lingkungan pembelajaran virtual berbasis Moodle (MOOC) menggunakan teknik machine learning [16].

Teknik machine learning digunakan oleh Kemper et al. [17] untuk memprediksi siswa yang berpotensi drop out dengan algoritma Decision Tree dan Logistic Regression, dan hasilnya walaupun membutuhkan dataset yang lebih kompleks, namun Decision Tree menjanjikan hasil yang selaras dengan prediksinya.

Penelitian oleh Waheed et al. [18] membandingkan algoritma Artificial Neural Network (ANN), Support Vector Machine (SVM), dan Logistic Regression dalam memprediksi performa akademik siswa menggunakan VLE Big Data. Hasilnya menunjukkan bahwa deep ANN memiliki akurasi klasifikasi terbaik, mencapai 84%-93%, melebihi akurasi SVM (79.95% - 89.14%) dan logistic regression (79.82% - 85.60%) [18].

Penelitian oleh Vijayalakshmi & Venkatachalapathy [19] menunjukkan bahwa Deep Neural Network (DNN) mampu memproses dan memprediksi keunggulan akademik dan kinerja siswa dengan akurasi 85% menggunakan Bahasa pemrograman R. Mereka menggunakan model DNN dengan dua lapisan tersembunyi, di mana lapisan pertama memiliki sepuluh neuron dan lapisan kedua memiliki lima neuron [19].

Penelitian oleh Alqurashi (2019) menggunakan model prediktif untuk memprediksi kepuasan 167 siswa dalam pembelajaran online dengan memeriksa variabel konten kursus dan interaksi. Hasilnya menunjukkan bahwa self-efficacy, LCI, dan LII secara signifikan dapat memprediksi dan menjelaskan

tingkat kepuasan siswa serta pembelajaran yang dirasakan (Alqurashi, 2019).

Penelitian oleh Ranjeeth et al. [20] membandingkan performa model jaringan syaraf tiruan ANNQ3H dengan beberapa algoritma lain untuk memprediksi kinerja siswa tingkat sekolah menengah. Hasilnya menunjukkan bahwa tiga kinerja terbaik adalah ANNQ3H (93,3%), J48 (90,5%), dan SVM (86,7%), di mana ANNQ3H merupakan yang terbaik.

Penelitian oleh Mahmoud & Zohair (2019) menggunakan algoritma visualisasi dan clustering untuk mengolah dataset kecil dalam memprediksi kinerja siswa yang berisiko. Metode ini terbukti efisien dengan tingkat akurasi dan reliabilitas klasifikasi yang dapat diterima. Mereka melakukan simulasi kuantitatif dengan berbagai algoritma klasifikasi dan memilih model paling akurat untuk memprediksi kinerja mahasiswa. Penelitian ini bertujuan untuk meningkatkan peringkat dan reputasi universitas dengan memastikan penyediaan sumber belajar dan pengalaman yang baik bagi siswa.

Penelitian oleh Burgos et al. (2018) menggunakan Teknik Logistics Regression untuk membuat model referensi untuk nilai semua murid dalam dataset training. Hasilnya menunjukkan bahwa model yang digunakan bersama dengan perencanaan pengajaran sangat bermanfaat dalam mengurangi tingkat drop-out siswa pada tahun pelajaran 2014/2015 [22].

III. METODE

A. EDM

EDM merupakan alat untuk analisis akademis dan pemahaman mendalam tentang siswa. Mulai dikenal pada 2005, EDM memungkinkan analisis kinerja siswa dan pengembangan strategi masa depan. Fokus utama EDM adalah memprediksi kinerja siswa dengan efektif. Implementasinya merupakan pengembangan dari teknik penambangan data tradisional, memungkinkan pemecahan masalah di bidang pendidikan. Meskipun sebagian besar teknik data mining dapat diterapkan langsung, beberapa memerlukan penyesuaian karena kekhususan lingkungan pendidikan. Dengan menggunakan basis data sistem pendidikan, EDM dapat merancang kebijakan pendidikan yang lebih baik untuk meningkatkan kinerja akademik siswa dan mengurangi tingkat kegagalan [2], [13], [23], [24].

B. SVR

Sementara itu, SVR adalah turunan SVM untuk regresi dalam machine learning dan data mining. Berbeda dengan SVM yang membagi dataset menjadi dua zona dengan hyperplane, SVR menangani regresi linier dan non-linier, menghasilkan output bilangan riil. Algoritma ini efektif mengatasi overfitting, menemukan hyperplane terbaik dengan mengukur margin antara hyperplane dan data terdekat disebut support vector. Ini memungkinkan prediksi variabel bernilai nyata yang berkelanjutan dan efisien pada data berdimensi tinggi, meminimalkan nilai epsilon untuk memastikan semua dataset masuk ke dalam satu zona [25], [26].

C. ANN

ANN merupakan salah satu jenis dari metode Supervised Learning. Artificial Neural Network (ANN) atau Jaringan Syaraf Tiruan (JST) merupakan replika dari sistem syaraf (neuron) pada jaringan otak manusia. Otak manusia terdiri dari milyaran neuron yang terkoneksi dan saling berhubungan satu sama lain (Wulandhari, 2017). Menurut Wulandhari (2017) dalam tulisannya, neuron biologis terdiri dari tiga komponen utama, diantaranya yaitu Dendrit yang merupakan saluran sinyal masukan (input) yang kekuatannya terkoneksi ke nucleus dan dipengaruhi oleh bobot sinapsis. Kemudian Cell Body (Badan Sel) sebagai tempat perhitungan sinyal input dan bobot yang menghasilkan sinyal keluaran yang selanjutnya akan dikirimkan ke neuron lain. Terakhir, Axon yang merupakan bagian yang mengirimkan sinyal output ke neuron lain yang terhubung dengannya. Berdasarkan representasi ini, JST dibagi menjadi beberapa elemen sebagai berikut, yaitu (1) Input Layer, lapisan masukan ini terdiri dari neuron-neuron yang merepresentasikan fitur-fitur atau variabel untuk pemecahan masalah, (2) Output Layer, lapisan keluaran ini terdiri dari neuron-neuron yang merepresentasikan hasil perhitungan dalam badan sel, (3) Weight, merupakan bobot yang mengindikasikan kekuatan koneksi antar neuron, (4) Activation Function, fungsi aktivasi untuk mendapatkan keluaran sesuai dengan masalah yang akan diselesaikan, (5) Learning Function, sebagai fungsi pembelajaran atau fungsi optimasi guna mendapatkan kesalahan minimum dengan memperbaiki bobot, (6) Hidden Layer, adalah lapisan tersembunyi, bersifat opsional tergantung pada arsitektur Jaringan Syaraf Tiruan.

ANN atau JST dapat diimplementasikan dalam Regression untuk mengestimasi hubungan antara variabel independen dengan variabel dependen untuk mendapatkan fungsi yang sesuai dengan data. Kemudian, Classification untuk memprediksi kelas dari data input sesuai dengan target atau data label. Ketiga adalah Clustering untuk mengelompokkan sekumpulan data tanpa label data atau target. Terakhir, Forecasting untuk memprediksi kondisi masa depan menggunakan data masa kini dan masa lalu atau dikenal sebagai data deret waktu.

Fungsi aktivasi dalam JST menerjemahkan input menjadi keluaran. Bobot yang ditentukan secara acak sebagai nilai awal mempengaruhi sinyal dari satu neuron ke neuron lainnya [19].

Bobot merupakan parameter krusial dalam proses pembelajaran karena menentukan keberhasilan pengenalan pola. Untuk memperoleh bobot yang efektif, dilakukan perhitungan pada data input dan bobot. Perhitungan ini melibatkan kombinasi linear antara input dan bobot, diikuti dengan penerapan fungsi aktivasi (Wulandhari, 2017).

Beberapa jenis fungsi aktivasi pada ANN adalah Linear Function, Step Function, Logistic Sigmoid Function (Logsig), Hyperbolic tangent sigmoid Function (tanh/tansig), Rectified Linear Unit Function (ReLU), dan Gaussian Function.

D. SVR

Mean Absolute Error (MAE) adalah metrik yang sering digunakan dalam mengevaluasi keakuratan model statistik. Mean Absolute Error (MAE) mengukur rata-rata perbedaan

absolut antara nilai yang diprediksi dan nilai aktual. Mean Absolute Error (MAE) dinyatakan oleh Yusa et al., [27] sebagai ukuran keakuratan prediksi; semakin kecil nilai MAE, semakin kuat algoritma tersebut. Dalam sebuah pembelajaran machine learning, tahap pengukuran tingkat keakuratan kinerja sebuah metode sangat penting untuk dilakukan dengan tujuan untuk mengukur kinerja suatu algoritma. Dalam bukunya ((Wulandhari, 2017). Rumus MAE sebagai berikut :

$$MAE(y, \bar{y}) = \frac{1}{n \text{ Sample}} \quad (1)$$

$$\sum_{i=0}^{n \text{ Sample}} |y_i - \bar{y}_i| \quad (2)$$

Dimana :

n: adalah ukuran sampel

y_i : adalah nilai data actual ke-i

\bar{y}_i adalah nilai data peramalan ke- i.

Hasil perhitungan pada MAE selalu bernilai positif karena terdapat tanda mutlak (II). $|y_i - \bar{y}_i|$, artinya menunjukkan nilai kesalahan atau perbedaan nilai atau selisih nilai antara nilai aktual (nilai sebenarnya) dengan nilai prediksi (nilai peramalan). MAE yang ideal dan diharapkan dalam setiap penelitian adalah MAE dengan nilai kesalahan yang terkecil. Nilai MAE ini umumnya digunakan untuk membandingkan dua model dalam melakukan peramalan.

IV. HASIL DAN PEMBAHASAN

Langkah awal yang dilakukan dalam penelitian ini, setelah proses pengumpulan data selesai dilakukan, adalah menggabungkan data file raport dalam bentuk Excel dari tiap angkatan. Total seluruh file yang diolah sebanyak 36 file, lalu seluruh file tersebut digabungkan menjadi 1 file, yang dirubah dari format Excel (.xlsx) menjadi format CSV (Comma Separated Value) agar dapat diolah kedalam program Python. Hasil heatmap correlation yang didapat dari python dipindahkan ke dalam bentuk Excel. Kemudian dilakukan pengelompokkan mata pelajaran sesuai dengan tingkatan korelasinya. Tabel 1 menunjukkan rentang nilai hubungan korelasi dan warna yang mewakilinya.

Tabel 1. Rentang Nilai Hubungan Korelasi

RENTANG NILAI	KORELASI
> 0,75 - 0,99	Sangat Kuat
> 0,5 - 0,75	Kuat
> 0,25 - 0,5	Cukup Kuat
> 0 - 0,25	Sangat Lemah
0	Tidak Ada Korelasi
< 0	< -0,5
< -0,5	> -0,5
-0,5 s/d -1	Nilai Itu Sendiri

Setelah proses pengklasifikasian dengan kode warna selesai, maka langkah selanjutnya adalah dengan melihat dan mengamati secara manual. Jika di dalam 1 (satu) deret horizontal terlihat ada lebih dari 2 (dua) angka yang berkorelasi tinggi,

maka mata pelajaran tersebut berkorelasi tinggi terhadap bulan ketersediaan. Pada Tabel 2 dan 3 merupakan tampilan proses pengklasifikasian korelasi data nilai raport semester 1 dan 2 angkatan tahun 2014-2015.

Berdasarkan hasil tersebut maka didapatkan insight pertama, yaitu pada semester 1 dan 2 terdapat tiga fitur mata pelajaran yang berkorelasi sangat kuat dan kuat. Sementara pada semester 3 dan 4 terdapat empat fitur mata pelajaran yang berkorelasi sangat kuat dan kuat. Terakhir, pada semester 5 dan 6 terdapat dua fitur mata pelajaran yang berkorelasi sangat kuat dan kuat. Dengan demikian, perlu diperhatikan agar guru yang mengajar di semester 3 dan 4 harus lebih memperhatikan siswanya, terutama dalam hal ketuntasan belajar dan prestasi akademiknya, karena semester ini sangat penting.

Tabel 2. Rentang Nilai Hubungan Korelasi

SEMESTER 12 2014-2015													
		MUATAN NASIONAL											
NO	MAPEL	1	2	3	4	5	6	7	8	9	10	11	12
1	AVPAGP12	1,00	0,18	0,30	0,32	0,33	0,27	0,32	0,16	0,28	0,16	0,38	0,30
2	AVPAGK12	0,18	1,00	-0,49	-0,41	-0,53	0,25	0,05	0,54	-0,11	-0,30	-0,04	0,08
3	AVPKNP12	0,30	-0,49	1,00	0,71	0,76	-0,07	0,34	-0,35	0,34	0,44	0,24	0,08
4	AVPKNK12	0,32	-0,41	0,71	1,00	0,68	0,12	0,29	-0,28	0,29	0,49	0,37	0,24
5	AVBIDP12	0,33	-0,53	0,76	0,68	1,00	0,11	0,50	-0,44	0,29	0,51	0,28	0,08
6	AVBIDK12	0,27	0,25	-0,07	0,12	0,11	1,00	0,14	0,31	0,07	0,21	0,14	0,21
7	AVMATP12	0,32	0,05	0,34	0,29	0,50	0,14	1,00	0,08	0,20	0,18	0,33	0,21
8	AVMATK12	0,16	0,54	-0,35	-0,28	-0,44	0,31	0,08	1,00	0,19	-0,20	0,19	0,30
9	AVSJR12	0,28	-0,11	0,34	0,29	0,29	0,07	0,20	0,19	1,00	0,19	0,48	0,31
10	AVSJRK12	0,16	-0,30	0,44	0,49	0,51	0,21	0,18	-0,20	0,19	1,00	0,15	0,10
11	AVBIGP12	0,35	0,02	0,24	0,37	0,25	0,14	0,35	0,19	0,45	0,15	1,00	0,88
12	AVBIGK12	0,30	0,08	0,08	0,24	0,08	0,21	0,21	0,30	0,31	0,10	0,88	1,00

Tabel 3. Rentang Nilai Hubungan Korelasi

TABEL FITUR MATA PELAJARAN YANG BERKORELASI SANGAT KUAT & KUAT			
NO	FITUR	MATA PELAJARAN	ASPEK PENGETAHUAN/KETERAMPILAN
SEMESTER 1 & 2			
1	AVPAGK12	Pendidikan Agama	(Keterampilan / Praktek)
2	AVSENK12	Seni Budaya	(Keterampilan / Praktek)
3	AVPKWP12	Pendidikan Kewirausahaan	(Pengetahuan / Teori)
SEMESTER 3 & 4			
1	AVBIGK34	Bahasa Inggris	(Keterampilan / Praktek)
2	AVPROP34	Pemrograman Dasar	(Pengetahuan / Teori)
3	AVPROK34	Pemrograman Dasar	(Keterampilan / Praktek)
4	AVSOJK34	Sistem Operasi Jaringan	(Keterampilan / Praktek)
SEMESTER 5 & 6			
1	AVTSJK56	Troubleshooting Jaringan	(Keterampilan / Praktek)
2	AVKPTK56	Kerja Praktek TKJ	(Keterampilan / Praktek)

Hasil penelitian juga memberikan insight kedua bahwa terdapat empat fitur mata pelajaran normatif adaptif (pelajaran umum) yang berkorelasi sangat kuat dan kuat. Selain itu, juga terdapat lima fitur mata pelajaran produktif (pelajaran kejuruan) yang berkorelasi sangat kuat dan kuat. Berdasarkan hasil penelitian tersebut dapat dikatakan bahwa mata pelajaran

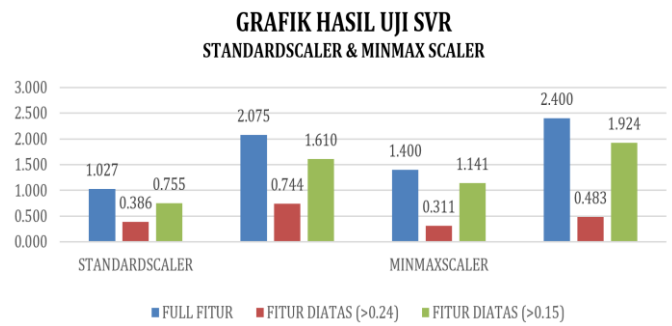
produktif (kejuruan) merupakan mata pelajaran yang lebih penting dibandingkan mata pelajaran umum untuk siswa dapat terserap di dunia industri. Hal tersebut untuk mendukung tingkat keprofesionalitas siswa saat terjun ke dunia industri (Febrianti et al., 2023).

Insight ketiga, berdasarkan hasil penelitian yaitu terdapat dua fitur mata pelajaran dalam aspek pengetahuan atau teori yang berkorelasi sangat kuat dan kuat. Kemudian, terdapat tujuh fitur mata pelajaran dalam aspek keterampilan / praktek yang berkorelasi sangat kuat dan kuat. Berdasarkan hasil tersebut maka mata pelajaran kejuruan harus lebih diperbanyak lagi porsi jam belajarnya. Hal tersebut dilakukan agar siswa lebih banyak melakukan praktik kejuruan. Dengan lebih mahir di bidang kejuruan, siswa tidak hanya dapat meningkatkan kualitas dan nilai raportnya, siswa juga pada akhirnya dapat meningkatkan angka keterserapan kerja di industri.

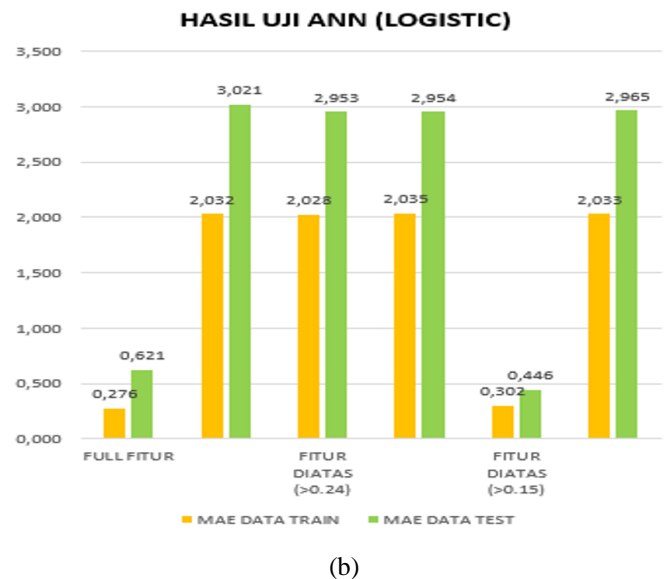
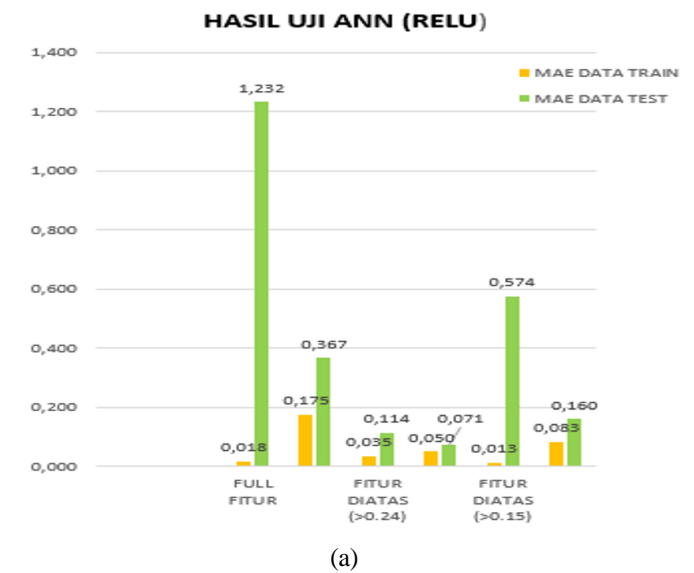
Jika pada penjelasan sebelumnya telah dijelaskan cara mencari korelasi antar mata pelajaran menggunakan heatmap correlation pada peranti lunak Python, lalu memindahkannya ke dalam bentuk Excel. Kemudian dilakukan proses pemilihan fitur dengan cara manual untuk dapat diambil insight datanya. Langkah berikutnya sebelum melakukan percobaan pada model adalah proses pemilihan fitur (feature selection) dengan menggunakan Python. Fitur yang digunakan adalah seluruh fitur (full feature) sebanyak 60 fitur, fitur > 0,24 sebanyak 10 fitur dan fitur > 0,15 sebanyak 33 fitur. Setelah semua data dan fitur yang dipilih telah siap, maka langkah berikutnya adalah melakukan percobaan pada model SVR dan ANN.

Model SVR dan ANN digunakan untuk dapat memprediksi di bulan ke-berapa siswa lulusan bisa terserap di dunia kerja, dengan output berbentuk angka. Kedua model ini diuji performanya dengan menggunakan Bahasa pemrograman Python, kemudian dicatat hasilnya, lalu dibandingkan. Hasil mana yang lebih bagus antara SVR dan ANN. Manakah yang menghasilkan hasil pengujian evaluasi data latih (data train) paling kecil, maka itulah model terbaik yang akan dipilih. Percobaan pada model SVR membandingkan normalisasi StandardScaler dan MinMaxScaler untuk mencari hasil terbaik. Pada fitur > 0,24 dengan 10 fitur, Coefisien 1,0, dan Epsilon 0,01 dengan MinMaxScaler menunjukkan kinerja terbaik: data train 0,311 dan data test 0,483 dibandingkan dengan Standard Scaler.

Hasil uji ANN yang divisualisasikan pada Gambar 1, menunjukkan bahwa algoritma aktivasi Relu dengan fitur di atas 0,15 (>0,15) sebanyak 33 fitur diolah menggunakan normalisasi Standard Scaler dengan parameter jumlah neurons 256 dan iter max 100 menghasilkan performa terbaik yaitu pada MAE data train sebesar 0,013 dan pada data test sebesar 0,574. Sedangkan pada algoritma aktifasi Logistic dengan seluruh fitur (full feature) sebanyak 60 fitur diolah menggunakan normalisasi StandardScaler, dengan parameter jumlah neurons 26 dan iter max 200 menghasilkan performa terbaik yaitu pada MAE data train sebesar 0,276 dan pada data test sebesar 0,621.



Gambar 1. Grafik Hasil Uji SVR



Gambar 2. Grafik perbandingan Hasil Uji ANN. (a) Relu & (b) Logistic

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dipaparkan di atas memberikan simpulan bahwa dengan pendekatan Educational Data mining dapat diketahui jenis mata pelajaran, yang diajarkan di semester berapa saja, serta aspek pengetahuan atau keterampilan yang paling berpengaruh terhadap angka keterserapan kerja siswa. Hal tersebut dapat dijadikan sebagai landasan pengetahuan bagi manajemen kurikulum di SMK Negeri 26 Jakarta. Mata pelajaran yang memiliki korelasi yang kuat dan sangat kuat terdapat pada mata pelajaran kejuruan (produktif). Mata pelajaran ini siswa SMK dapatkan sejak semester 3 sampai dengan semester 6, di mana aspek keterampilan (praktik) turut berpengaruh besar terhadap keterserapan kerja. Mata pelajaran produktif (kejuruan), paling banyak berpengaruh terhadap keterserapan kerja siswa. Berdasarkan hasil penelitian, didapat sebesar 0,5 bulan sebagai angka keterserapan siswa dalam hitungan bulan setelah kelulusan. Siswa yang belum terserap pada dunia industri di kisaran 2 bulan masih dianggap wajar. Hasil penelitian ini menunjukkan bahwa masa awal kelas X di semester 1 dan 2 adalah masa yang paling fundamental (pembentukan dasar), maka sekolah diharapkan bisa lebih memberikan pembinaan dan perhatian khusus yang lebih intens kepada siswa kelas X di jurusan Sistem Informatika Jaringan dan Aplikasi (SIJA), dalam hal mata pelajaran normatif adaptif dan mata pelajaran produktif. Sekolah juga perlu untuk lebih menekankan dan mengembangkan aspek mata pelajaran praktek (aspek keterampilan) sejak semester 1 sampai semester 6, agar keterserapan kerja dapat terus meningkat. SMK Negeri 26 Jakarta dan untuk SMK di seluruh Indonesia, diharapkan dapat melakukan penelitian dan pengembangan educational data mining dengan menggunakan fitur dan variabel-variabel yang lain agar bisa lebih bermanfaat untuk dapat mengambil insight data dari kumpulan data di sekolah yang sangat banyak dan belum pernah teruji dan tergalai manfaatnya selama ini.

DAFTAR PUSTAKA

- [1] A. A. Dudhe and S. R. Sakhare, "Teacher Ranking System To Rank of Teacher As Per Specific Domain.," *ICTACT J. Soft Comput.*, vol. 6956, no. January, pp. 1589–1596, 2018, doi: 10.21917/ijsc.2018.0222.
- [2] E. Y. Obsie and S. A. Adem, "Prediction of Student Academic Performance using Neural Network, Linear Regression and Support Vector Regression: A Case Study," *Int. J. Comput. Appl.*, vol. 180, no. 40, pp. 39–47, 2018, doi: 10.5120/ijca2018917057.
- [3] Y. K. Salal, S. M. Abdullaev, and Y. M. Kumar, "Educational Data Mining: Study Performance Prediction in Academic," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 4C, pp. 54–59, 2019.
- [4] P. Thakar, A. Mehta, and Manisha, "A Unified Model of Clustering and Classification to Improve Students' Employability Prediction," *I.J. Intell. Syst. Appl.*, vol. 9, no. September, pp. 10–18, 2017, doi: 10.5815/ijisa.2017.09.02.
- [5] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining: Second Edition*. 2014.
- [6] S. Susanto and D. Suryadi, *Pengantar Data Mining*. 2010.
- [7] E. I. A. Warih and Y. Rahayu, "Penerapan Data Mining untuk Menentukan Estimasi Produktivitas Tanaman Tebu dengan Menggunakan Algoritma Linier Regresi Berganda di Kabupaten Rembang," pp. 1–5, 2014.
- [8] P. Akulwar, S. Pardeshi, and A. Kamble, "Survey on Different Data Mining Techniques for Prediction," in *Proceeding of the 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2018, pp. 513–519, doi: 10.1109/I-SMAC.2018.8653734.
- [9] J. Hao and T. K. Ho, "Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language," *J. Educ. Behav. Stat.*, vol. XX, no. X, pp. 1–14, 2019, doi: 10.3102/1076998619832248.
- [10] K. R. Srinath, "Python -The Fastest Growing Programming Language," *Int. Res. J. Eng. Technol.*, vol. 4, no. 12, pp. 354–357, 2017.
- [11] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 11, pp. 1–19, 2022, doi: 10.1186/s40561-022-00192-z.
- [12] F. Qiu et al., "Predicting students' performance in e-learning using learning process and behaviour data," *Sci. Rep.*, no. 0123456789, pp. 1–15, 2022, doi: 10.1038/s41598-021-03867-8.
- [13] S. Amjad, M. Younas, M. Anwar, Q. Shaheen, M. Shiraz, and A. Gani, "Data Mining Techniques to Analyze the Impact of Social Media on Academic Performance of High School Students," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/9299115.
- [14] K. Nahar, B. I. Shova, T. Ria, H. B. Rashid, and A. H. S. Islam, "Mining educational data to predict students' A comparative study of data mining techniques," *Educ. Inf. Technol.*, 2021, doi: <https://doi.org/10.1007/s10639-021-10575-3> Mining.
- [15] H. Turabieh et al., "Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance," *Computing*, 2021, doi: 10.1007/s00607-020-00894-7.
- [16] A. Rivas, A. González-briones, G. Hernández, J. Prieto, and P. Chamoso, "Neurocomputing Artificial neural network analysis of the academic performance of students in virtual learning environments," *Neurocomputing*, no. xxx, 2020, doi: 10.1016/j.neucom.2020.02.125.
- [17] L. Kemper, G. Vorhoff, and B. U. Wigger, "Predicting student dropout: A machine learning approach Predicting student dropout: A machine learning approach," *Eur. J. High. Educ.*, vol. 0, no. 0, pp. 1–20, 2020, doi: 10.1080/21568235.2020.1718520.
- [18] H. Waheed, S. U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Comput. Human Behav.*, vol. 104, 2020, doi: 10.1016/j.chb.2019.106189.
- [19] V. Vijayalakshmi and K. Venkatchalapathy, "Deep Neural Network for Multi-Class Prediction of Student Performance in Educational Data," *Int. J. Recent Technol. Eng.*, vol. 3878, no. 2, pp. 5073–5081, 2019, doi: 10.35940/ijrte.B2155.078219.
- [20] S. Ranjeeth, T. P. Latchoumi, M. Sivaram, A. Jayanthiladevi, and T. S. Kumar, "Predicting Student Performance with ANNQ3H: A Case Study in Secondary Education," *Proc. 2019 Int. Conf. Comput. Intell. Knowl. Econ. ICCIKE 2019*, pp. 603–607, 2019, doi: 10.1109/ICCIKE47802.2019.9004387.
- [21] L. M. Abu Zohair, "Prediction of Student's performance by modelling small dataset size," *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, 2019, doi: 10.1186/s41239-019-0160-3.
- [22] C. Burgos, M. L. Campanario, D. de la Peña, J. A. Lara, D. Lizcano, and M. A. Martínez, "Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout," *Comput. Electr. Eng.*, vol. 66, pp. 541–556, 2018, doi: 10.1016/j.compeleceng.2017.03.005.
- [23] A. Mueen, B. Zafar, and U. Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques," *I.J. Mod. Educ. Comput. Sci.*, no. November, pp. 36–42, 2016, doi: 10.5815/ijmecs.2016.11.05.
- [24] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. Van Erven, "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil," *J. Bus. Res.*, vol. 94, no. February, pp. 335–343, 2019, doi: 10.1016/j.jbusres.2018.02.012.
- [25] F. Zhang and L. J. O. Donnell, "Support vector regression," in *Machine Learning*, Elsevier Inc., 2020, pp. 123–140.
- [26] F. Zhang and L. J. O'Donnell, *Support vector regression*. Elsevier Inc., 2019.
- [27] M. Yusa, E. Utami, and E. T. Luthfi, "Analisis Komparatif Evaluasi Performa Algoritma Klasifikasi pada Readmisi Pasien Diabetes," *J. Buana Inform.*, vol. 7, no. 4, pp. 293–302, 2016, doi: 10.24002/jbi.v7i4.770.