

Model Prediksi Prestasi Mahasiswa Berdasarkan Evaluasi Pembelajaran Menggunakan Pendekatan Data Science

Tommy^{1*}, Amir Mahmud Husein²

¹Universitas Harapan Medan, Fakultas Teknik dan Komputer, Teknik Informatika, Indonesia

²Universitas Prima Indonesia, Fakultas Teknologi dan Ilmu Komputer, Teknik Informatika, Indonesia

¹omshirakawa@gmail.com, ²amirmahmud@unprimdn.ac.id

Received: xx xx xxxx

Accepted: 9 May 2019

Published: 14 May 2019



*Tommy

Keywords: Data Science, Model Prediksi, Prestasi Mahasiswa, Prediksi Prestasi Mahasiswa, Evaluasi Pembelajaran

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Perguruan tinggi merupakan satuan penyelenggara pendidikan tinggi sebagai tingkat lanjut jenjang pendidikan menengah di jalur pendidikan formal. Aspek prestasi belajar merupakan salah satu aspek penilaian keberhasilan perguruan tinggi dalam proses belajar. Dalam makalah ini menyajikan hasil analisis hubungan antara pembelajaran dengan prestasi mahasiswa dimana tahapan yang dilakukan menggunakan pendekatan data science. Berdasarkan Analisis data terdapat tiga indikator penting dalam penilaian prestasi belajar yaitu pedagogi, profesional dan kepribadian. Ketiga fitur digunakan sebagai variabel dependen untuk memprediksi prestasi belajar dimana algoritma *DecisionTree* menghasilkan akurasi lebih baik dari pada model *k-nearest neighbors (KNN)*, *Logistic Regression*, *Support Vector Machine*, *Naive Bayes* dan dengan tingkat akurasi 68%, kemudian KNN dengan akurasi 66% dan lainnya sebesar 55% pada masing-masing algoritma yang diusulkan.

PENDAHULUAN

Perguruan tinggi merupakan satuan penyelenggara pendidikan tinggi sebagai tingkat lanjut jenjang pendidikan menengah di jalur pendidikan formal. Salah satu aspek yang berpengaruh ialah mahasiswa dan lulusan (Putri, Anggreani, and Wibawa 2020). Kualitas lulusan sangat berpengaruh terhadap proses pembelajaran dimana interaksi antara mahasiswa dan dosen sangat diperlukan sehingga mahasiswa dapat lebih mudah memahami capaian pembelajaran (Bernacki, Chavez, and Uesbeck 2020). Persepsi pembelajaran merupakan kegiatan evaluasi pembelajaran yang di laksanakan pada tiap akhir semester dari tiap mahasiswa terhadap perkuliahan yang di tempuh ketika mahasiswa telah menuntaskan pembelajaran matakuliah tersebut (Nuñez and Padilla 2020). Teknik yang di lakukan dalam mengumpulkan evaluasi pembelajaran adalah dengan meminta mahasiswa mengisi kuisisioner online sebagai syarat untuk mahasiswa dapat melihat nilai dari matakuliah tersebut. Dari hasil kuisisioner tersebut maka ditemukan hasil persepsi dari tiap mahasiswa terhadap matakuliah yang ditempuh beserta performance dari dosen yang mengajar, apakah mereka memahami atau tidak memahami. Hasil dari evaluasi pembelajaran ini sebenarnya sebagai masukan bagi dosen dalam meningkatkan teknik atau metode pembelajaran.

Prestasi belajar merupakan nilai capaian kegiatan pembelajaran dari mahasiswa yang diberikan dosen matakuliah yang bersangkutan berdasarkan kriteria penilaian yang telah ditetapkan sebelumnya (Nur et al. 2020). Data prestasi belajar mahasiswa merupakan data akademik yang rutin dilaporkan ke PDDIKTI. Kedua data tersebut sangat banyak ditampung dalam pangkalan data sebuah perguruan tinggi, dimana pemanfaatannya masih bersifat statistical murni untuk mengetahui tingkat atau sebaran suatu hal. Dari kedua data inilah kemudian akan digali mengenai keterhubungan apakah yang mungkin dapat mengkaitkan atau menjelaskan fenomena yang terjadi. Keterhubungan apa yang mungkin dapat ditemukan atau fenomena apa

yang bisa di jelaskan dari data persepsi pembelajaran dengan data prestasi belajar dan kebutuhan prediksi prestasi mahasiswa berdasarkan data persepsi pembelajaran. Salah satu ukuran kinerja dari setiap sistem pendidikan adalah prestasi akademik siswa. (Lolo and Ricambi 2019) Prestasi akademik adalah kemampuan untuk membuktikan prestasi akademik dalam perolehan hasil yang direncanakan (Wibrowski, Matthews, and Kitsantas 2017). Namun, memiliki kecerdasan yang tinggi tidak menjamin prestasi akademik, dan individu perlu menyadari gaya belajar mereka (Fong et al. 2017). Gaya belajar adalah metode belajar yang diterapkan siswa dalam mencapai, menganalisis, dan menginternalisasi pengetahuan yang baru diperolehnya

Makalah ini menyajikan analisis hubungan antara pembelajaran dan prestasi mahasiswa dengan menggunakan pendekatan data science. Tujuan bisnis yang ingin dicapai dalam penelitian ini adalah memberikan pemahaman yang lebih mendalam terhadap interpretasi data evaluasi pembelajaran dan prestasi belajar serta membangun model prediksi prestasi mahasiswa berdasarkan evaluasi pembelajaran. Hasil analisis dan prediksi ini diharapkan dapat dijadikan masukan bagi evaluasi pembelajaran itu sendiri dalam menyusun strategi evaluasi pembelajaran di masa depan.

TINJAUAN LITERATUR

Sebagian pekerjaan yang berhubungan dengan prediksi prestasi belajar siswa fokus pada faktor-faktor pengaruh prestasi belajar dan umumnya mereka meneliti pada kasus pembelajaran jarak jauh. Pendekatan pembelajaran mesin merupakan salah satu model yang paling banyak diterapkan (Alyahyan and Düştegör 2020). sebagian peneliti menerapkan teknik penambangan data (Kabakchieva 2013) kemudian diterapkan pada pembelajaran mesin (Anuradha and Velmurugan 2015), diantaranya Murad, D et al (Murad et al. 2019) memprediksi faktor-faktor yang mempengaruhi hasil belajar siswa pembelajaran jarak jauh menggunakan dataset catatan nilai siswa tahun 2016 dan 2017 sebanyak 16000 data. Mereka menggunakan 4 indikator yaitu Forum discussion (FOD), Personal assignment (PAS), TAS and Quiz (QIZ) dengan pendekatan Naïve Bayes pada aplikasi Rapid Minner dimana akurasi model ini mencapai 84.03%. Selanjutnya, Fan, K; et al (Fan et al. 2013) mengusulkan metode motivasi, kognitif, dan PLS (Partial Least Square) prediksi prestasi pembelajaran bermakna. Model yang diusulkan difokuskan pada model pengajaran teknologi informasi dan komunikasi, pembelajaran bermakna. Pada penelitian Nabizadeh S, et al (Nabizadeh et al. 2019) memprediksi prestasi akademik berdasarkan strategi motivasi belajar dan harapan hasil berdasarkan model teoritis yang fokus pada mahasiswa kesehatan.

Penerapan data mining untuk menambang data nilai siswa diusulkan oleh Peling, I et al (Peling et al. 2017) kemudian menerapkan algoritma Naïve Bayes untuk memprediksi periode belajar siswa. Makalah ini lebih menyoroti kegagalan mahasiswa adalah mahasiswa yang lulus tidak tepat waktu. Mahasiswa yang lulus tidak tepat waktu adalah mahasiswa yang tidak dapat menyelesaikan studinya sesuai dengan ketentuan waktu yang diberikan. Naïve Bayes mampu mengklasifikasikan data testing yang benar rata-rata sebesar 86,16% dan error 13,84%. Kemudian (YILDIZ and BÖREKÇİ 2020) mengembangkan metodologi untuk membangun wawasan wawasan tentang data pendidikan yang dikumpulkan dari siswa kelas sembilan dengan menggunakan metode data mining. Hasil pengumpulan data berisi informasi demografis tentang siswa dan keluarganya, rutinitas belajar, perilaku menghadiri kegiatan pembelajaran, dan keyakinan epistemologis mereka tentang sains. Terakhir mereka melakukan klasifikasi dengan pembelajaran mesin berbeda dimana Neural Network algoritma (98,6%) memiliki skor tertinggi. Tingkat akurasi algoritma lainnya adalah kNN (86,2%), Regresi Logistik (78,4%), SVM (90,3%), Decision Tree (91,9%), Random Forest (90,0%), dan Naive Bayes (81,7%).

BAHAN DAN METODE

Bahan

Penelitian ini menggunakan pendekatan data science dimana pemahanan bisnis merupakan salah bagian penting dalam membangun model prediksi. Proses pengumpulan data bersumber dari hasil catatan nilai siswa di salah satu perguruan tinggi yang ada di Kota Medan. Pada tabel 1 merupakan sebagian dataset yang

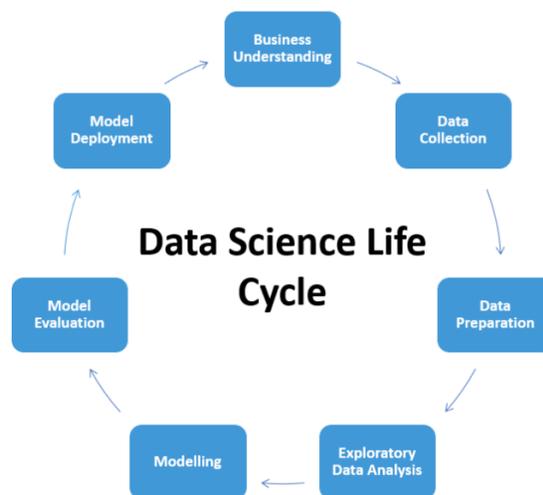
digunakan. Dataset yang digunakan dalam membangun model pada project ini adalah dataset penilaian persepsi dosen pengampu dan nilai mahasiswa yang diampu dari masing – masing dosen.

<i>id_nilai</i>	<i>id_matkul</i>	<i>nilai</i>	<i>id_dosen1</i>	<i>id_dosen2</i>	<i>id_prodi</i>	<i>pedagogik</i>	<i>profesional</i>	<i>kepribadian</i>
1	22	51	10	1	1	3.15	3.26	3.26
2	22	51	10	1	1	3.15	3.26	3.26
3	22	51	10	1	1	3.15	3.26	3.26
4	22	51	10	1	1	3.15	3.26	3.26
5	22	73	10	1	1	3.15	3.26	3.26
...
1753	16	75	9	11	1	3.88	3.87	4.09
1754	16	75	9	11	1	3.88	3.87	4.09
1755	16	0	9	11	1	3.88	3.87	4.09
1756	5	83	6	3	1	4.04	3.98	4.17
1757	6	83	6	3	1	3.93	3.92	4.16

Kumpulan data ini berisi informasi tentang id makul, nilai, id dosen1, id dosen 2, id_prodi, pedagogi, profesional dan kepribadian. Dataset ini terdiri dari 1.575 baris dan 9 kolom.

Metodologi yang diusulkan.

Metodolgi yang digunakan untuk membangun model prediksi terdiri dari Identifikasi Permasalahan Bisnis, Data Collection, Data Preparation, Exploratory Data Analysis, Modelling, Evaluation dan Deployment. Lebih jelaskan dapat dilihat pada gambar 1.



Gambar 1 Metodologi yang diusulkan

1. Identifikasi Permasalahan Bisnis

Keterhubungan apa yang mungkin dapat di temukan atau fenomena apa yang bisa di jelaskan dari data persepsi pembelajaran dengan data prestasi belajar dan kebutuhan prediksi prestasi mahasiswa berdasarkan data persepsi pembelajaran.

2. Data Collection

Setelah pemahaman bisnis, langkah selanjutnya adalah Pemahaman data. Langkah ini melibatkan pengumpulan semua data yang tersedia.

3. Data Preparation

Setelah langkah Pemahaman Data, langkah selanjutnya adalah Persiapan Data. Langkah ini juga dikenal sebagai Pembersihan Data atau Data Wrangling. Ini mencakup langkah-langkah seperti memilih data yang relevan, mengintegrasikan data dengan menggabungkan set data, membersihkannya, menangani nilai yang hilang dengan menghapusnya atau memasukkannya dengan data yang relevan, memperlakukan data yang salah dengan menghapusnya, juga memeriksa pencilan dan menanganinya. Membuat data baru, mendapatkan fitur baru dari yang sudah ada dengan menggunakan rekayasa fitur. Format data ke dalam struktur yang diinginkan, hapus kolom dan fitur yang tidak diinginkan. Persiapan data adalah yang paling memakan waktu karena memakan waktu hingga 70% -90% dari waktu dalam pembangunan model.

4. Exploratory Data Analysis

Analisis Data Eksplorasi (EDA) memainkan peran penting pada tahap ini karena ringkasan data bersih membantu dalam mengidentifikasi struktur, pencilan, anomali, dan pola yang ada dalam data. Wawasan ini dapat membantu menemukan kumpulan fitur yang tepat, algoritme yang akan digunakan untuk pembuatan model, dan membangun model.

5. Modelling

Pemodelan data dianggap sebagai jantung dari analisis data. Model mengambil data yang telah disiapkan dari langkah sebelumnya (Persiapan Data) sebagai masukan dan memberikan keluaran yang diinginkan. Langkah ini termasuk memilih jenis model yang sesuai, apakah masalahnya adalah masalah klasifikasi, atau masalah regresi atau masalah pengelompokan. Setelah memilih model, di antara berbagai algoritma yang ada. Pada penelitian kami mengusulkan lima model yaitu *Logistic Regression*, *Support Vector Machine*, *DecisionTree*, *Naive Bayes* dan *k-nearest neighbors (KNN)*

6. Evaluation

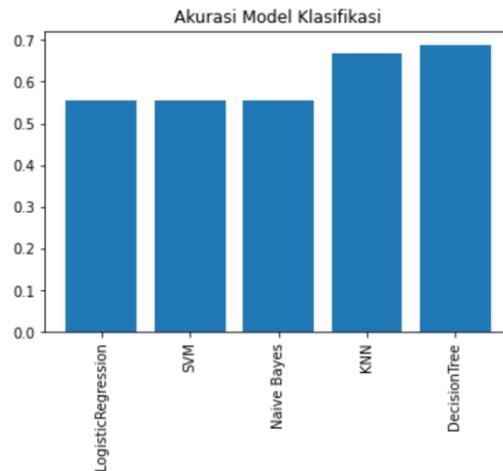
Setelah pembangunan model, tahapan selanjutnya adalah mengevaluasi model dengan mengukur keakuratan (Seberapa baik kinerja model yaitu apakah ia mendeskripsikan data secara akurat) dan relevansi (Apakah itu menjawab pertanyaan asli yang ditetapkan untuk dijawab). Dalam hal ini perlu memastikan ada keseimbangan yang benar antara performa dan generalisasi, yang berarti model yang dibuat tidak boleh bias dan harus menjadi model yang digeneralisasi.

7. Deployment

Model setelah evaluasi yang ketat akhirnya diterapkan dalam format dan saluran yang diinginkan. Ini adalah langkah terakhir dalam siklus hidup data science yang diusulkan.

HASIL PENELITIAN

Hasil pengujian yang dilakukan pada penelitian ini terdiri dari beberapa hasil temuan dari berbagai metode yaitu *Logistic Regression*, *Support Vector Machine*, *DecisionTree*, *Naive Bayes* dan *k-nearest neighbors (KNN)*. Semua metode yang diusulkan akan dievaluasi berdasarkan tingkat akurasi, namun tahapan rekayasa fitur sangat penting dalam kinerja model pembelajaran mesin. Berdasarkan pengamatan hasil rekayasa fitur, kami menetapkan fitur pedagogic, profesional dan kepribadian memberikan distribusi sebaran yang lebih baik terhadap fitur nilai mahasiswa. Setelah menentukan fitur penting, maka tahapan selanjutnya adalah pelabelan dataset dengan menggunakan empat kelompok nilai yang di encode secara numerik mulai dari nilai 1 sampai 4 pada masing-masing label Buruk, Sedang, Baik dan Sangat Baik. Akhirnya dilakukan pengujian pada 5 model yang diusulkan dan hasilnya dapat dilihat pada gambar 2.



Gambar 2 Hasil Akurasi Model Klasifikasi

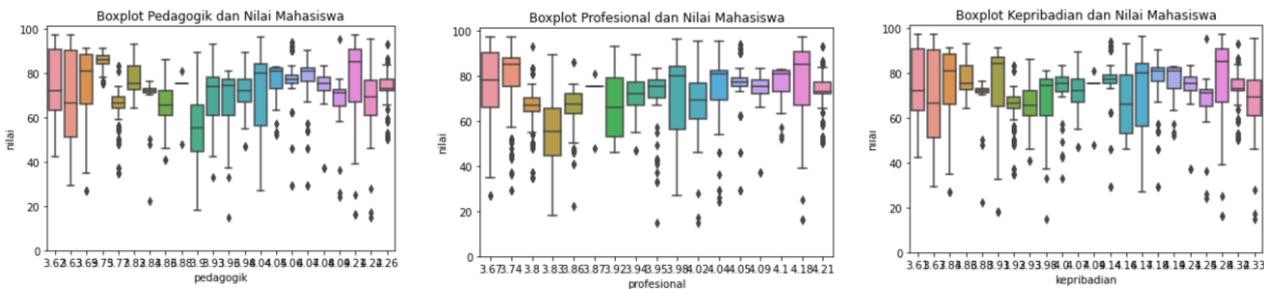
Pada gambar 2 dapat dilihat model DecisionTree lebih unggul dibanding model lainnya, kemudian KNN sedang ketiga model menghasilkan akurasi rata-rata 50%. Model DecisionTree dan KNN menghasilkan akurasi 68,81% dan 66,67 % sehingga model DecisionTree akan ditetapkan dalam proses deployment selanjutnya.

Tabel 2 Perbandingan Akurasi Model Klasifikasi

Classifier	Accuracy
LogisticRegression	0.5549
SVM	0.5549
DecisionTree	0.6881
Naive Bayes	0.5549
KNN	0.6677

Diskusi

Secara hipotesa, nilai evaluasi pembelajaran dosen pengampu akan memiliki korelasi yang kuat terhadap prestasi mahasiswa, dimana tingginya nilai evaluasi pembelajaran akan menghasilkan prestasi mahasiswa yang baik. Namun, berdasarkan analisa dataset yang telah dilakukan, tidak selamanya nilai evaluasi pembelajaran memiliki korelasi yang kuat terhadap prestasi mahasiswa. Hal ini dapat dilihat dari sebaran regresi linear dari nilai mahasiswa terhadap evaluasi pembelajaran yang tidak positif. Diagram boxplot juga menunjukkan banyaknya outlier nilai mahasiswa terhadap nilai evaluasi pembelajaran.



Gambar 2 Diagram boxplot Seleksi Fitur

Oleh karena itu, kami membangun lima model untuk prediksi prestasi mahasiswa berdasarkan evaluasi pembelajaran dosen pengampu. Dengan menggunakan model pembelajaran mesin, akurasi prediksi dapat mencapai nilai 1.0 yang dapat dilihat pada confusion matrix pengujian walaupun korelasi antara evaluasi pembelajaran dengan prestasi mahasiswa tidak selalu kuat. Berdasarkan hasil evaluasi dari kelima model yang

diusulkan DecisionTree menghasilkan akurasi sebesar 68 % lebih baik dari model lain sehingga model ini telah di-deploy dan dapat beroperasi dengan baik sehingga dapat digunakan untuk kebutuhan prediksi. Untuk pengembangan yang akan datang, dibutuhkan analisa yang lebih mendalam terhadap variabel – variabel pendukung lain yang masih memiliki keterkaitan dengan evaluasi pembelajaran dosen pengampu terhadap prediksi prestasi mahasiswa.

KESIMPULAN

Penelitian ini fokus pada analisi hubungan antara pembelajaran terhadap prestasi mahasiswa. Dari hasil analisis nilai evaluasi pembelajaran dosen pengampu akan memiliki korelasi yang kuat terhadap prestasi mahasiswa, dimana tingginya nilai evaluasi pembelajaran akan menghasilkan prestasi mahasiswa yang baik, namun kondisi ini sangat berpengaruh pada kurikulum dan aturan penilaian akademik karena setiap perguruan tinggi memiliki kriteria berbeda dalam penilaian keberhasilan dosen sehingga sangat penting dilakukan penelitian di masa dengan dengan kondisi dan struktur data yang berbeda.

Supplementary Materials (optional)

Tidak tersedia secara terbuka dan sangat diharapkan apabila

Kontribusi Peneliti

Tommy: Konseptualisasi,; metodologi,; perangkat lunak, validasi; analisis formal, visualisasi,; pengawasan; AMH: penyidikan, sumber daya, akurasi data, tulisan—persiapan draf asli; menulis—ulasan dan penyuntingan. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Konflik kepentingan

Para penulis menyatakan tidak ada konflik kepentingan

REFERENCES

- Alyahyan, Eyman, and Dilek Düşteğör. 2020. "Predicting Academic Success in Higher Education: Literature Review and Best Practices." *International Journal of Educational Technology in Higher Education* 17(1):1–21.
- Anuradha, C., and T. Velmurugan. 2015. "A Comparative Analysis on the Evaluation of Classification Algorithms in the Prediction of Students Performance." *Indian Journal of Science and Technology* 8(15). doi: 10.17485/ijst/2015/v8i15/74555.
- Bernacki, Matthew L., Michelle M. Chavez, and P. Merlin Uesbeck. 2020. "Predicting Achievement and Providing Support before STEM Majors Begin to Fail." *Computers and Education* 158(702):103999. doi: 10.1016/j.compedu.2020.103999.
- Fan, Kuo Kuang, Chung Ho Su, Shuh Yeuan Deng, and Wei Jhung Wang. 2013. "An Achievement Prediction Model of Meaningful Learning, Motivation, and Cognitive on SPANI: Partial Least Square Analysis." *Mathematical Problems in Engineering* 2013. doi: 10.1155/2013/961963.
- Fong, Carlton J., Coreen W. Davis, Yughi Kim, Young Won Kim, Lauren Marriott, and Soo Yeon Kim. 2017. "Psychosocial Factors and Community College Student Success: A Meta-Analytic Investigation." *Review of Educational Research* 87(2):388–424. doi: 10.3102/0034654316653479.
- Kabakchieva, Dorina. 2013. "Predicting Student Performance by Using Data Mining Methods for Classification." *Cybernetics and Information Technologies* 13(1):61–72. doi: 10.2478/cait-2013-0006.
- Lolo, Suhartono, and Stopira Ricambi. 2019. "An Analysis of Bayesian Algorithm in Graduates To Study Further Into University." 871–79.
- Murad, Dina Fitria, Bambang Dwi Wijanarko, Erick Fernando, Willy Johan Widjaja Saputra, Taufik Darwis, and Lena. 2019. "Prediction Learning Achievement Indicators in Distance Learning Students." *TALE 2019 - 2019*

- IEEE International Conference on Engineering, Technology and Education* 1–6. doi: 10.1109/TALE48000.2019.9225897.
- Nabizadeh, Sakineh, Sepideh Hajian, Zohre Sheikhan, and Fatemeh Rafiei. 2019. "Prediction of Academic Achievement Based on Learning Strategies and Outcome Expectations among Medical Students." *BMC Medical Education* 19(1):1–11. doi: 10.1186/s12909-019-1527-9.
- Nuñez, Sara, and Jesus Mendoza Padilla. 2020. "Industry 4.0: An Integrated Distance Learning Solution." *2020 6th International Conference on Science in Information Technology: Embracing Industry 4.0: Towards Innovation in Disaster Management, ICSITech 2020* 29–34. doi: 10.1109/ICSITech49800.2020.9392072.
- Nur, Amri Muliawan, Muhammad Farid Wazdi, Bambang Harianto, and Muhammad Farid Zaini. 2020. "Implementation of Naive Bayes Algorithm in Analyzing Acceptance of Poor Student Assistance." *Journal of Physics: Conference Series* 1539(1). doi: 10.1088/1742-6596/1539/1/012018.
- Peling, Ida Bagus Adisimakrisna, I. Nyoman Arnawan, I. Putu Arich Arthawan, and I. Gusti Ngurah Janardana. 2017. "Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm." *International Journal of Engineering and Emerging Technology* 2(1):53. doi: 10.24843/ijeet.2017.v02.i01.p11.
- Putri, Desy Pratiwi Ika, Desi Anggreani, and Aji prasetya Wibawa. 2020. "Prediksi Lama Studi Menggunakan Naïve Bayes." *Mobile and Forensics* 2(1):41–50.
- Wibrowski, Connie R., Wendy K. Matthews, and Anastasia Kitsantas. 2017. "The Role of a Skills Learning Support Program on First-Generation College Students' Self-Regulation, Motivation, and Academic Achievement: A Longitudinal Study." *Journal of College Student Retention: Research, Theory and Practice* 19(3):317–32. doi: 10.1177/1521025116629152.
- YILDIZ, Muhammed, and Caner BÖREKÇİ. 2020. "Predicting Academic Achievement with Machine Learning Algorithms." *Journal of Educational Technology and Online Learning* 3(3). doi: 10.31681/jetol.773206.