

Prediksi Lama Studi dan Predikat Kelulusan Mahasiswa Menggunakan Algoritma *Supervised Learning*

Nur Baiti Nasution^{1✉}, Dwi Hartanto², Dicky Januarizky Silitonga³, Lasimin⁴, Dewi Mardhiyana⁵

^{1,5} Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Pekalongan, Indonesia

² Indonesian Artificial Intelligence Society, Indonesia

³ Institut Teknologi Sumatera, Indonesia

⁴ Universitas Nahdlatul Ulama Al Ghazali, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 30-12-2022

Direvisi : 25-01-2023

Diterima : 31-01-2023

Kata Kunci:

Machine Learning,
Educational Data Mining,
Klasifikasi, Masa Studi,
Predikat Kelulusan.

Keywords :

Machine Learning,
Educational Data Mining,
Classification, Graduation,
Prediction

Corresponding Author :

Nur Baiti Nasution

Pendidikan Matematika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Pekalongan, Indonesia

Jalan Sriwijaya No 3 Pekalongan

Email: nurbaiti.nasution@unikal.ac.id

ABSTRAK

Untuk menjaga kualitas alumni di suatu universitas, perlu dilakukan prediksi mengenai masa studi dan IPK akhir mahasiswa. Prediksi masa studi dan predikat dapat digunakan sebagai dasar pemberian perlakuan agar mahasiswa dapat lulus tepat waktu dengan nilai yang baik. Lebih lanjut, prediksi tersebut juga dapat digunakan sebagai dasar perbaikan pengelolaan universitas. Penelitian ini bertujuan untuk mengembangkan model klasifikasi menggunakan algoritma *supervised learning* untuk memprediksi masa studi dan predikat IPK mahasiswa ketika lulus kuliah. Data yang digunakan adalah data alumni Universitas Pekalongan tahun 2018 yang terdiri dari 1208 alumni. Pemodelan dilakukan dengan menggunakan metodologi CRISP DM dengan menerapkan 3 model klasifikasi yaitu KNN, *Decision Tree*, dan *Support Vector Machine*. Metrik keberhasilan yang digunakan adalah akurasi. Dari hasil pemodelan, diperoleh bahwa model yang terbaik untuk variabel masa studi adalah SVM dengan nilai akurasi 0.70 dan model terbaik untuk variabel predikat adalah KNN dengan nilai akurasi 0.51.

ABSTRACT

To maintain the quality of alumni at a university, it is necessary to predict the study period and final GPA of students. Prediction of study period and predicate can be used as a basis for providing treatment so that students can graduate on time with good grades. Furthermore, the prediction can also be used as a basis for improving university management. This research aims to develop a classification model using supervised learning algorithm to predict the study period and GPA predicate of students when they graduate. The data used is Pekalongan University alumni data in 2018 consisting of 1208 alumni. Modeling is done using CRISP DM methodology by applying 3 classification models namely KNN, Decision Tree, and Support Vector Machine. The success metric used is accuracy. From the modeling results, it was found that the best model for the study period variable was SVM with an accuracy value of 0.70 and the best model for the predicate variable was KNN with an accuracy value of 0.51.

PENDAHULUAN

Terdapat berbagai variabel yang dapat digunakan sebagai ukuran keberhasilan studi mahasiswa di perguruan tinggi. Salah satunya adalah masa studi lulusan yaitu waktu yang diperlukan mahasiswa untuk menempuh studi mulai dari semester 1 hingga lulus menjadi alumni. Alumni S1 dikatakan lulus tepat waktu jika masa studinya kurang dari atau sama dengan 4 tahun. Sedangkan alumni D3 dikatakan lulus tepat waktu jika masa studinya kurang dari atau sama dengan 3 tahun.

Selain masa studi, indikator lain yang sering digunakan adalah IPK lulusan. IPK atau Indek Prestasi Kumulatif adalah jumlah seluruh nilai akhir yang diperoleh oleh alumni untuk keseluruhan mata kuliah dibagi dengan jumlah SKS yang telah ditempuh selama studi. IPK biasanya dinyatakan dalam skala 4 dengan nilai minimal 0 dan maksimal 4.00. Besaran IPK biasanya dikategorikan lagi menjadi 3 yaitu Memuaskan, Sangat Memuaskan, dan Dengan Pujian (*cumlaude*). Predikat ini dapat berbeda di tiap universitas. Nilai IPK yang termasuk ke dalam masing-masing kategori pun dapat berbeda sesuai kebijakan masing-masing universitas dan program studi.

Masa studi dan IPK mahasiswa juga menjadi indikator kinerja program studi pada Instrumen Akreditasi Program Studi (IAPS 4.0) yang dikeluarkan oleh BAN PT (2019). Dalam instrumen tersebut dinyatakan bahwa prodi yang dapat meluluskan tepat waktu dengan rata-rata IPK lebih dari 3.00 dapat memperoleh nilai akreditasi yang baik. Oleh sebab itu, perguruan tinggi berusaha agar masa studi dan IPK lulusannya sesuai dengan kriteria yang diperlukan untuk keperluan akreditasi. Idealnya, perguruan tinggi berusaha agar mahasiswa dapat lulus tepat waktu dengan IPK yang baik.

Akan tetapi, proses meluluskan mahasiswa tidak selalu mudah dan lancar. Banyak kasus menunjukkan bahwa mahasiswa sering terkendala pada penyusunan tugas akhir sehingga masa studi menjadi lama (Malfasari dkk., 2019; Modouw & Nugroho, 2021; Pratiwi & Roosyanti, 2019; Zain dkk., 2021). Beberapa studi menyatakan bahwa kendala yang mungkin ditemui adalah adanya kesulitan komunikasi dengan dosen pembimbing (Kocimaheni dkk., 2020; Malfasari dkk., 2019; Pratiwi & Roosyanti, 2019) dan kurangnya pemahaman mengenai substansi tugas akhir (Daniel & Taneo, 2019). Kendala-kendala tersebut sebenarnya dapat diminimalisir jika sejak awal prodi dapat memprediksi kelancaran studi mahasiswa yang kemudian data tersebut digunakan sebagai dasar memberikan perlakuan agar mahasiswa dapat lulus tepat waktu.

Untuk itu, perlu diselidiki variabel apa saja yang berpengaruh terhadap lama studi dan IPK mahasiswa selama belajar di universitas. Setelah diperoleh variabel-variabel tersebut, perlu dikembangkan suatu model yang dapat memprediksi bilamana seorang mahasiswa akan mencapai ketepatan waktu selama studi dan apakah mahasiswa tersebut akan mendapatkan nilai IPK sesuai dengan target atau tidak. Model yang diperoleh dapat digunakan sebagai dasar pengambilan keputusan mengenai pengelolaan universitas atau program studi seperti pengadaan program matrikulasi, pemberian tambahan SKS untuk mata kuliah tertentu, dan sebagainya.

Pada artikel ini akan dikembangkan suatu model *supervised learning* yang dapat digunakan untuk memprediksi masa studi dan IPK mahasiswa berdasarkan variabel-variabel masukan. Selama ini, telah banyak dikembangkan aplikasi pemodelan yang dapat digunakan untuk memprediksi keberhasilan studi mahasiswa berdasarkan variabel seperti masa studi (Azahari dkk., 2020; Fadrial, 2021; Saputro & Sari, 2020; Windarti & Suradi, 2019) dan kinerja (Azahari dkk., 2020). Walaupun demikian, sebagian dari studi tersebut masih menggunakan algoritma sistem pakar yang berbasis aturan dan belum banyak yang menggunakan algoritma *machine learning*. Dengan demikian, masih terdapat ruang yang cukup besar untuk mengimplementasikan *machine learning* khususnya yaitu *supervised learning* dalam menyelesaikan masalah prediksi masa studi dan IPK mahasiswa.

Machine learning adalah bagian dari *artificial intelligence* yang memfokuskan pada pembuatan model agar komputer dapat menyelesaikan suatu masalah regresi, klasifikasi, atau clustering. Pengembangan model *machine learning* dapat berupa pemodelan untuk *supervised*

learning, unsupervised learning, dan reinforcement learning (Roihan dkk., 2020). *Supervised learning* adalah pemodelan *machine learning* jika nilai variabel terikat telah diketahui sebelumnya. Beberapa algoritma supervised learning yang sering digunakan adalah KNN, *Decision Tree*, Naïve Bayes, dan *Support Vector Machine*. Adapun tujuan dari penelitian ini adalah untuk mengembangkan beberapa model *machine learning* kemudian membandingkan nilai akurasinya untuk dapat memperoleh model yang paling baik untuk memprediksi masa studi dan predikat IPK mahasiswa.

METODE PENELITIAN

Penelitian ini menggunakan metodologi standar yang biasa diegunakan pada penelitian data mining yaitu metode Crisp DM, yaitu *business understanding, data understanding, data preparation, modeling, evaluation, dan deployment* (Schröer dkk., 2021). Data yang digunakan pada penelitian ini adalah data alumni Universitas Pekalongan tahun 2018 yaitu sebanyak 1208 alumni. Data yang digunakan merupakan data primer yang diperoleh dari Bagian Akademik Universitas Pekalongan. Pada data mentah, terdapat 24 atribut (variabel) yang mana masih perlu dilakukan pemilihan variabel mana saja yang dapat digunakan sebagai variabel bebas untuk digunakan sebagai prediktor.

Pada tahapan *business understanding*, dilakukan kajian mengenai atribut data yang mana yang dapat digunakan untuk memprediksi masa studi dan IPK mahasiswa. Kajian dilakukan berdasarkan ketersediaan dan kecukupan data. Hasil pemodelan pada tahap akhir nanti dapat menjadi dasar dalam perencanaan atau pengambilan keputusan adalah dalam hal penerimaan mahasiswa baru, pengembangan metode pembelajaran yang diselaraskan dengan profil latar belakang pendidikan mahasiswa sebelumnya, dan lain sebagainya.

Pada tahap *data understanding*, dilakukan pengumpulan data dan proses data understanding atau lebih dikenal dengan *exploratory data analysis*. Pada tahapan ini dilakukan visualisasi untuk setiap atribut yang digunakan dan juga analisis statistik deskriptif untuk atribut yang bertipe numerik. Pada tahap *data preparation*, dilakukan transformasi masing-masing atribut sehingga sesuai dengan input model yang diinginkan. Selain itu, juga dilakukan *handling* terhadap data-data yang hilang (*missing*), invalid, atau *outlier*.

Selanjutnya, tahapan modelling dimana akan dibangun model untuk data yang telah diproses pada bagian sebelumnya. Terdapat 3 model klasifikasi *machine learning* yang akan dicoba yaitu K-Nearest Neighbour Classifier, Decision Tree Classifier, dan Support Vector Machine. Untuk setiap model tersebut, apabila memungkinkan, juga akan dilakukan parameter tuning. Proses pemodelan dilakukan dengan menggunakan 70% dari keseluruhan data sebagai data *training* dan 30% sisanya akan digunakan menjadi data *testing*. Pada tahap *evaluation*, ketiga model tersebut dibandingkan dengan menghitung metrik keberhasilan yang sesuai. Pengembangan model *machine learning* pada penelitian ini menggunakan bahasa pemrograman Python 3.

Model yang dikembangkan pada penelitian ini memiliki perbedaan dengan model yang sudah pernah dikembangkan sebelumnya khususnya pada proses pemilihan variabel yang dilibatkan. Seperti yang telah disebutkan sebelumnya bahwa data pada penelitian ini merupakan data wisuda yang tidak sepenuhnya “rapi” dan sesuai dengan kebutuhan model. Oleh sebab itu, banyak waktu penelitian digunakan untuk proses *cleaning data* sehingga menjadi sesuai kebutuhan pemodelan.

HASIL DAN PEMBAHASAN

Data yang diperoleh dari Bagian Akademik Universitas Pekalongan memiliki 24 atribut yaitu No Urut, No. Seri Ijasah, No. Seri Transkrip, NIK, NPM, Nama Lengkap di Buku Ijasah, Nama Ijasah, Tempat Lahir, Tanggal Lahir dalam format tanggal, Tanggal Lahir dalam format teks, TTL, Jenis Kelamin, Tahun Masuk, Status (mahasiswa baru atau pindahan), No Telp/HP, Judul Skripsi, Tanggal Lulus dalam format tanggal, Tanggal Lulus dalam format teks, Total SKS, Nilai Kumulatif, IPK, Asal Sekolah, Pekerjaan Orang Tua, Instansi Orang Tua, dan Predikat. Data terdiri dari 1208 *record* alumni. Karena belum berbasis sistem informasi dan pengumpulan

data hanya berasal dari Google Form, maka banyak data yang bersifat opsional dan tidak diisi oleh alumni. Data yang diperoleh pun masih berupa dua file excel (1 file untuk setiap periode wisuda) yang masing-masing berisi 14 sheet (setiap sheet berisi data alumni untuk setiap prodi). Dari ke-24 atribut tersebut, akan ditetapkan variabel yang akan digunakan sebagai prediktor pada tahapan *business understanding*.

Business Understanding

Selanjutnya dilakukan pengkajian atas masalah prediksi yang akan diselesaikan sekaligus dipilih variabel yang akan digunakan sebagai variabel prediktor. Model yang dikembangkan diharapkan dapat memprediksi masa studi mahasiswa. Awalnya ditetapkan bahwa model seharusnya dapat memprediksi lama studi mahasiswa dalam satuan bulan (model regresi karena variabel terikat merupakan data interval). Akan tetapi, mengingat banyak mahasiswa lulus mendekati masa wisuda maka banyak mahasiswa yang memiliki masa studi yang sama. Dengan kata lain, keragaman data yang ada tidaklah terlalu besar. Jika data seperti ini dipakai untuk membuat model regresi maka dikhawatirkan model tidak akan terlalu informatif. Akibatnya, hasil prediksi yang diharapkan cukup hanya berupa apakah mahasiswa tersebut akan lulus tepat waktu atau tidak dengan mempertimbangkan jenjang pendidikannya (S1 atau D3). Dengan kata lain, model yang dikembangkan untuk memprediksi masa studi merupakan model klasifikasi.

Mengingat pada data mentah tidak ada variabel masa studi maka perlu dimunculkan variabel masa studi (dengan satuan bulan) yang dapat diperoleh dari :

$$\text{masa studi} = \text{tanggal lulus} - \text{tanggal masuk}$$

Selanjutnya, dibuat lagi variabel baru yaitu variabel Kategori Masa Studi (selanjutnya disebut dengan masa studi) yang terdiri dari 2 nilai yaitu Lulus Tepat Waktu dan Tidak Lulus Tepat Waktu sesuai dengan aturan yang ditunjukkan pada Tabel 1.

Tabel 1. Kategori Masa Studi

Jenjang	Masa Studi (X)	Kategori
S1	$X > 48$ bulan	Tidak Lulus Tepat Waktu
	$X \leq 48$ bulan	Lulus Tepat Waktu
D3	$X > 36$ bulan	Tidak Lulus Tepat Waktu
	$X \leq 36$ bulan	Lulus Tepat Waktu

Agar memudahkan ketika memprediksi IPK mahasiswa, ditetapkan juga bahwa model yang dikembangkan untuk memprediksi nilai IPK juga model klasifikasi. Untuk itu, perlu dibuat klasifikasi nilai IPK mahasiswa. Klasifikasi yang dimaksud dilakukan sesuai dengan Peraturan Akademik Universitas Pekalongan yang ditunjukkan pada Tabel 1. Dengan ketentuan tersebut, selanjutnya, tidak akan digunakan istilah IPK, melainkan diganti dengan variabel Predikat Kelulusan (selanjutnya cukup disebut predikat). Perlu diketahui bahwa predikat terendah adalah Memuaskan yang diperoleh jika nilai IPK berkisar antara 2.75 hingga 3.00. Pada data yang digunakan, tidak ditemui alumni dengan IPK yang kurang dari 2.75.

Tabel 2. Kategori IPK Mahasiswa

Rentang IPK	Kolom 2
$2.75 \leq IPK \leq 3.00$	Memuaskan
$3.00 < IPK \leq 3.50$	Sangat Memuaskan
$3.50 < IPK \leq 4.00$	Dengan Pujian

Selanjutnya, ditentukan variabel bebas atau prediktor yang akan digunakan. Uraian sebelumnya menyatakan bahwa hasil prediksi akan dipergunakan untuk memberikan perlakuan kepada mahasiswa agar dapat mencapai predikat Sangat Memuaskan (minimal) dan dapat lulus tepat waktu. Perlakuan tersebut dapat berupa pembenahan kurikulum, pemberian matrikulasi,

atau perubahan model pembelajaran. Dengan demikian, idealnya, prediksi dilakukan dengan menggunakan atribut-atribut yang sejak awal telah dimiliki mahasiswa sejak masuk kuliah yaitu asal sekolah menengah (SMA/SMK) dan juga jenis kelamin.

Universitas Pekalongan merupakan universitas swasta yang mana mekanisme penerimaan mahasiswa baru nya belum terlalu ketat. Dari sudut pandang kualitas input mahasiswa baru, latar belakang pendidikan jenjang sekolah menengah atas ini dapat mempengaruhi performa akademik di kampus. Sebagai contoh, mahasiswa yang berasal dari latar belakang pendidikan SMA akan dapat dengan mudah beradaptasi pada prodi-prodi akademik. Hal tersebut dikarenakan materi pada prodi akademik lebih banyak diisi dengan materi teoretis dasar yang sudah dipelajari di SMA. Hal berbeda akan berlaku pada para mahasiswa yang berlatar belakang SMK. Mahasiswa yang berasal dari SMK mungkin saja akan kesulitan menempuh prodi akademik dan lebih mudah menempuh pembelajaran di prodi vokasi atau prodi yang lebih menekankan pada praktek. Kemampuan adaptasi mahasiswa dalam menerima materi perkuliahan ini tentunya berpengaruh pada prestasi mahasiswa dan juga lama studinya.

Di samping asal sekolah menengahnya, jenis kelamin juga dinilai dapat mempengaruhi performa akademik mahasiswa. Keterampilan dan minat jenis kelamin tertentu memiliki kecenderungan tertentu. Misalnya, minat terhadap ilmu keteknikan akan berbeda antara perempuan dan laki-laki, begitu juga dalam beberapa kategori keilmuan yang lain. Namun demikian, ada juga bidang-bidang yang tidak terlalu terpengaruh dengan jenis kelamin sehingga performa mahasiswa perempuan dan laki-laki bisa dikatakan sama. Selain asal sekolah dan jenis kelamin, atribut lain yang akan digunakan sebagai prediktor adalah fakultas. Pemilihan atribut ini didasarkan pada fakta bahwa setiap fakultas kemungkinan memiliki target dan cara pembelajaran yang berbeda satu sama lain sehingga dapat mempengaruhi lama studi dan juga pencapaian IPK mahasiswa. Rangkuman hasil yang diperoleh pada tahapan business understanding, berupa variabel yang akan digunakan dalam pemodelan, peran variabel tersebut (sebagai variabel bebas atau terikat), dan jenis variabelnya ditunjukkan pada Tabel 3.

Tabel 3. Variabel yang Terlibat Pada Pemodelan

Nama Variabel	Berperan Sebagai	Jenis Variabel
Masa Studi	Variabel Terikat	Nominal
Predikat	Variabel Terikat	Kategorikal
Fakultas	Variabel Bebas (Prediktor)	Kategorikal
Jenis Kelamin	Variabel Bebas (Prediktor)	Nominal
Asal Sekolah	Variabel Bebas (Prediktor)	Nominal

Data Understanding dan Data Preparation

Pada tahapan ini dilakukan analisis mengenai data yang akan digunakan dalam pemodelan meliputi pengecekan data yang tidak seimbang, *outlier*, dan *missing value*. Ditemukan beberapa kejanggalan pada beberapa variabel. Daftar kejanggalan tersebut dan cara penanganan yang perlu dilakukan untuk mengatasinya ditunjukkan pada Tabel 4.

Tabel 4. Hasil Data Understanding

Variabel	Permasalahan	Cara Menangani
Jenis Kelamin	Terdapat 28 <i>record</i> data yang belum memiliki nilai untuk variabel ini Beberapa <i>record</i> tertulis “Laki-laki” dan <i>record</i> lain tertulis “Laki-Laki”	Menuliskan secara manual data yang hilang. Menyeragamkan dengan menggunakan fungsi Replace
Predikat	Terdapat 60 <i>record</i> yang tertulis “Cumlaude” dan <i>record</i> lain tertulis “Dengan Pujian”	Menyeragamkan dengan menggunakan fungsi Replace
Fakultas	Fakultas Teknik hanya memiliki 6 orang alumni. Jumlah ini relatif kecil jika dibandingkan dengan keseluruhan. Variabel asal sekolah pada data bernilai nama sekolah pada jenjang sebelum kuliah. Adapun yang dibutuhkan hanyalah jenis sekolahnya (SMA/SMK)	Alumni Fakultas Teknik tidak diikuti dalam pembuatan model Mengekstrak asal sekolah yang dibutuhkan (SMA/SMK) dengan bantuan fungsi manipulasi string
Asal Sekolah	Terdapat alumni yang asal sekolahnya adalah MA, padahal yang dibutuhkan adalah SMA atau SMK Alumni Prodi Profesi Ners memiliki asal sekolah bukan SMA/SMK melainkan STIKES	Mengubah nilai “MA” menjadi “SMA” Meningat masa studi profesi hanya 1 tahun dan jumlah alumninya hanya sedikit maka alumni prodi ini tidak diikuti dalam pemodelan

Gambar 1 menunjukkan cuplikan data yang telah dilakukan pembersihan seperti yang ditunjukkan pada Tabel 4. Gambar 2 menunjukkan banyak alumni untuk masing-masing kategori pada variabel predikat. Adapun Gambar 3 menunjukkan banyaknya alumni untuk masing-masing kategori pada kedua variabel masa studi. Hasil pembersihan memberikan sebanyak 1130 data alumni yang siap digunakan sebagai bahan untuk pemodelan.

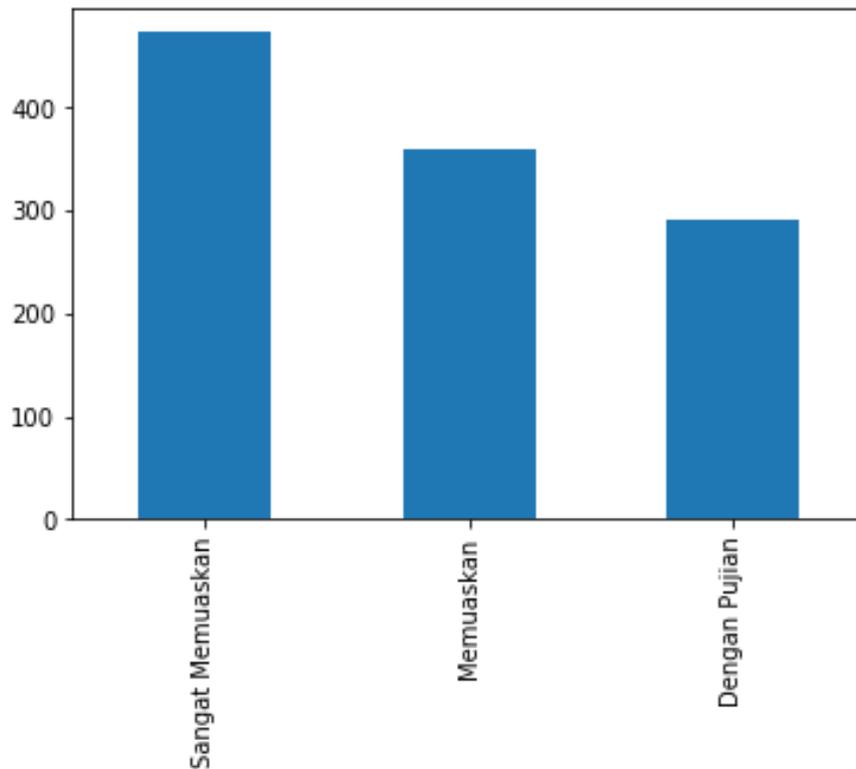
	gender	fakultas	jenjang	predikat	kategori_masa_studi
0	Perempuan	Ekonomi	SMA/MA	Dengan Pujian	Tidak Tepat Waktu
1	Perempuan	Ekonomi	SMK	Sangat Memuaskan	Tidak Tepat Waktu
2	Laki-Laki	Ekonomi	SMK	Memuaskan	Tidak Tepat Waktu
3	Perempuan	Ekonomi	SMK	Sangat Memuaskan	Tidak Tepat Waktu
4	Perempuan	Ekonomi	SMK	Sangat Memuaskan	Tidak Tepat Waktu

Gambar 1. Cuplikan Data Setelah Dibersihkan

Data Preparation

Selanjutnya data yang telah dibersihkan tersebut dipersiapkan untuk dapat memasuki tahap pemodelan. Atribut yang akan digunakan dalam pemodelan semuanya merupakan atribut kategorikal yang bernilai string. Sebelum memulai pemodelan, semua atribut tersebut diubah ke dalam bentuk numerik, atau di-encode. Untuk variabel bernilai binary atau nominal, *encoding* seharusnya tidak menunjukkan urutan atau dengan kata lain, perlu dilakukan *one hot encoding*. Akan tetapi, hal ini tidak dilakukan karena terdapat atribut fakultas yang memiliki 7 nilai yang berbeda. Jika dilakukan *one hot encoding* pada atribut fakultas maka akan muncul 7 variabel baru yang berfungsi sebagai vektor yang menyatakan asal fakultas suatu data. Jika ditambahkan dengan hasil *one hot encoding* atribut jenis kelamin dan asal sekolah maka akan muncul 11 atribut baru. Hal

ini dikhawatirkan dapat mengurangi informasi yang dapat terserap dari data dan kemungkinan untuk terjadinya *overfitting*. Oleh sebab itu, *encoding* yang dilakukan hanya sebatas mengubah ke bentuk numerik dengan menggunakan LabelEncoding.



Gambar 2. Visualisasi Variabel Predikat

Adapun untuk variabel terikat, yaitu masa studi dan predikat, keduanya dianggap variabel kategorikal yang bersifat ordinal (memiliki urutan) maka proses *encoding* pun dilakukan dengan memperhatikan hal tersebut. Pada variabel masa studi, nilai 0 menunjukkan kategori Tidak Tepat Waktu dan nilai 1 menunjukkan kategori Tepat Waktu. Sedangkan pada variabel predikat, nilai 0 menotasikan Memuaskan, nilai 1 menotasikan Sangat Memuaskan, dan 2 menotasikan Dengan Pujian. Cuplikan data setelah dilakukan proses encoding dapat dilihat pada Gambar 4.



Gambar 3. Visualisasi Variabel Masa Studi

Modelling

Algoritma model yang dapat digunakan untuk menyelesaikan tujuan bisnis pada makalah ini adalah semua algoritma *supervised learning* yang bertujuan untuk klasifikasi. Beberapa model yang akan dicoba adalah K-Nearest Neighbour Classifier, Decision Tree Classifier, dan Support Vector Machine. Pemilihan model ini didasarkan pada penerapan model pada kasus-kasus serupa dan mendapatkan hasil yang cukup memuaskan (Alsayed dkk., 2021; Windarti & Suradi, 2019). Dataset yang digunakan akan dibagi menjadi data training dan testing dengan perbandingan 70:30. Dengan demikian diperoleh sebanyak 786 data training dan 338 data testing. Adapun kriteria performansi yang akan digunakan adalah nilai akurasi. Hasil perolehan akurasi dari setiap model untuk masing-masing variabel terikat dapat dilihat pada Tabel 5.

	gender	fakultas	jenjang	predikat	kategori_masa_studi
0	1	0	0	2	0
1	1	0	1	1	0
2	0	0	1	0	0
3	1	0	1	1	0
4	1	0	1	1	0

Gambar 4. Cuplikan Data Setelah *Encoding*

Dari Tabel 5 dapat dilihat bahwa untuk variabel masa studi, nilai akurasi terbesar didapatkan ketika menerapkan model Support Vector Machine dengan tuning parameter. Hal ini sejalan dengan hasil pemodelan yang dilakukan oleh Kesumawati (2018). Hal ini disebabkan karena algoritma SVM memunculkan *hyperplane* yang menjadi pembatas antara 2 kategori. Disebutkan juga bahwa SVM merupakan model klasifikasi yang sangat sesuai untuk data yang hanya terdiri dari 2 kategori (Pang dkk., 2017). Pada pemodelan di artikel ini, tuning parameter SVM dilakukan dengan metode Grid Search dan memperoleh hasil terbaik pada parameter dengan rincian $C = 10$, $\gamma = 1$, dan kernel = rbf.

Tabel 5. Hasil Akurasi Setiap Model

Model Klasifikasi	Variabel Masa Studi		Variabel Predikat	
	Nilai Akurasi	Kesimpulan	Nilai Akurasi	Kesimpulan
KNN tanpa tuning	0.64	Tidak Dipakai	0.44	Tidak Dipakai
KNN dengan tuning	0.68	Tidak Dipakai	0.51	Dipakai
Decision Tree tanpa tuning	0.67	Tidak Dipakai	0.46	Tidak Dipakai
Decision Tree dengan tuning	0.67	Tidak Dipakai	0.46	Tidak Dipakai
SVM tanpa tuning	0.69	Tidak Dipakai	0.48	Tidak Dipakai
SVM dengan tuning	0.70	Dipakai	0.49	Tidak Dipakai

Untuk variable predikat, Tabel 5 menunjukkan bahwa nilai akurasi tertinggi hanya sebesar 0.51 yang diperoleh ketika model yang dipakai adalah model KNN dengan tuning. Hasil tuning parameter menunjukkan bahwa nilai n (*neighbour*) yang mengakibatkan nilai akurasi terbesar adalah 66. Dalam kasus ini, nilai akurasi model dengan KNN merupakan yang terbesar meskipun belum di taraf memuaskan. Hasil ini menunjukkan bahwa KNN merupakan model yang cukup dapat diandalkan pada permasalahan prediksi IPK mahasiswa seperti yang ditunjukkan oleh Wiyono (2019) dan Mohammadi (2019).

Dalam hal ini, model KNN dapat memberikan hasil yang paling baik karena seperti yang telah diketahui bahwa model KNN melakukan klasifikasi dengan cara menghitung jarak data terdekat dan melakukan iterasi yang berawal dengan $n = 1$. Model ini sangat efektif jika data yang

digunakan berukuran besar seperti data pada artikel ini. Nilai n yang diperoleh pada model dengan nilai akurasi tertinggi pun masih cukup kecil yaitu 66.

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini bertujuan untuk mengembangkan suatu model yang dapat memprediksi masa studi dan IPK mahasiswa di Universitas Pekalongan. Model yang dikembangkan dapat digunakan sebagai dasar pengambilan keputusan untuk pengelolaan universitas selanjutnya.

Dari hasil pemodelan untuk kedua variabel terikat, diperoleh model yang memiliki nilai akurasi paling tinggi untuk memprediksi variabel masa studi adalah model SVM dengan parameter $C = 10$, $\gamma = 1$, kernel = rbf, dan nilai akurasi sebesar 0.70. Sedangkan model yang memiliki nilai akurasi paling tinggi dalam memprediksi variabel predikat adalah model KNN dengan nilai $n = 66$ dan nilai akurasi sebesar 0.51.

Model ini selanjutnya dapat digunakan untuk perbaikan *management* universitas atau program studi agar dapat memetakan mahasiswa mana yang memerlukan perlakuan khusus untuk dapat lulus tepat waktu dan dapat memperoleh IPK sesuai target.

Saran

Penelitian ini masih memiliki keterbatasan karena model yang digunakan hanya 3 macam dan variabel predictor yang dipilih masih belum bervariasi. Untuk penelitian selanjutnya, dapat dipertimbangkan dalam menggunakan model klasifikasi yang lain dan menggunakan data yang memiliki atribut yang lebih beragam. Selain itu, algoritma *unsupervised learning* juga dapat digunakan untuk melakukan segmentasi alumni kemudian menganalisis karakteristik masing-masing kelompok alumni.

UCAPAN TERIMA KASIH

Pemodelan yang dilakukan pada artikel ini merupakan hasil Thematic Academy : Pelatihan Artificial Intelligence untuk Dosen dan Instruktur yang diselenggarakan oleh KOMINFO RI. Untuk itu, Kami mengucapkan terima kasih yang sebesar-besarnya atas ilmu dan pengalaman yang diberikan.

REFERENSI

- Alsayed, A. O., Rahim, M. S. M., Albidewi, I., Hussain, M., Jabeen, S. H., Alromema, N., Hussain, S., & Jibril, M. L. (2021). Selection of the Right Undergraduate Major by Students Using Supervised Learning Techniques. *Applied Sciences* 2021, Vol. 11, Page 10639, 11(22), 10639. <https://doi.org/10.3390/APP112210639>
- Azahari, A., Yulindawati, Y., Rosita, D., & Mallala, S. (2020). Komparasi Data Mining Naive Bayes dan Neural Network memprediksi Masa Studi Mahasiswa S1. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(3), 443. <https://doi.org/10.25126/jtiik.2020732093>
- Instrumen Akreditasi Program Studi*, 47 (2019) (testimony of BAN-PT).
- Daniel, F., & Taneo, P. N. L. (2019). Analisis Kesulitan Mahasiswa Dalam Penyusunan Proposal Penelitian Pendidikan Matematika. *JPMI (Jurnal Pendidikan Matematika Indonesia)*, 4(2), 79. <https://doi.org/10.26737/jpmi.v4i2.956>
- Fadrial, Y. E. (2021). Algoritma Naive Bayes Untuk Mencari Perkiraan Waktu Studi Mahasiswa. *INTECOMS: Journal of Information Technology and Computer Science*, 4(1), 20–29. <https://doi.org/10.31539/intecom.v4i1.2219>

- Kesumawati, A., & Utari, D. T. (2018). Predicting patterns of student graduation rates using Naive bayes classifier and support vector machine. *AIP Conference Proceedings, 2021*(October 2018). <https://doi.org/10.1063/1.5062769>
- Kocimaheni, A. A., Aminin, Z., & Kartika, A. D. (2020). Identifikasi Kesulitan Mahasiswa Fakultas Bahasa Dan Seni Universitas Negeri Surabaya Dalam Penyelesaian Studinya. *Paramasastra, 7*(2), 139. <https://doi.org/10.26740/paramasastra.v7n2.p139>
- Malfasari, E., Devita, Y., Erlin, F., & Filer, F. (2019). Faktor-Faktor Yang Mempengaruhi Kecemasan Mahasiswa Dalam Menyelesaikan Tugas Akhir Di Stikes Payung Negeri Pekanbaru. *Jurnal Ners Indonesia, 9*(1), 124. <https://doi.org/10.31258/jni.8.2.124-131>
- Modouw, H. B., & Nugroho, P. I. (2021). Faktor-Faktor Yang Mempengaruhi Penyelesaian Tugas Akhir Mahasiswa Program Studi Akuntansi Universitas Kristen Satya Wacana. *Jurnal Akuntansi Profesi, 12*(1), 59. <https://doi.org/10.23887/jap.v12i1.33283>
- Mohammadi, M., Dawodi, M., Tomohisa, W., & Ahmadi, N. (2019). Comparative study of supervised learning algorithms for student performance prediction. *1st International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2019, 124–127*. <https://doi.org/10.1109/ICAIIIC.2019.8669085>
- Pratiwi, D. E., & Roosyanti, A. (2019). Analisis faktor penghambat skripsi mahasiswa jurusan pendidikan guru sekolah dasar Universitas Wijaya Kusuma Surabaya. *Jurnal Pendidikan Dasar, 10*(1), 101–114. <http://journal.unj.ac.id/unj/index.php/jpd/article/view/11272>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology), 5*(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naive Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal, 6*(1), 1. <https://doi.org/10.24076/citec.2019v6i1.178>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science, 181*(2019), 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Windarti, M., & Suradi, A. (2019). Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining untuk Prediksi Masa Studi Mahasiswa. *Telematika, 12*(1), 14. <https://doi.org/10.35671/telematika.v12i1.778>
- Wiyono, S., Abidin, T., Wibowo, D. S., Hidayatullah, M. F., & Dairoh, D. (2019). Comparative Study of Machine Learning Knn, Svm, and Decision Tree Algorithm To Predict Student'S Performance. *International Journal of Research -GRANTHAALAYAH, 6*(1), 190–196. <https://doi.org/10.29121/granthaalayah.v7.i1.2019.1048>
- Zain, M. I., Radiusman, R., Syazali, M., Hasnawati, H., & Amrullah, L. W. Z. (2021). Identifikasi Kesulitan Mahasiswa Dalam Penyusunan Skripsi Prodi Pgsd Universitas Mataram. *Tunjuk Ajar: Jurnal Penelitian Ilmu Pendidikan, 4*(1), 73. <https://doi.org/10.31258/jta.v4i1.73-85>