

Prediksi Kinerja Calon Mahasiswa Berdasarkan Nilai Seleksi Masuk Menggunakan Pendekatan *Machine Learning*

Sisia Dika Ariyanto dan Lili Ayu Wulandhari

Universitas Bina Nusantara
Kampus Angrek, Jl. Kebon Jeruk Raya No. 27, Jakarta Barat, 11530
E-mail : sisia.ariyanto@binus.ac.id, lili.wulandhari@binus.ac.id

Abstrak

Politeknik Astra menerapkan tes seleksi masuk dimana salah satu tahapan utamanya adalah Tes Potensi Akademik (TPA). TPA terdiri dari tujuh subtes yang dianggap terlalu banyak dan membutuhkan waktu yang lama. Panitia seleksi berkeinginan untuk melakukan perubahan menjadi efektif dan efisien dengan mengurangi jumlah subtes. Untuk mengetahui subtes yang paling berpengaruh terhadap performa mahasiswa, digunakanlah *Machine Learning*. Penelitian ini bertujuan untuk mengetahui subtest paling berpengaruh terhadap performa mahasiswa yang didapatkan dengan membandingkan nilai seleksi masuk dengan indeks prestasi semester 1. Pendekatan dilakukan dengan algoritma klasifikasi dan regresi. Hasil dari klasifikasi, algoritma *Random Forest* memberikan hasil terbaik. Selanjutnya untuk melihat fitur yang paling berpengaruh terhadap kelulusan seleksi masuk, dilakukan seleksi fitur dengan metode filter dan *impurity based*. Tiga fitur terbaik diperoleh dari Prodi Manajemen Informatika dan non-Manajemen Informatika. Selanjutnya dilakukan regresi dengan dua algoritma, yaitu *Support Vector Regression (SVR)* dan *Neural Network Regression* dengan konfigurasi 3, 5, dan model 7 fitur. Hasil terbaik konsisten dengan dua model data yaitu algoritma SVR dengan *mean absolute error* untuk Prodi Manajemen Informatika 0.17 dan non-Manajemen Informatika 0.19. Hasilnya, model data dengan 3 fitur memiliki hasil terbaik untuk Prodi Manajemen Informatika, artinya TPA dapat disederhanakan dengan tiga fitur, sedangkan pada Prodi Non Manajemen Informatika, hasil terbaik pada tujuh fitur.

Kata kunci : Prediksi Performa Mahasiswa, Machine Learning, Klasifikasi, Feature Importance, Regresi.

Pendahuluan

Proses seleksi masuk mahasiswa ke perguruan tinggi adalah salah satu kegiatan rutin tahunan terbesar karena diikuti hampir seluruh lulusan Sekolah Menengah Atas (SMA)/Sekolah Menengah Kejuruan (SMK) se-Indonesia. Penentuan tahapan seleksi masuk juga dilakukan oleh Politeknik Astra. Tujuan utamanya adalah mencari calon mahasiswa berkarakter untuk menjadi lulusan siap kerja. Politeknik Astra memiliki enam Program Studi jenjang Diploma Tiga yaitu Pembuatan Peralatan Perkakas Produksi, Teknik Produksi dan Proses Manufaktur, Manajemen Informatika, Mesin Otomotif, Mekatronika, dan Teknologi Konstruksi Bangunan Gedung. Politeknik Astra saat ini memiliki kapasitas maksimal sebanyak 2000 mahasiswa, namun pada saat dilakukan penelitian ini kapasitas yang terisi baru sekitar 40%. Hal ini dikarenakan baru dilakukan peningkatan kapasitas

pada tahun 2021. Peningkatan kapasitas terjadi karena dibangunnya gedung baru untuk menampung kapasitas mahasiswa yang lebih banyak dari sebelumnya. Selain itu Politeknik Astra yang awalnya hanya meluluskan jenjang Diploma Tiga saat ini sedang melakukan pengembangan untuk Program Studi Diploma Empat/Sarjana Terapan. Politeknik Astra memiliki keunggulan dibandingkan dengan Institusi Pendidikan lainnya karena dapat terhubung langsung dengan 200 lebih anak perusahaan Astra Grup. Politeknik Astra membuktikan daya tariknya dengan adanya pendaftar program beasiswa dan program reguler mencapai 8000 pendaftar setiap tahunnya. Politeknik Astra juga menawarkan beasiswa Astra kepada mahasiswa berpotensi dan berprestasi yang berjumlah 20-30% dari kapasitas maksimal mahasiswa. Untuk mendapatkan beasiswa Astra setiap siswa harus mengikuti rangkaian seleksi tes.

Seleksi tes Politeknik Astra terdiri dari 4 bagian,

yaitu tes administrasi, tes potensi akademik, tes wawancara, dan tes kesehatan. Pada tes administrasi, siswa akan dilihat berdasarkan nilai akademik selama masa SMA/SMK, pendapatan orang tua dan juga jurusan pada masa SMA/SMK. Pada seleksi potensi akademik terdiri dari beberapa subtes yang harus dikerjakan setiap siswa. Jika lulus pada seleksi potensi akademik maka siswa akan melanjutkan untuk tes wawancara. Dalam tes wawancara terdapat dua nilai secara umum, yaitu nilai terkait kepribadian siswa dan nilai kesesuaian program studi yang dipilih dengan potensi siswa. Setelah dinyatakan lulus dalam tahap seleksi wawancara, peserta akan diminta untuk melakukan tes kesehatan. Tes kesehatan ini bertujuan untuk memastikan bahwa setiap mahasiswa siap secara fisik untuk mengikuti pola perkuliahan di Politeknik Astra. Pada seleksi program reguler, tidak diperlukan seleksi wawancara.

Rangkaian tes seleksi masuk Politeknik Astra melalui program reguler & beasiswa sudah diselenggarakan lebih dari 10 tahun. Namun, khusus pada tes seleksi potensi akademik sering mengalami perubahan susunan subtes. Perubahan tersebut dikarenakan panitia seleksi mengalami kesulitan dalam menentukan subtes mana saja yang efektif dan efisien untuk menyeleksi atau menyaring calon mahasiswa baru. Efektif dan efisien yang dimaksud adalah, dengan jumlah subtes seminimal mungkin dapat menyaring calon-calon mahasiswa terbaik sesuai kebutuhan setiap program studi. Total seluruh subtes sampai saat ini yang pernah digunakan dan sedang digunakan sebanyak tujuh subtes. Subtes pertama yaitu matematika dasar, yang kedua ada fisika untuk Prodi Teknik dan PAT untuk Prodi Manajemen Informatika, selanjutnya ada subtes Bahasa Inggris. Untuk ketiga subtes di atas termasuk ke dalam golongan tes akademik. Kemudian ada 4 subtes yang lainnya yang masuk ke dalam golongan psikotes yang memiliki tujuan untuk mengetahui karakter dari calon mahasiswa. Keempat subtes tersebut yaitu *spatial visualization*, *perceptual speed*, *assembly*, dan *symbolic reasoning*. Komposisi dari subtes tersebut sering diubah pada setiap tahunnya.

Pada tujuh tahun terakhir sudah dilakukan empat kali perubahan pada susunan subtes dengan adanya penambahan dan pengurangan subtes yang ada. Berdasarkan hasil perubahan tersebut, performa mahasiswa yang sudah diterima ternyata masih belum menunjukkan sesuai yang diinginkan. Panitia seleksi masih berkeinginan untuk melakukan perubahan, karena semakin banyak calon mahasiswa yang diterima maka semakin besar pula potensi pendaftar calon mahasiswa baru dan peserta tes potensi akademik akan meningkat. Sehingga dapat disimpulkan bahwa adanya kebutuhan untuk membuat tes potensi akademik menjadi efektif dan efisien. Salah satu cara memenuhi kebutuhan tersebut adalah mengetahui subtes yang paling berpengaruh terhadap kelulusan. Untuk menge-

tau subtes mana yang paling berpengaruh maka dikembangkan model klasifikasi. Selain itu, dibutuhkan pula prediksi kinerja dari hasil seleksi masuk. Tujuannya adalah untuk mengetahui subtes yang paling berpengaruh terhadap kinerja mahasiswa. Pada tahap ini akan dikembangkan dengan prediksi regresi. Sehingga akan diketahui subtes mana yang paling mempengaruhi kinerja mahasiswa.

Penggunaan pendekatan *Machine Learning* di dunia pendidikan sudah dilakukan pada penelitian-penelitian sebelumnya. Penggunaan *Machine Learning* salah satunya bertujuan untuk melakukan prediksi performa. Penggunaan algoritma beragam ada yang menggunakan klasifikasi ada pula regresi. Beberapa algoritma klasifikasi yang menunjukkan hasil dengan *accuracy* baik pada penelitian yang dilakukan oleh [1] dan [2] yaitu *Decision Tree*, *Random Forest*, *Support Vector Machine* dan *Neural Network*. Untuk regresi 2 algoritma yang cukup sering digunakan dengan nilai error kecil berdasarkan penelitian yang dilakukan oleh [3] yaitu *Support Vector Regression* dan *Neural Network Regression*. Prediksi performa yang dilakukan dengan klasifikasi dan regresi belum ada yang berfokus pada prediksi berdasarkan nilai seleksi masuk untuk melihat performa mahasiswa.

Beberapa penelitian terkait diantaranya adalah, *Machine Learning* berguna dalam memantau dan menganalisis proses pembelajaran di sekolah, memeriksa efisiensi dan efektivitas metode pembelajaran, memprediksi kinerja siswa dengan menawarkan bantuan akademik yang diperlukan, memberikan umpan balik yang berarti bagi guru dan siswa dan lain-lain [4]. Terdapat juga sebuah penelitian yang mengimplementasikan *Machine Learning* dalam dunia pendidikan yang menggunakan metode algoritma SVM dengan menggunakan Parameter *Grid*. Hasil akurasi yang diperoleh dari pengujian adalah 99,64% yang berarti metode SVM dengan menggunakan Parameter *Grid* lebih efisien dalam memprediksi performa siswa [5]. Sebuah penelitian lainnya juga bertujuan untuk menguji suatu algoritma klasifikasi sebagai klasifikasi hybrid yang menggunakan jaringan Radial Basis Function, Multilayer Perceptron, C4.5 dan Algoritma Random Forest untuk klasifikasinya. Hasil yang ditemukan pada penelitian ini adalah algoritma klasifikasi hybrid memberikan akurasi lebih dibandingkan algoritma klasifikasi individual [2]. Selain yang sudah disebutkan, terdapat juga penelitian lain yang menggunakan pendekatan *Machine Learning* dan fitur-fitur utama untuk memprediksi performa siswa. Sebanyak 39 studi terpilih yang kemudian dievaluasi merupakan studi yang dipublikasikan antara tahun 2015 hingga 2021. Hasilnya menunjukkan bahwa enam model *Machine Learning* yang paling banyak digunakan: *Artificial Neural Network (ANN)*, *Linear Regression (LinR)*, *Decision Tree (DT)*, *Support Vector Machine (SVM)*, *K-Nearest Neighbor*

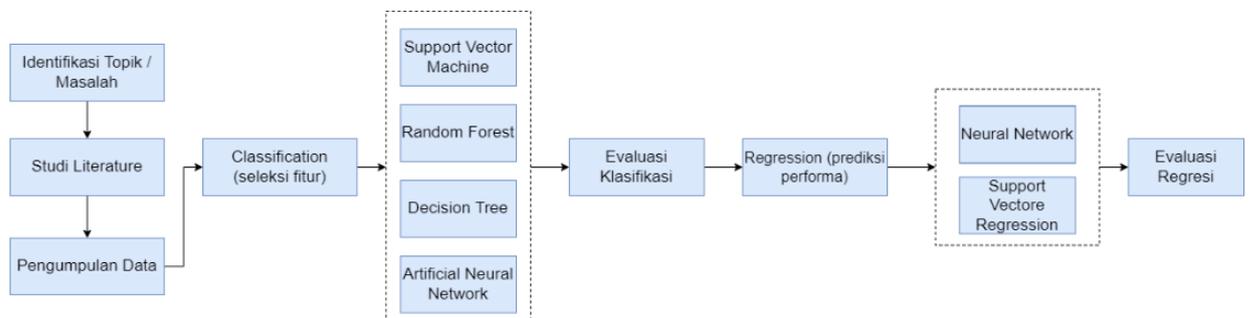
(KNN), dan *Naïve Bayes (NB)*. Hasilnya juga menunjukkan bahwa model ANN mengungguli model lainnya dengan tingkat akurasi yang lebih tinggi. Atribut yang paling dominan sebagai variabel input adalah atribut akademik, demografi, penilaian internal dan keluarga. Analisis mengungkapkan bahwa terdapat peningkatan jumlah penelitian di domain ini dan berbagai algoritma Machine Learning yang diterapkan. Dengan kata lain, ini menunjukkan bukti bahwa *Machine Learning* berguna dalam mengidentifikasi dan meningkatkan berbagai bidang kinerja akademik [6].

Sebuah studi di tempat lain bertujuan untuk mengusulkan pendekatan cerdas baru untuk memprediksi kinerja siswa menggunakan *Support Vector Regression (SVR)* yang dioptimalkan dengan algoritma duel yang disempurnakan (IDA). Model yang diusulkan untuk memprediksi kinerja siswa adalah model hybrid IDA-SVR. Berdasarkan analisis, penelitian ini memberikan kontribusi sebagai berikut: (1) mengusulkan pendekatan cerdas baru untuk memprediksi kinerja siswa menggunakan regresi vektor dukungan (di mana beberapa eksperimen menggunakan data nilai matematika siswa. Hasilnya menunjukkan bahwa model yang diusulkan memiliki kinerja yang sangat baik. kinerja dalam memecahkan masalah prediksi performa siswa, (2) Algoritma duel yang disempurnakan dirancang untuk mengoptimalkan parameter kernel SVR dan fitur-fitur pilihan. Jika dibandingkan dengan metode optimasi parameter lainnya, algoritma IDA dapat secara efektif menghindari pencarian optima lokal dan buta serta meningkatkan konvergensi kecepatan menuju solusi optimal [3]. Meramalkan keberhasilan siswa dapat memungkinkan guru untuk mencegah siswa dari risiko putus sekolah sebelum ujian akhir dengan mengidentifikasi siapa yang membutuhkan bantuan tambahan dan meningkatkan peringkat dan prestise institusi. Penelitian ini membandingkan kinerja beberapa *Supervised Machine Learning Decision*

Trees, Naïve Bayes, Logistic Regression, Support Vector Machines, K-Nearest Neighbors, Sequential Minimal Optimization dan *Neural Networks*. Hasil penelitian ini menunjukkan bahwa pengklasifikasi regresi logistik paling akurat dalam memprediksi nilai akhir siswa [1].

Berdasarkan permasalahan di atas dan penelitian sebelumnya, pemanfaatan teknologi Machine Learning untuk menentukan subtes mana yang paling berpengaruh terhadap performa mahasiswa merupakan satu pendekatan yang akan dilakukan pada penelitian ini. Sehingga proses seleksi pada Politeknik Astra menjadi lebih efektif dan efisien. Pemilihan teknologi Machine Learning dikarenakan ketersediaan data tes potensi akademik yang kurang lebih mencapai 18.818 baris data. Data tersebut berasal dari proses seleksi pada tahun 2016-2021. Selanjutnya, data akan diolah dan dibandingkan dengan data IP semester 1 mahasiswa tahun angkatan 2016-2021, sehingga dapat diketahui subtes mana yang paling berpengaruh terhadap performa mahasiswa.

Tujuan utama penelitian ini adalah untuk mengurangi jumlah subtes sehingga tes seleksi menjadi lebih efektif dan efisien, dari 7 subtest ditargetkan menjadi 3 subtest. Untuk mencapai tujuan utama dilakukan tahapan dengan tujuan lebih detail. Menentukan subtes seleksi yang paling berpengaruh terhadap kelulusan seleksi calon mahasiswa berdasarkan feature importance yang diperoleh menggunakan Machine Learning, Mengembangkan model klasifikasi kelulusan calon mahasiswa berdasarkan tes seleksi yang dilakukan dengan menggunakan pendekatan *Machine Learning* dengan *algoritma Decision Tree, Random Forest, Support Vector Machine*, dan *Artificial Neural Network*, Melakukan prediksi regresi performa mahasiswa berdasarkan nilai tes seleksi masuk dengan menggunakan pendekatan *Machine Learning* dengan *algoritma Support Vector Regression* dan *Neural Network Regression*.



Gambar 1: Metodologi Penelitian

Metode Penelitian

Metodologi penelitian merupakan metodologi atau tahapan kerja dalam penelitian ini. Gambar 1

merupakan metodologi penelitian yang dijadikan acuan tahapan pengerjaan. Tahap ini fokus pada topik *Machine Learning* dengan penggunaan 2 model yaitu klasifikasi dan regresi.

Alur penelitian diawali dengan identifikasi topik/masalah. Pada penelitian kali ini, peneliti mengambil studi kasus pada Institusi Pendidikan Politeknik Astra. Studi kasus akan fokus pada prediksi kinerja mahasiswa berdasarkan nilai seleksi masuk menggunakan metode *Machine Learning* sehingga dapat diketahui subtes mana saja yang mempengaruhi kinerja siswa. Pada tahap selanjutnya akan dilakukan studi literatur. Penelusuran literatur akan berkaitan dengan topik prediksi kinerja siswa dengan Machine Learning yang telah dilakukan pada lembaga pendidikan. Sehingga dapat diketahui apa kelebihan dan kekurangan dari penelitian sebelumnya.

Persiapan dan Analisis Data

Selanjutnya data-data yang dibutuhkan untuk melakukan penelitian akan dikumpulkan. Data yang dibutuhkan adalah Data Nilai Seleksi Masuk Tahun 2016 – 2021. Rentang waktu tersebut diambil karena model tes/tahapan tes pada tahun tersebut relatif sama jadi potensi data redundan atau data kosong dan tidak lengkap lebih kecil. Untuk data seleksi masuk secara lokasi terbagi menjadi dua pada tahun 2016-2018 data masih dikelola menggunakan aplikasi pengolahan data Microsoft Excel. Pada tahun 2019-2021 karena tes sudah menggunakan aplikasi web, data dikelola dengan DBMS Microsoft SQL Server. Sementara untuk data IP Semester 1 dari tahun 2016 – 2021 juga dikelola dengan DBMS Microsoft SQL Server. Pengumpulan data dilakukan dengan mengambil dokumentasi yang ada di Politeknik Astra. Setelah data terkumpul maka data tersebut akan diolah terlebih dahulu. Pra-pemrosesan data bertujuan untuk mempersiapkan data yang akan diolah dari data mentah menjadi data siap diolah sehingga hasil analisis data menjadi lebih akurat. Dalam pra-pemrosesan data terdapat tiga proses utama yaitu pembersihan data, penanganan nilai yang hilang, dan transformasi data.

Klasifikasi

Proses klasifikasi akan dilakukan dengan empat algoritma yaitu *Decision Tree*, *Artificial Neural Network (ANN)*, *Random Forest* dan *Support Vector Machine (SVM)*. Pemilihan empat algoritma bertujuan untuk mendapatkan hasil terbaik dari proses yang dilakukan. Dalam teknik SVM, penulis mencoba mencari fungsi pengklasifikasi terbaik di antara fungsi yang jumlahnya tidak terbatas untuk memisahkan dua jenis objek. Pemisah pada dua dimensi berupa garis, pada tiga dimensi berupa bidang, sedangkan pada dimensi yang lebih tinggi disebut bidang hiper, pada dimensi tiga dimensi disebut bidang, sedangkan pada dimensi yang lebih tinggi disebut bidang hiper. Kelebihan SVM antara lain efektif pada ruang berdimensi tinggi, dan efektif pada kasus dimana jumlah dimensi

lebih besar dibandingkan jumlah sampel [7]. *Artificial Neural Network (ANN)* merupakan suatu pendekatan pemrosesan informasi yang terinspirasi dari sistem jaringan syaraf makhluk hidup khususnya pada sel otak manusia dalam memproses informasi [8]. *Neural Network* terdiri dari banyak elemen pemrosesan yang saling berhubungan dan kemudian bekerja secara paralel untuk menyelesaikan suatu masalah yang umumnya merupakan masalah klasifikasi atau prediksi. Penggunaan teknik ANN memiliki kelebihan antara lain, (i) pembelajaran adaptif karena kemampuan belajar melakukan tugas berdasarkan data yang diberikan, (ii) *self-organization*, yaitu dapat membangun representasi informasi yang diterimanya selama pembelajaran proses belajar mandiri, (iii) waktu operasi nyata karena perhitungannya sendiri dapat dilakukan secara paralel sehingga waktu komputasi lebih cepat [9]. *Decision Tree* merupakan diagram alir berbentuk pohon yang mempunyai simpul akar untuk mengumpulkan data, simpul dalam yang berisi pertanyaan-pertanyaan mengenai data dan simpul daun yang berguna untuk memecahkan masalah dan mengambil keputusan. *Decision Tree* itu sendiri dapat mencegah kita melakukan overfitting pada kumpulan data saat mencapai akurasi maksimum. *Random Forest* merupakan kombinasi dari setiap pohon yang baik, kemudian digabungkan menjadi sebuah model [2]. Tiga poin utama dalam metode *Random Forest* adalah (i) bootstrap sampling untuk membangun pohon prediksi, (ii) setiap pohon keputusan melakukan prediksi dengan prediktor acak, (iii) kemudian *Random Forest* melakukan prediksi dengan menggabungkan hasil dari setiap pohon keputusan. Prediksinya berdasarkan suara terbanyak untuk klasifikasi, sedangkan regresi menggunakan rata-rata [10].

Seleksi fitur akan dilakukan dengan algoritma terbaik dari proses klasifikasi yang telah dilakukan sebelumnya. Seleksi fitur bertujuan untuk mengetahui 3 fitur yang paling berpengaruh terhadap kelulusan calon mahasiswa pada tahapan tes potensi akademik. Seleksi fitur menggunakan metode filter dalam pemilihan fitur yang paling berpengaruh dan untuk korelasi secara keseluruhan menggunakan metode impurity based.

Seleksi fitur akan membentuk beberapa konfigurasi data untuk data MI dan non-MI. Data akan dikonfigurasi menjadi 3 yaitu data dengan 3 fitur, 5 fitur dan *full feature* / 7 fitur. Data tersebut akan dilakukan proses klasifikasi ulang. Sebelum dilakukan regresi. Proses klasifikasi ulang untuk melihat mana konfigurasi terbaik dilihat dari hasil evaluasi klasifikasi.

Evaluasi Klasifikasi

Evaluasi performa dilakukan terhadap setiap algoritma dan konfigurasi data. Evaluasi klasifikasi menggunakan confusion matrix. Ada 4 hal yang

akan dievaluasi yaitu *accuracy*, *precision score*, *recall score*, dan *F1 Score*. Pada kasus normal dalam arti jumlah data memadai akan digunakan *accuracy* sebagai tolok ukur utama, namun pada kasus ini akan digunakan *F1 Score* karena jumlah data tidak begitu besar. Perhitungan evaluasi keakuratan sebuah metode pengelompokan dapat dilakukan dengan formula sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$F1Score = \frac{2 * (Recall \times Precision)}{Recall + Precision} \quad (4)$$

dimana

TP : jumlah data positif yang terdeteksi benar

TN : jumlah data negatif yang terdeteksi dengan benar

FP : jumlah data negatif namun terdeteksi sebagai data positif

FN : jumlah data positif namun terdeteksi sebagai data negatif

Evaluasi pengujian didasarkan pada tinjauan positif dan negatif yang dihitung dengan pendekatan yang diusulkan. Notasi TP menunjukkan *True Positives*, FP menunjukkan *False Positives*, TN menunjukkan *True Negatives* dan FN menunjukkan *False Negative*.

Regresi

Pada tahap selanjutnya akan dilakukan dengan metode regresi. Pemilihan metode regresi selain dikarenakan sudah ada kriteria/label juga karena data yang dimiliki sesuai dengan data regresi yaitu data IP semester 1. Untuk melakukan regresi digunakan dua algoritma regresi yaitu *Support Vector Regression* dan *Neural Network*. Dua model tersebut diharapkan mampu menghasilkan akurasi yang baik dalam proses prediksi regresi.

Proses prediksi regresi dilakukan dengan dua cara. Cara pertama adalah data tes potensi akademik yang telah diseleksi fitur akan diproyeksikan terhadap data IP semester 1. Kemudian cara kedua adalah data tes potensi akademik yang tidak terpilih atau tidak lulus seleksi fitur (konfigurasi data dengan variasi jumlah fitur) akan diproyeksikan terhadap data IP semester 1. Kedua proses tersebut akan dibandingkan mana yang memiliki hasil lebih baik.

Hasil yang diharapkan adalah menemukan fitur atau subtes yang paling berpengaruh terhadap performa mahasiswa. Jika subtes tersebut berasal dari seleksi fitur pada proses klasifikasi maka selama ini subtes yang digunakan telah sesuai atau subtes tersebut dapat menjadi acuan utama dalam penentuan kelulusan calon mahasiswa. Jika subtes yang paling berpengaruh tidak berasal dari seleksi fitur maka penentuan kelulusan yang dilakukan masih belum maksimal. Sehingga perlu dilakukan prioritas terhadap subtes yang paling berpengaruh.

Evaluasi Regresi

Pada metode regresi, setiap modelnya akan dilakukan evaluasi dengan *Mean Absolute Error* (MAE).

$$MAE = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n} \quad (5)$$

dimana,

Y_i : output sebenarnya (ground truth) dari data ke-i.

\hat{Y}_i : nilai prediksi ke-i

n : banyaknya data

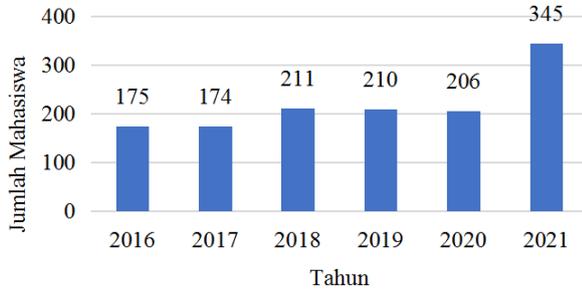
Perhitungan tersebut akan melihat proyeksi IP Semester 1 dari data test kemudian dihitung selisih dari setiap prediksi yang dilakukan dan kemudian akan dirata-rata. Adapun target MAE untuk proses regresi yaitu dibawah 0.2. Sehingga jika IP asli seorang mahasiswa adalah 3.5, maka hasil prediksi masih pada range 3.3-3.7, hal tersebut masih dianggap wajar sesuai dengan kebutuhan studi kasus. Hasil analisis data akan terdiri dari beberapa hal. Hal pertama adalah fitur seleksi yang paling berpengaruh terhadap status kelulusan mahasiswa. Kedua adalah hasil prediksi kinerja dengan regresi sehingga akan diketahui fitur atau subtes mana yang paling besar pengaruhnya terhadap kinerja mahasiswa. Selanjutnya setiap algoritma yang digunakan akan dibandingkan, algoritma mana yang mempunyai akurasi terbaik dalam hal ini.

Hasil dan Pembahasan

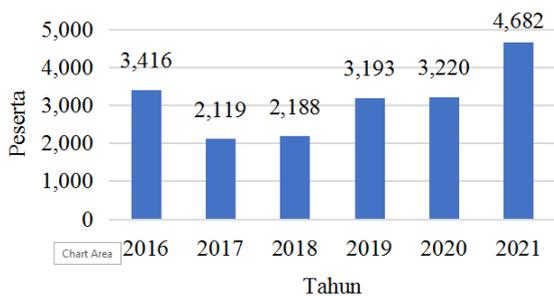
Data yang akan menjadi fokus pada penelitian ini adalah data nilai seleksi Tes Potensi Akademik dan Data IP Semester 1. Pada awalnya data nilai seleksi TPA akan menjadi satu buah model data namun ternyata ada perbedaan fitur antara Prodi Manajemen Informatika dengan program studi lainnya. Perbedaan terletak pada subtes yang diujikan. Pada Prodi MI yang diujikan terdapat subtes PAT, sementara pada Prodi Non-MI digantikan dengan mata kuliah fisika. Sehingga model dibagi dua yaitu data TPA tahun 2016-2018 Prodi MI, dan 2016-2018 Prodi Non-MI atau keteknikan.

Selanjutnya data yang telah terkumpul akan direkapitulasi untuk mengetahui jumlah total data

yang terkumpul. Untuk data seleksi yang masuk dari tahun 2016 – 2018 berjumlah 7.723 baris data. Rincian jumlah data seleksi yang masuk setiap tahunnya dapat dilihat pada Gambar 2. Sedangkan untuk data IP Semester 1 terdapat 560 data. Rincian jumlah data IP Semester pada setiap tahunnya dapat dilihat pada Gambar 3.



Gambar 2: Diagram Batang IP Mahasiswa Semester 1 Politeknik Astra



Gambar 3: Peserta Tes Potensi Akademik Politeknik Astra

Data awal yang didapatkan dalam bentuk format file Microsoft Excel. Data masih terdapat seluruh atribut data pendaftaran seperti data alamat, sma, Prodi pilihan dan lainnya. Contoh data mentah dapat dilihat pada tabel 3.1. Langkah selanjut-

nya menghapus kolom-kolom data yang tidak dibutuhkan. Selanjutnya, akan difilter untuk Prodi MI dan Non-MI. Keduanya akan dipisah pada file yang berbeda dengan format csv. Pada kedua model data terdapat dengan format decimal atau float.

Pada proses pengolahan data menggunakan Jupyter Notebook Anaconda 3 akan dilakukan ekspor model data dengan format csv. Setelah dilakukan ekspor, perlu dipastikan tidak ada missing data dan melihat overview data yang berhasil diekspor. Selanjutnya perlu dilihat tipe data dari setiap kolom yang ada. Pada model data yang memiliki nilai desimal masih terbaca sebagai object, artinya pemisah desimal “.” tidak dibaca sebagai pemisah. Hal tersebut perlu dilakukan penggantian pemisah desimal menjadi “.”. Selanjutnya data hasil masih berupa string yaitu “pass” dan “failed”. Untuk menjadikan data numerical maka akan digunakan Teknik feature engineering yaitu one hot encoding. Di mana nilai string akan langsung dipetakan dengan nilai numerical tertentu. Pada model data nilai TPA, hasil pass akan dipetakan menjadi nilai 1, dan failed akan dipetakan menjadi nilai 0. Selanjutnya datanya akan dibagi menjadi dua yaitu untuk Program Studi Manajemen Informatika dan Program Studi Non Manajemen Informatika. Langkah ini diambil karena terdapat perbedaan fitur pada masing-masing model data. Pada Program Studi Manajemen Informatika terdapat subtes PAT, namun pada Program Studi Non Manajemen Informatika tesnya diganti dengan tes fisika. Hasilnya dapat dilihat pada Tabel 1.

Tabel 1: Contoh Data

Nama	Pilihan_I	PAT	BHS_I NGGR IS	ASSEMBLY	RESULT
Mahasiswa 1	MI	53	6	6	1
Mahasiswa 2	MI	1	1.4	1	0
Mahasiswa 3	MI	35	1.4	6	1
Mahasiswa 4	MI	4	3.4	4	0

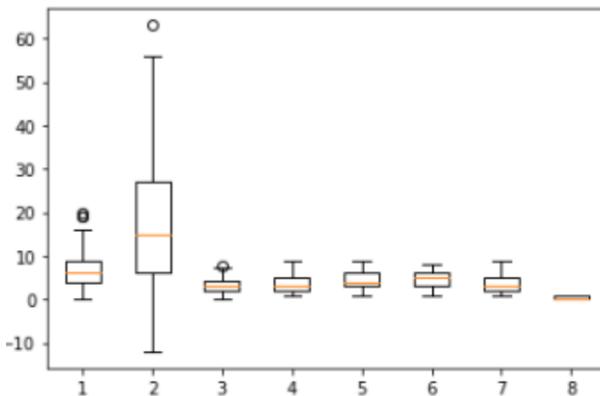
Tabel 2: Deskripsi Data MI

	MAT	PAT	BAHASA INGGRIS	SPATIAL_V	P_SPEED	ASSEMBLY	SYMBOLIC REASONING	RESULT
count	371	371	371	371	371	371	371	371
mean	6.61	17.22	3.20	3.78	4.32	4.23	3.55	0.26
std	3.45	14.96	1.51	2.19	2.18	1.88	2.02	0.44
min	0	-12	0	1	1	1	1	0
25%	4	6	2	2	3	3	2	0
50%	6	15	3	3	4	5	3	0
75%	9	27	4.2	5	6	6	5	1
max	20	63	7.6	9	9	8	9	1

Tabel 3: Deskripsi Data Non-MI

	MAT	FISIKA	BHS INGGRIS	SPATIAL V	P SPEED	ASSEMBLY	SYMBOLIC REASONING	RESULT
count	1688	1688	1688	1688	1688	1688	1688	1688
mean	5.97	16.13	2.84	3.90	3.90	4.22	3.20	0.25
std	3.27	4.55	1.45	2.13	2.18	1.98	1.97	0.43
min	0	0	0	1	1	1	1	0
25%	4	13	1.8	2	2	3	1	0
50%	6	16	2.6	4	4	5	3	0
75%	8	19	3.8	5	5	6	5	1
max	18	32	9.6	9	9	9	9	1

Jumlah data pada setiap model tergolong sedikit yaitu pada Prodi MI total 300 data dan non-MI sekitar 1000 data. Pada Prodi MI akan memiliki tujuh fitur yaitu Matematika, PAT, Bahasa Inggris, *Spatial Visualization*, *Perceptual Speed*, *Assembly*, dan *Symbolic Reasoning*. Untuk Prodi Non-MI sama kecuali subtes PAT digantikan dengan subtes fisika. Hal tersebut berisiko menurunkan tingkat akurasi pada proses pengolahan data. Sebelum diolah kita perlu memastikan kesiapan data dengan melihat tren dan sebaran data, hasilnya dapat dilihat pada Tabel 2 untuk Prodi MI dan Non-MI bisa dilihat pada Tabel 3.

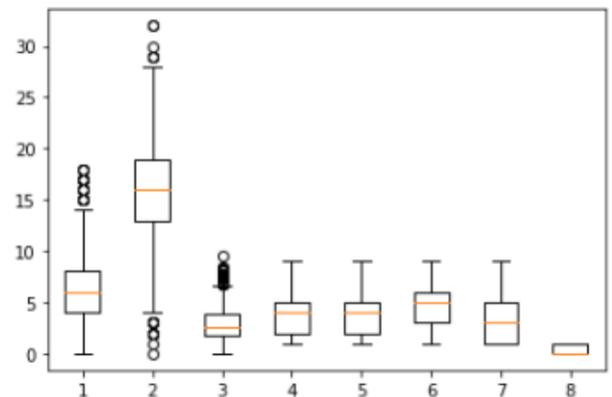


Gambar 4: Boxplot MI

Berdasarkan hasil pada kedua tabel tersebut, bisa dilihat rata-rata nilai, nilai maksimal, nilai minimal, dan quartil. Selanjutnya perlu dilihat apakah ada outlier dan seberapa banyak data outlier yang ada. Jika ada outlier hasil analisa data akan berpengaruh. Setelah dibuat grafik boxplot terdapat beberapa outlier namun masih dalam jumlah kecil sehingga tidak begitu berpengaruh terhadap hasil analisa. Hasil boxplot Prodi MI dapat dilihat pada Gambar 4 dan non-MI dapat dilihat pada Gambar 5.

Proses klasifikasi dimulai dengan menyiapkan atau membagi data menjadi data training dan data tes. Jumlah data size pada uji coba yang dilakukan berjumlah 10% dari seluruh data. Penentuan jumlah data tes didasarkan pada jumlah keseluruhan data yang tidak begitu besar. *Classification* menggunakan 4 metode algoritma yaitu *Decision Tree*, *Random Forest*, *Support Vector Machine*, dan *Neural Network*. Pada setiap algoritma akan dilakukan setting atau tuning parameter untuk dapat menghasilkan F1 Score terbaik. Pemilihan F1 Score sebagai alat ukur utama menentukan

keberhasilan karena jumlah data tidak begitu besar akan berisiko tidak valid hasilnya jika menggunakan accuracy saja. Adapun target yang ingin dicapai adalah F1 Score di atas 90%. Percobaan pertama untuk *classification* menggunakan model algoritma *Decision Tree*. Hasil terbaik yang didapatkan dapat dilihat pada Tabel 4. Hasil tersebut didapatkan dari *tuning parameter criterion = gini*, dan *max_depth=default*.



Gambar 5: Boxplot non-MI

Tabel 4: *Classification Decision Tree*

Jenis Prodi	Parameter	Hasil
NON Manajemen Informatika	Accuracy	0.9586
	Recall	0.9130
	Precision	0.9333
	F1	0.9231
Manajemen Informatika	Accuracy	0.9189
	Recall	0.7143
	Precision	0.8333
	F1	0.7692

Percobaan Kedua untuk classification menggunakan model algoritma *Random Forest*. Hasil terbaik yang didapatkan dapat dilihat pada Tabel 5. Hasil tersebut didapatkan dari tuning parameter *criterion="entropy",max_depth=8, random_state=123, max_features="sqrt", class_weight="balanced"*.

Tabel 5: *Classification Random Forest*

Jenis Prodi	Parameter	Hasil
NON Manajemen Informatika	Accuracy	0.9763
	Recall	10,000
	Precision	0.9200
	F1	0.9583
Manajemen Informatika	Accuracy	0.9730
	Recall	0.8571
	Precision	10,000
	F1	0.9231

Percobaan ketiga untuk *classification* menggunakan model algoritma *Support Vector Machine*. Hasil terbaik yang didapatkan dapat dilihat pada Tabel 6. Hasil tersebut didapatkan dari parameter *kernel='rbf', gamma='scale', probability='false'*.

Tabel 6: *Classification Support Vector Machine*

Jenis Prodi	Parameter	Hasil
NON Manajemen Informatika	Accuracy	0.8817
	Recall	0.7174
	Precision	0.8250
	F1	0.7674
Manajemen Informatika	Accuracy	0.8649
	Recall	0.5714
	Precision	0.6667
	F1	0.6154

Percobaan keempat untuk *classification* menggunakan model algoritma *Neural Network*. Hasil terbaik yang didapatkan dapat dilihat pada Tabel 7. Hasil tersebut didapatkan dari parameter *random_state=1, max_iter=2000, activation="logistic"*.

Tabel 7: *Classification Neural Network*

Jenis Prodi	Parameter	Hasil
NON Manajemen Informatika	Accuracy	0.9112
	Recall	0.7609
	Precision	0.8974
	F1	0.8235
Manajemen Informatika	Accuracy	0.9189
	Recall	0.8571
	Precision	0.7500
	F1	0.8000

Percobaan dengan keempat model algoritma menghasilkan bahwa model algoritma *Random For-*

est dapat menghasilkan F1 score terbaik. Hal tersebut juga tampak pada *Accuracy, Recall, dan Precision*. Hasil tersebut konsisten pada kedua jenis model data yaitu MI dan Non-MI. Sehingga semakin memperkuat bahwa pada studi kasus ini dan dengan jumlah data kurang dari 1000 model *Random Forest* dapat memberikan hasil terbaik. Untuk Prodi MI pada beberapa model algoritma hasilnya dibawah Non-MI karena jumlah data terlalu sedikit untuk melakukan training. Hasil pengujian model Manajemen Informatika dapat dilihat pada Tabel 8. Kemudian hasil pengujian model Non Manajemen Informatika dapat dilihat pada Tabel 9.

Tabel 8: Hasil Klasifikasi Manajemen Informatika

No.	Algoritma	Accuracy Score	Recall Score	Precision Score	F1-Score
1	Decision Tree	0.9189	0.7143	0.8333	0.7692
2	Random Forest	0.9730	0.8571	1.0000	0.9231
3	Support Vector Machine	0.8649	0.5714	0.6667	0.6154
4	Neural Network	0.9189	0.8571	0.7500	0.8000

Tabel 9: Hasil Klasifikasi Non Manajemen Informatika

No.	Algoritma	Accuracy Score	Recall Score	Precision Score	F1-Score
1	Decision Tree	0.9586	0.9130	0.9333	0.9231
2	Random Forest	0.9763	1.0000	0.9200	0.9583
3	Support Vector Machine	0.8817	0.7174	0.8250	0.7674
4	Neural Network	0.9408	0.8261	0.9500	0.8837

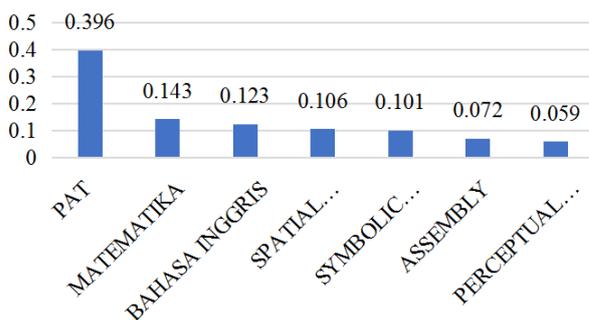
Hasil yang baik dari algoritma *Random Forest* diperoleh setelah tuning parameter. Kriteria terbaik dalam hal ini adalah entropi. Kriteria entropi menghitung entropi Shannon dari kelas-kelas yang mungkin. Nilai kedalaman maksimum adalah 8 yang memberikan nilai terbaik pada jenis data dan fitur ini. Jumlah kedalaman maksimal mungkin berbeda dalam kasus yang berbeda. Penyetelan terakhir adalah bobot kelas. Bobot yang dikaitkan dengan kelas dalam bentuk $\{class_label:weight\}$. Kalau tidak diberikan nilai, maka akan memberikan nilai default dimana semua kelas seharusnya berbobot satu. Untuk masalah multi-output, daftar dicts dapat diberikan dalam urutan yang sama dengan kolom y. Mode "balanced" menggunakan nilai y untuk secara otomatis menyesuaikan bobot yang berbanding terbalik dengan frekuensi kelas dalam data masukan sebagai $n_samples / (n_classes * np.bincount(y))$.

Feature selection digunakan untuk mengetahui *feature importance* dari beberapa *feature* yang ada pada model data nilai TPA, sehingga dapat dilihat subtes mana yang mempengaruhi kelulusan calon mahasiswa. Metode seleksi fitur yang digu-

nakan adalah filter untuk mengetahui feature importance. Feature akan dilihat tiga tertinggi dari masing masing model data. Selanjutnya akan dilihat lima feature dengan skor tertinggi dari masing-masing model. Proses ini dilakukan untuk membagi setiap model menjadi tiga model data sehingga akan dimiliki enam model data. Pembagian bertujuan untuk memperkuat proses analisis terhadap data yang dimiliki dan juga sebagai tolok ukur seberapa banyak subtes yang dibutuhkan untuk dapat menyaring calon mahasiswa dengan hasil terbaik.

Setiap model yang telah dihasilkan berdasarkan pembagian nilai dari feature selection akan dilakukan percobaan classification ulang. Percobaan ulang dilakukan untuk melihat model data terbaik. Hal tersebut juga akan berlaku pada proses selanjutnya yaitu regression. Dari proses berulang dan membagi model menjadi beberapa bagian, dapat diketahui dengan benar mana saja subtes yang sebenarnya berpengaruh terhadap proses seleksi dan performa mahasiswa untuk menjawab kebutuhan dari studi kasus. Feature selection dilakukan menggunakan hasil dari algoritma Random Forest.

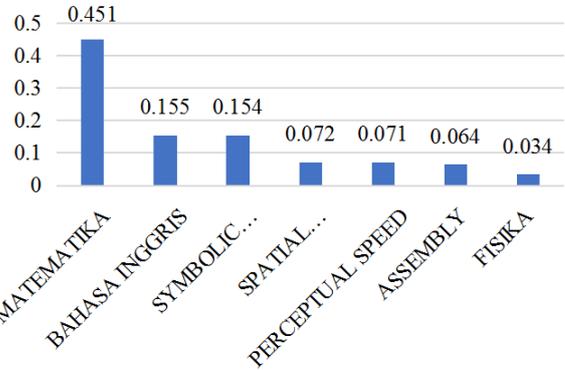
Hasil feature selection dari model Prodi MI dapat dilihat pada Gambar 6. Berdasarkan hasil tersebut untuk model data MI dapat dilihat 3 feature yang memiliki nilai tertinggi yaitu PAT, matematika dan Bahasa Inggris. Nilai tertinggi dan dominan diperoleh PAT dibandingkan dengan materi subtes lainnya. Pada model data MI akan dibagi menjadi 3 model, yaitu model pertama memiliki tujuh feature dengan seluruh subtes menjadi feature, model kedua dengan lima feature (PAT, Matematika, Bahasa Inggris, spatial visualization, dan assembly), dan model ketiga dengan tiga feature (PAT, Matematika, Bahasa Inggris).



Gambar 6: Feature Importance Manajemen Informatika

Hasil feature selection pada Prodi Non-MI dapat dilihat pada Gambar 7. Berdasarkan hasil tersebut, untuk model data Non-MI dapat dilihat 3 feature yang memiliki nilai tertinggi yaitu Matematika, Bahasa Inggris dan *Symbolic Reasoning*. Nilai tertinggi dan dominan diperoleh Matematika dibandingkan dengan materi subtes lainnya. Pada model data non-MI akan dibagi menjadi 3 model,

yaitu model pertama memiliki tujuh feature dengan seluruh subtes menjadi feature, model kedua dengan lima feature (Matematika, Bahasa Inggris, *Symbolic Reasoning*, *Spatial Visualization*, *Perceptual Speed*), dan model ketiga dengan tiga feature (Matematika, Bahasa Inggris dan *Symbolic Reasoning*).



Gambar 7: Feature Importance Non Manajemen Informatika

Berdasarkan hasil pengujian di atas terlihat bahwa pendekatan *Machine Learning* dapat menghasilkan model agar tes seleksi masuk perguruan tinggi menjadi lebih efektif dan efisien. Hasil seleksi fitur menunjukkan tiga tes yang paling berpengaruh terhadap kelulusan, yang awalnya terdiri dari tujuh subtes. Awalnya membutuhkan waktu kurang lebih 180 menit untuk menyelesaikan seluruh subtes, namun dengan hasil di atas maka kita akan mencoba tiga subtes saja yang berarti hanya membutuhkan waktu kurang lebih 90 menit untuk menyelesaikan seluruh subtes. Uji coba akan dilakukan dengan data TPA angkatan 2019 berdasarkan model yang telah dihasilkan.

Dalam hal ini diketahui bahwa Politeknik Astra merupakan lembaga pendidikan yang menghasilkan seorang insinyur. Dari hasil seleksi fitur di atas terdapat kesamaan pada subtes matematika. Hal ini menunjukkan bahwa matematika merupakan salah satu keterampilan dasar yang harus dimiliki oleh seorang insinyur. Sampai tahap ini, matematika dapat dikatakan mempunyai pengaruh paling besar terhadap tes kelulusan siswa. Untuk mengetahui apakah hal tersebut juga mempengaruhi kinerja siswa akan dilakukan pada tahap selanjutnya.

Model data nilai seleksi TPA akan digabungkan dengan data IP semester. Data IP semester akan menjadi target yang dituju dalam percobaan regression. Percobaan regression ingin melihat seberapa berpengaruh nilai seleksi TPA terhadap performa mahasiswa, sehingga dapat dilihat potensi performa mahasiswa dari seleksi masuk. Selain itu pada regression juga dapat dilihat nilai seleksi TPA/subtes mana yang paling berpengaruh terhadap performa mahasiswa. Pada proses regression ada target error test dan error training ≤ 0.2 . Hal tersebut berarti jika seorang mahasiswa pada aktualnya mendapat

atkan IP 3.5 maka seburuk-buruknya IPnya akan diprediksi menjadi 3.3-3.7. Nilai tersebut sudah cukup relevan untuk menggambarkan performa mahasiswa.

Pada proses regression akan digunakan dua model algoritma yaitu Support Vector Regression, dan Neural Network Regression. Penggunaan dua model bertujuan untuk mendapatkan hasil terbaik atau nilai error paling rendah. Percobaan dengan ketiga model data akan dilakukan dengan 6 model data yang telah dibagi berdasarkan seleksi feature. Tujuannya untuk melihat pengaruh disederhanakannya subtes dari seleksi TPA. Jumlah data tes yang digunakan adalah 10% dari jumlah keseluruhan data.

Pada setiap algoritma akan dilakukan tuning parameter untuk mendapatkan hasil maksimal. Tuning yang dilakukan pada algoritma Support Vector Regression dengan memberikan nilai pada parameter kernel = 'rbf', gamma = 'scale', regularization parameter/c = '1.0', dan epsilon = '0.1'. Tuning parameter tersebut menghasilkan hasil terbaik pada algoritma Support Vector Regression. Untuk algoritma Neural Network, tuning parameter yang dilakukan adalah memberikan nilai pada parameter activation = 'relu' (fungsi linier mengembalikan nilai $f(x) = \max(0,x)$, solver='adam', max_iter='500'. Hasil dari setiap algoritma dan model dapat dilihat pada Tabel 10. Hasil terbaik ditunjukkan oleh model algoritma Support Vector Regression untuk kedua model MI dan Non-MI. Pada data MI hasil terbaik dihasilkan model data dengan 3 feature sementara untuk Prodi Non-MI ada 7 feature. Target error test dapat tercapai pada percobaan menggunakan ini.

Tabel 10: Hasil Regresi

No.	Algoritma	Feature	NON Manajemen Informatika		Manajemen Informatika	
			Error Train	Error Test	Error Train	Error Test
1	Support Vector Regression	7	0.2195	0.1958	0.1718	0.1810
2	Support Vector Regression	3	0.2120	0.2723	0.1853	0.1827
3	Neural Network Regression	7	0.2339	0.2087	0.2408	0.2175
4	Neural Network Regression	3	0.2137	0.2591	0.1903	0.2052

Untuk Program Studi Manajemen Informatika, hasil terbaik ditunjukkan dengan penggunaan 3 fitur. Sedangkan untuk Program Studi Non Manajemen Informatika hasil terbaik ditunjukkan dengan penggunaan 7 fitur. Artinya pemilihan Program Studi Non Manajemen Informatika tidak bisa disederhanakan. Jika dilihat kembali dari hasil pemilihan fitur, perbedaan nilai tiap fitur tidak terlalu signifikan. Dimana setiap nilai masih saling mempengaruhi terhadap kelulusan dan juga kinerja siswa.

Penutup

Hasil klasifikasi menunjukkan bahwa 7 subtes yang telah digunakan dalam tes seleksi memberikan skor F1 yang tinggi pada evaluasi. Skor F1 digunakan sebagai parameter karena jumlah data yang sedikit. Model dengan Skor F1 tertinggi adalah Random Forest. Hasilnya konsisten dalam dua data model dengan subtes yang berbeda.

Pemilihan fitur menunjukkan bahwa subtes yang penting dalam Manajemen Informatika adalah PAT, Matematika, dan Bahasa Inggris. Nilai PAT menjadi sangat dominan karena PAT sendiri merupakan tes untuk melihat potensi seorang programmer. Sementara untuk Prodi Non Manajemen Informatika adalah Matematika, Bahasa Inggris, Symbolic Reasoning. Dari seleksi fitur diperoleh kesimpulan bahwa seorang engineer harus memiliki kemampuan matematika dasar sesuai dengan jurusan yang ada pada Politeknik Astra.

Hasil terbaik konsisten dengan 2 model data yaitu algoritma Support Vector Regression. Namun terdapat perbedaan bahwa untuk Program Studi Manajemen Informatika model data dengan 3 fitur memiliki hasil terbaik, artinya tes potensi akademik dapat disederhanakan dengan 3 fitur tersebut. Sedangkan pada Prodi Non Manajemen Informatika, hasil terbaik masih pada 7 fitur. Subtes perlu didesain ulang agar dapat memfilter dengan baik.

Penelitian ini berfokus pada prediksi performa mahasiswa berdasarkan nilai seleksi masuk dan mendapatkan error test rata-rata 0.18. Hasil tersebut sangat baik namun hasil performa dari mahasiswa dapat dipengaruhi oleh banyak faktor. Faktor-faktor lainnya perlu dilihat juga untuk melihat pemetaan dari calon mahasiswa. Contoh faktor-faktor lain misalkan data diri, data keluarga, dan data-data kualitatif lainnya yang bisa mempengaruhi mahasiswa.

Jadi, pada penelitian selanjutnya perlu dilakukan analisa data bukan hanya data dari data kuantitatif tetapi juga dengan data kualitatif. Selain itu untuk Prodi MI perlu dicoba untuk menggunakan tiga subtes dengan pengaruh tertinggi dalam seleksi masuk. Sementara untuk Prodi Non-MI harus melakukan evaluasi subtes agar mendapatkan hasil seleksi lebih sesuai untuk menyaring calon mahasiswa sehingga bisa relevan dengan performa mahasiswa.

Ucapan Terimakasih

Saya berterima kasih kepada Pembimbing, Ibu Lili Ayu Wulandhari atas informasi dan pengetahuan yang diberikan kepada saya. Saya juga mengucapkan terima kasih kepada Politeknik Astra yang sudah berkenan menyediakan informasi yang diperlukan untuk penelitian ini. Selain itu, saya juga mengucapkan terima kasih kepada keluarga saya

yang selalu mendukung saya saat mengerjakan jurnal.

Daftar Pustaka

- [1] A. S. Hashim, W. A. Awadh and A. K. Hamoud, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms", in 2nd International Scientific Conference of Al-Ayen University (ISCAU-2020), vol. 928, doi:10.1088/1757-899X/928/3/032019, 2020.
- [2] A. D. Kumar, V. Selvam and R. P. Palanisamy, "Hybrid Classification Algorithms for Predicting Student Performance", in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), pp. 1074–1079, doi: 10.1109/ICAIS50930.2021.9395974, 2021.
- [3] H. Xu, "Prediction of Students' Performance Based on the Hybrid IDA-SVR Model", Complexity, vol. 2022, p. 11, doi.org/10.1155/2022/1845571, 2022.
- [4] S. Hussain, Z. F. Muhsin, Y. K. Salal, P. Theodorou, F. Kurtoglu and G. C. Hazarika, "Prediction Model on Student Performance based on Internal Assessment using Deep Learning", Int. J. Emerg. Technol. Learn., vol. 14, no. 8, pp. 4–22, doi: 10.3991/ijet.v14i08.10001, 2019.
- [5] A. A. Setiawan, "Prediction Multiclass SVM dengan Optimize Parameter Grid Untuk Mem-
- prediksi Performance Student", Indones. J. Informatics Res., vol. 3 No. 1, pp. 36–41, 2022.
- [6] Y. A. Alsariera, Y. Baashar, G. Alkaws, A. Mustafa, A. A. Alkahtani, and N. Ali, "Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance", Comput. Intell. Neurosci., vol. 2022, pp. 1–11, doi.org/10.1155/2022/4151487%0A, 2022.
- [7] J. D. Boissberranger, "1.4. Support Vector Machines — scikit-learn 0.20.2 Documentation", Scikit-Learn, 2020 diakses daring pada : <https://scikitlearn.org/stable/modules/svm.html>
- [8] A. Yanuar, "Artificial Neural Network (Video)", Universitas Gajah Mada, 2018 diakses daring pada :<https://machinelearning.mipa.ugm.ac.id/2018/09/29/artificial-neural-network-video/>
- [9] H. D. Widiputra, "Artificial Neural Network", Perbanas Institute Jakarta, 2016 diakses daring pada : <https://dosen.perbanas.id/artificial-neural-network/>
- [10] Muhammad Rivza Adrian, Muhammad Papuandivitama Putra, Muhammad Hilman Rafialdy dan Nur Aini Rakhmawati, "Perbandingan Metode Klasifikasi Random Forest dan SVM Pada Analisis Sentimen PSBB", Jurnal Informatika Upgris, Vol.7, no.1 June, pp. 6–11, doi: 10.26877/jiu.v7i1.7099, 2021

Halaman ini sengaja dikosongkan.