## Perbandingan Algoritma Klasifikasi Decision Trees dan Support Vector Machine dalam Memprediksi Keberhasilan Akademis dan Drop-Out Mahasiswa

###### Lian Wira Manuel Maharaja1,

1 Sistem Informasi, Fakultas Teknik Informatika,

, 1lian.wira@student.umn.ac.id

***Abstract***-This study delves into a comparative analysis of Decision Trees and Support Vector Machines (SVM) as predictive modeling tools for anticipating student academic outcomes in higher education. Utilizing a comprehensive dataset encompassing diverse factors such as demographic information, socio-economic indicators, academic performance metrics, and regional economic variables, we investigate the effectiveness of these machine learning algorithms. Decision Trees and SVM are evaluated based on their performance in predicting academic success and potential student drop-out. The study aims to provide valuable insights into the strengths and weaknesses of each algorithm, offering a nuanced understanding of their applicability in the context of diverse undergraduate programs. The findings contribute to the ongoing discourse on leveraging machine learning for educational analytics, assisting educators and institutions in enhancing strategies for student retention and academic success across various disciplines.

***Keyword :*** *machine learning, decision trees, support vector machine, student academic performance, comparative study*

1. **PENDAHULUAN**

**1.1 Latar Belakang Dan Tujuan**

Mahasiswa merupakan elemen yang signifikan dalam

penilaian kesuksesan pelaksanaan program studi di perguruan tinggi. Secara umum, mahasiswa memiliki variasi kualitas yang sangat beragam di antara satu sama lain. Keberagaman kualitas ini memiliki dampak yang signifikan pada perjalanan studi mahasiswa. Individu dengan tingkat kualitas tertentu mungkin memiliki risiko tinggi mengalami drop out [1]. Pendidikan tinggi juga merupakan aspek kritis dalam pembentukan sumber daya manusia berkualitas dan berkelanjutan. Di era digital ini, penting untuk memahami faktor-faktor yang mempengaruhi keberhasilan akademis mahasiswa dan potensi drop-out guna mengembangkan strategi intervensi yang efektif. Dalam kerangka ini, algoritma pembelajaran mesin, seperti Decision Trees dan Support Vector Machines (SVM), menawarkan potensi untuk memberikan pemahaman yang mendalam tentang pola kompleks yang dapat memprediksi hasil akademis.

Meskipun telah ada penelitian sebelumnya tentang pemodelan prediktif untuk mahasiswa, masih ada kebutuhan untuk mendalami lebih jauh pada perbandingan kinerja antara Decision Trees dan SVM, terutama dalam konteks program sarjana yang beragam. Laporan yang komprehensif tentang perbandingan ini dapat memberikan pandangan yang lebih tajam terhadap kelebihan dan kelemahan masing-masing algoritma. Oleh karena itu, penelitian ini bertujuan untuk menyelidiki dan membandingkan kinerja Decision Trees dan

SVM dalam memprediksi hasil akademis mahasiswa di lingkungan pendidikan tinggi. Dengan menggali data yang mencakup informasi demografis, indikator sosial-ekonomi, kinerja akademis, dan faktor ekonomi regional, penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam tentang faktor-faktor yang berkontribusi terhadap keberhasilan akademis mahasiswa dan potensi drop-out.

Dengan penelitian ini, berharap dapat memberikan kontribusi pada literatur tentang analisis pendidikan tinggi dan menghasilkan temuan yang dapat membantu pengambilan keputusan di lembaga pendidikan dan pemerintah untuk meningkatkan efektivitas strategi pendukung mahasiswa dan mendukung keberhasilan akademis di berbagai disiplin studi.

1. **STUDI LITERATURE**

**2.1 Klasifikasi**

Klasifikasi merupakan proses yang melibatkan pencarian dan identifikasi sekumpulan model, pola, atau fungsi yang dapat menggambarkan serta membedakan objek data. Tujuan utama dari klasifikasi adalah mengelompokkan objek data ke dalam kelas-kelas tertentu dari sejumlah kelas yang tersedia. Dengan menggunakan berbagai metode dan algoritma klasifikasi, dapat dilakukan identifikasi pola yang memungkinkan pengelompokan yang akurat berdasarkan karakteristik atau atribut tertentu [2].

**2.1 Algoritma Decision Trees**

Decision Tree disebut juga dengan pohon keputusan merupakan model dari klasifikasi. Bentuknya yang seperti struktur pohon merepresentasikan atribut setiap data yang diproses. Decision Tree yaitu mendeskripsikan tiap-tiap kelas untuk menemukan pola atau fungsi [3].

Decision Tree juga sering digunakan untuk mengumpulkan informasi guna mendukung proses pengambilan keputusan. Algoritma ini memungkinkan pembelajaran pola klasifikasi dan prediksi dari data, serta menggambarkan hubungan antara variabel atribut *x* dan variabel target *y* dalam bentuk pohon [19]. Keunggulan Decision Tree meliputi transformasi daerah pengambilan keputusan yang sebelumnya kompleks dan global menjadi lebih sederhana dan spesifik, eliminasi perhitungan yang tidak diperlukan dengan menguji sampel berdasarkan kriteria atau kelas tertentu, serta fleksibilitas dalam memilih fitur dari node internal yang berbeda. Pemilihan fitur ini memungkinkan perbedaan kriteria dalam suatu node dibandingkan dengan kriteria lainnya. Kefleksibelan dari metode pohon keputusan ini berkontribusi pada peningkatan kualitas keputusan yang dihasilkan, terutama jika dibandingkan dengan metode penghitungan satu tahap yang lebih konvensional. Dalam analisis multivarian dengan kriteria dan kelas yang jumlahnya sangat banyak, seorang pengguna biasanya perlu melakukan estimasi terhadap distribusi dimensi tinggi atau parameter tertentu dari distribusi kelas tersebut [20]. Dengan demikian, Decision Tree muncul sebagai alat yang efektif dalam menyederhanakan proses pengambilan keputusan dan meningkatkan kualitas analisis dalam konteks dataset yang kompleks.

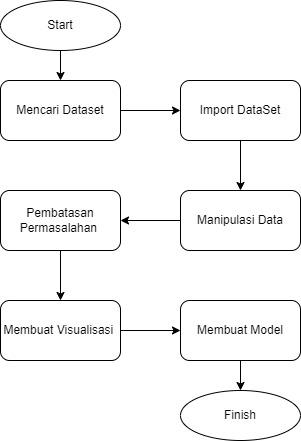
**2.2 Algoritma Support Vector Machine**

Algoritma Support Vector Machine (SVM) merupakan metode pembelajaran mesin yang memiliki tujuan utama, yaitu menemukan hyperplane maksimal, suatu fungsi matematis yang efektif memisahkan dua kelas dalam suatu dataset. Dalam prosesnya, SVM sangat memfokuskan diri pada maksimasi margin atau jarak optimal antara pola pelatihan dan batas keputusan [4].

Keunggulan utama dari SVM mencakup performa yang optimal, terlepas dari skala dataset, sehingga dapat memberikan hasil yang memuaskan baik pada dataset berskala kecil maupun besar. Selain itu, SVM terbukti sangat handal dalam menangani dataset dengan jumlah atribut yang signifikan, serta relatif mudah diimplementasikan [5]. SVM awalnya dikembangkan untuk melakukan klasifikasi biner, namun, sejalan dengan evolusi dan penelitian lebih lanjut, algoritma ini mengalami perluasan kemampuannya. Kini, SVM dapat digunakan untuk mengklasifikasi beberapa kelas sekaligus, menunjukkan fleksibilitas yang tinggi dalam tugas klasifikasi [5]. Kemampuan SVM tidak hanya terbatas pada klasifikasi; algoritma ini juga dapat diaplikasikan dalam konteks regresi dan deteksi outlier. Dengan demikian, SVM menyajikan alat analisis yang serbaguna dan dapat diandalkan untuk berbagai tugas di bidang pembelajaran mesin [5]. Ketangguhan SVM dalam menangani kompleksitas dataset dan memberikan hasil yang baik terutama pada dataset dengan dimensi yang tinggi membuatnya menjadi pilihan yang populer. Kefleksibelan algoritma ini memperluas ruang aplikasinya dari pemrosesan data klasifikasi hingga tugas analisis yang lebih kompleks, memastikan relevansinya dalam berbagai konteks [5]. Dengan demikian, SVM tidak hanya menjadi algoritma pembelajaran mesin yang efisien tetapi juga mendefinisikan kemajuan dalam pemodelan prediktif dan analisis data.

1. **METODOLOGI**

Metodologi penelitian ini mencakup serangkaian tahap yang sistematis untuk mencapai tujuan analisis data. Berikut gambar alur penelitian dibawah ini :



Gambar 1. Alur Penelitian

Alur penelitian ini dirancang untuk memastikan bahwa setiap tahap dilakukan dengan hati-hati dan sistematis, sehingga hasil analisis dapat memberikan wawasan yang berarti terkait permasalahan atau tujuan penelitian yang telah ditetapkan.

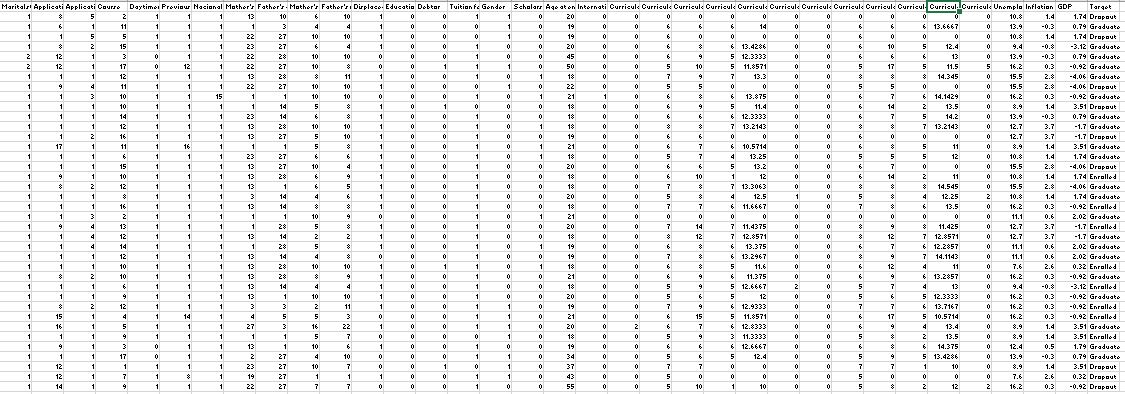
**3.1. Mencari DataSet**

Dalam tahap pertama penelitian ini, peneliti telah melakukan pencarian dataset yang sesuai dengan kebutuhan penelitian. Dataset yang ditemukan memiliki nama "Predict students' dropout and academic success" dan dapat diakses melalui link berikut:

<https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention/data>.

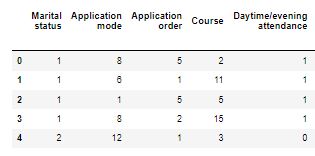
Dataset ini memberikan pandangan komprehensif mengenai mahasiswa yang terdaftar dalam berbagai program sarjana yang ditawarkan oleh lembaga pendidikan tinggi. Informasi yang tercakup melibatkan rincian demografis, indikator sosial-ekonomi, dan data kinerja akademis, memungkinkan eksplorasi faktor-faktor potensial yang mempengaruhi tingkat putus sekolah dan pencapaian akademis mahasiswa. Dataset ini terdiri dari basis data yang berbeda dengan informasi yang relevan yang dikumpulkan pada saat pendaftaran, termasuk mode aplikasi, status perkawinan, kursus yang dipilih, dan lainnya.

Selain itu, dataset memfasilitasi penilaian kinerja mahasiswa secara keseluruhan pada akhir setiap semester dengan mempertimbangkan jumlah unit kurikuler yang dikreditkan/diikuti/ dievaluasi/diakui serta nilai-nilai yang sesuai. Selain itu, dataset ini juga menggabungkan indikator ekonomi regional seperti tingkat pengangguran, tingkat inflasi, dan GDP, yang memberikan wawasan tentang bagaimana variabel ekonomi dapat berkontribusi pada variasi tingkat putus sekolah atau keberhasilan akademis mahasiswa.

Dengan menyajikan kolom-kolom terkait, seperti status perkawinan, mode aplikasi, urutan aplikasi, kursus, dan lainnya, dataset ini menjadi sumber informasi yang kaya untuk mengeksplorasi faktor-faktor potensial yang memengaruhi retensi mahasiswa. Kolom-kolom numerik seperti umur pada saat pendaftaran, jumlah unit kurikuler yang dikreditkan, diikuti, dievaluasi, dan memberikan dimensi analisis yang lebih terperinci.

Gambar 2. Cuplikan Isi DataSet

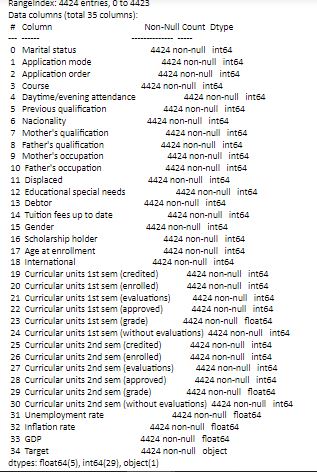
**3.2. Import Dataset**

Setelah menemukan dataset yang diperlukan, tahap berikutnya, peneliti melakukan *import* dataset ke dalam Jupyter Notebook. 

Gambar 3. Cuplikan Isi DataSet

**3.3. Manipulasi Data**

Setelah berhasil mengimpor dataset, langkah selanjutnya dalam penelitian ini melibatkan manipulasi data pada dataset. Pemeriksaan awal dilakukan untuk memastikan bahwa kolom yang akan digunakan tidak mengandung nilai null. Berikut adalah langkah-langkah yang diambil untuk memanipulasi data:



Gambar 4. Pemeriksaan awal dengan pemeriksaan pada kolom tabel

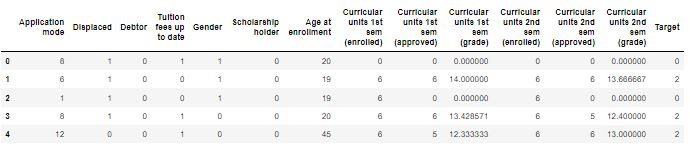
Peneliti melihat bahwa kolom Target masih berupa object dan setelah dilakukan identifikasi nilai unik pada target masih terdapat nilai null pada kolom tersebut maka peneliti perlu mengganti nilai yang tidak sesuai dengan 'Dropout' dengan 'Unknown' (atau nilai lain yang sesuai)



Gambar 5. Code Menangani Nilai Null(Tidak ada)

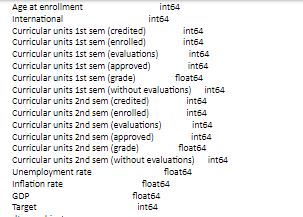


Gambar 6. Code Mengganti nilai yang tidak sesuai



Gambar 7. Kolom Tabel Nilai Yang sudah Diganti

Setelah melakukan penggantian nilai pada kolom 'Target', langkah ini bertujuan untuk mengonversi nilai kategorikal 'Dropout', 'Enrolled', dan 'Graduate' menjadi nilai numerik. Dengan pemetaan tersebut, nilai 'Dropout' diwakili oleh angka 0, 'Enrolled' diwakili oleh angka 1, dan 'Graduate' diwakili oleh angka 2. Proses ini memungkinkan penggunaan kolom 'Target' dalam analisis dan pemodelan yang melibatkan algoritma machine learning yang memerlukan input numerik. Dan pada data yang memiliki nilai object pada kolom target sudah tidak lagi nilai object melainkan sudah menjadi numerik untuk mengolah data selanjutnya

Gambar 8. Analisis statistik data 

**3.4. Pembatasan Permasalahan**

Setelah selesai memanipulasi data, peneliti mulai melakukan tahap selanjutnya, yaitu pembatasan permasalahan. Terdapat pemilihan kolom-kolom yang dianggap relevan sebagai input (fitur) dan output (target) dalam pengembangan model prediksi. Proses ini bertujuan untuk menyusun ke Data Frame baru yang hanya berisi kolom-kolom yang dianggap memiliki dampak signifikan terhadap variabel target seperti gambar berikut :



Gambar 9. Variabel Target

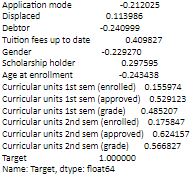
Pada pemilihan kolom-kolom untuk DataFrame student\_df, tampaknya ada sejumlah fitur yang dipilih sebagai input (fitur) dalam analisis. Berikut adalah penjelasan singkat untuk setiap fitur yang dipilih:

1. Application Mode: Menunjukkan metode aplikasi yang digunakan oleh mahasiswa ketika mendaftar.
2. Displaced: Menunjukkan apakah mahasiswa termasuk dalam kategori "Displaced" atau bukan.
3. Debtor: Menunjukkan apakah mahasiswa merupakan peminjam atau tidak.
4. Tuition Fees Up to Date: Menunjukkan apakah biaya kuliah mahasiswa sudah dibayar hingga saat ini.
5. Gender: Menunjukkan jenis kelamin mahasiswa.
6. Scholarship Holder: Menunjukkan apakah mahasiswa adalah penerima beasiswa atau tidak.
7. Age at Enrollment: Menunjukkan usia mahasiswa pada saat pendaftaran.
8. Curricular Units 1st Sem (Enrolled): Jumlah unit kurikuler yang diambil pada semester pertama.
9. Curricular Units 1st Sem (Approved): Jumlah unit kurikuler yang disetujui pada semester pertama.
10. Curricular Units 1st Sem (Grade): Nilai rata-rata dari unit kurikuler pada semester pertama.
11. Curricular Units 2nd Sem (Enrolled): Jumlah unit kurikuler yang diambil pada semester kedua.
12. Curricular Units 2nd Sem (Approved): Jumlah unit kurikuler yang disetujui pada semester kedua.
13. Curricular Units 2nd Sem (Grade): Nilai rata-rata dari unit kurikuler pada semester kedua.
14. Target: Variabel target yang menunjukkan status mahasiswa, dapat berupa "Dropout", "Enrolled", atau "Graduate".

DataFrame student\_df ini mencakup fitur-fitur yang akan digunakan dalam tahap pengembangan model prediksi untuk memahami faktor-faktor yang memengaruhi dari status mahasiswa-nya.



Gambar 10. Filter Data Jumlah putus sekolah, terdaftar & lulusan yang ada di kolom Target

Pada gambar diatas Metode ini digunakan untuk menghitung jumlah kemunculan setiap nilai unik dalam kolom tersebut. Hasilnya kemudian ditampilkan dalam bentuk Data Frame dengan dua kolom: satu kolom berisi nilai unik dari 'Target' dan kolom kedua berisi jumlah kemunculan setiap nilai tersebut. Dari hasil tersebut, terdapat tiga nilai unik dalam kolom 'Target'. 

Gambar 11. korelasi antara variabel dependen (target) dengan variabel independen(fitur atau atribut lainnya)

Seperti gambar diatas Peneliti melihat korelasi antara kolom 'Target' (variabel dependen atau target) dengan variabel independen (fitur atau atribut lainnya) dalam DataFrame student\_df. Nilai korelasi ini berkisar dari -1 hingga 1, dengan interpretasi sebagai berikut:

1. Nilai 1: Korelasi sempurna (positif).
2. Nilai -1: Korelasi sempurna (negatif).
3. Nilai 0: Tidak ada korelasi.

Sebagai contoh, jika nilai korelasi antara 'Target' dan suatu fitur adalah 0.5, ini menunjukkan bahwa ada korelasi positif moderat antara fitur tersebut dengan 'Target'. Sebaliknya, jika nilai korelasinya adalah -0.5, ini menunjukkan korelasi negatif moderat. Dalam hasil tersebut, kita lihat bahwa kolom 'Target' memiliki korelasi positif yang kuat dengan beberapa fitur tertentu:

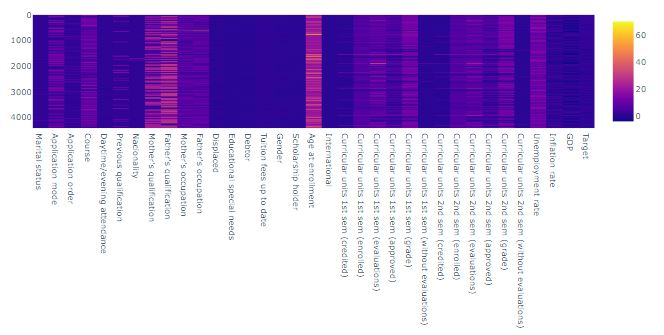
1. ‘Tuition fees up to date' (0.409827): Korelasi positif moderat.
2. ‘Scholarship holder' (0.297595): Korelasi positif moderat.
3. 'Curricular units 1st sem (approved)' (0.529123): Korelasi positif kuat.
4. 'Curricular units 1st sem (grade)' (0.485207): Korelasi positif kuat.
5. 'Curricular units 2nd sem (approved)' (0.624157): Korelasi positif kuat.
6. 'Curricular units 2nd sem (grade)' (0.566827): Korelasi positif kuat.

Ini menunjukkan bahwa semakin tinggi nilai fitur-fitur ini, semakin tinggi kemungkinan mahasiswa lulus ('Target' dengan nilai 2). Sebaliknya, beberapa fitur memiliki korelasi negatif dengan 'Target', seperti 'Application mode', 'Gender', dan 'Age at enrollment', meskipun nilainya tidak terlalu kuat. Korelasi negatif menunjukkan bahwa semakin rendah nilai

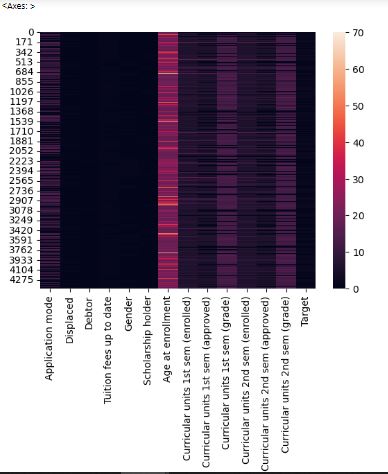
fitur-fitur ini, semakin tinggi kemungkinan mahasiswa putus sekolah ('Target' dengan nilai 0).

**3.5. Membuat Visualisasi**

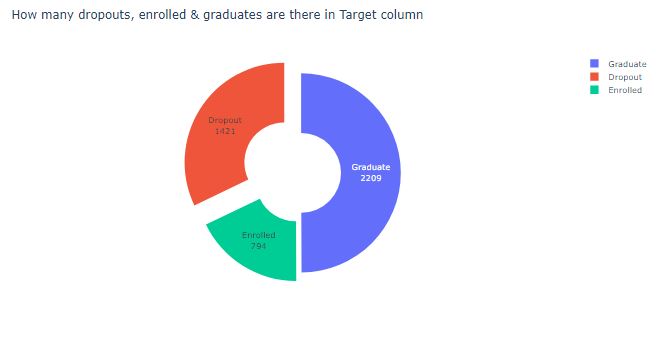
Dalam proses manipulasi data dan pembatasan permasalahan sebelumnya, terdapat beberapa Dataset yang digunakan untuk membuat visualisasi. Dataset tersebut adalah sebagai berikut:

1. Grafik heatmap (dihasilkan oleh px.imshow(student)) dapat membantu peneliti dalam visualisasi korelasi antara kolom-kolom dalam bentuk warna seperti gambar dibawah berikut : 

Gambar 12. Grafik heatmap Data student

1. Berikut grafik DataFrame student\_df untuk membuat heatmap (peta panas) dari korelasi antar-kolom dalam DataFrame :

Gambar 13. Grafik heatmap DataFrame student\_df

1. Grafik dibawah berikut menggunakan plot pie dengan menggunakan library Plotly Express (px.pie) untuk memvisualisasikan proporsi setiap kategori. Plot tersebut menunjukkan seberapa besar bagian dari keseluruhan populasi mahasiswa yang masuk ke dalam masing-masing kategori : 

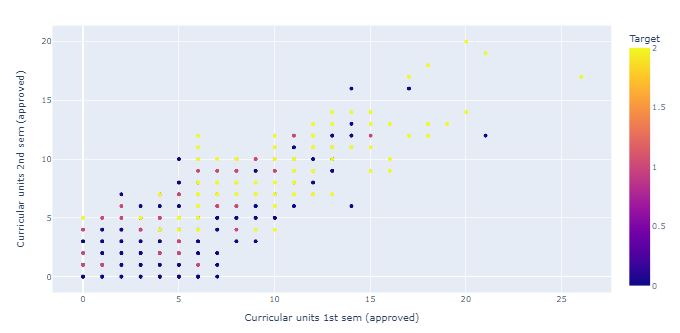
Gambar 14. Grafik Plot jumlah mahasiswa untuk setiap kategori Graduate, Dropout, Enrolled

Keterangan pada gambar tersebut peneliti menemukan Jumlah mahasiswa untuk setiap kategori:

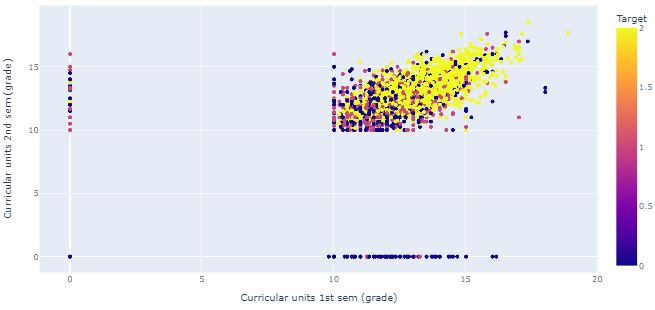
* Lulusan (Graduate): 2,209 mahasiswa
* Putus Sekolah (Dropout): 1,421 mahasiswa
* Terdaftar (Enrolled): 794 mahasiswa

Pada kolom Graduate Diperlihatkan dengan warna biru, menyusun sebagian besar dari keseluruhan populasi. Kolom Dropout: Diperlihatkan dengan warna orange, memiliki proporsi yang signifikan. Lalu kolom Enrolled Diperlihatkan dengan warna hijau, menyusun proporsi yang lebih kecil dibandingkan kedua kategori lainnya.

1. Gambar plot dibawah ini tersebut adalah scatter plot yang menunjukkan hubungan antara jumlah mata kuliah yang disetujui pada semester pertama dengan jumlah mata kuliah yang disetujui pada semester kedua. Setiap titik (data point) dalam scatter plot merepresentasikan seorang mahasiswa. Warna titik-titik tersebut dibedakan berdasarkan kategori dalam kolom 'Target', yaitu Graduate (Lulusan), Dropout (Putus Sekolah), dan Enrolled (Terdaftar). Scatter plot ini membantu melihat pola dan distribusi data terkait jumlah mata kuliah yang disetujui pada kedua semester untuk setiap kategori mahasiswa, Dengan melihat pola distribusi titik-titik berdasarkan warna, peneliti dapat mencari tahu apakah terdapat pola yang dapat membedakan mahasiswa yang lulus, putus sekolah, atau masih terdaftar berdasarkan jumlah mata kuliah yang disetujui pada kedua semester. Berikut gambar visualisasi-nya:

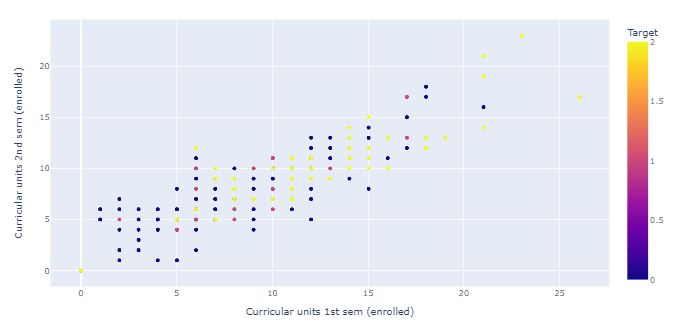


Gambar 15. Scatter plot pola dan distribusi data terkait jumlah mata kuliah yang disetujui pada kedua semester untuk setiap kategori mahasiswa.

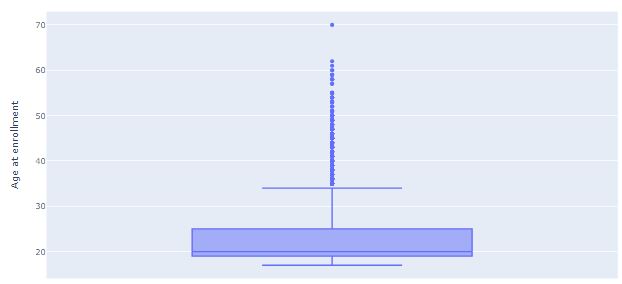
1. Selanjutnya, Scatter plot di bawah ini adalah representasi visual dari hubungan antara nilai mata kuliah pada semester pertama (sumbu x) dan nilai mata kuliah pada semester kedua (sumbu y) untuk setiap mahasiswa dalam dataset. Setiap titik dalam scatter plot mewakili seorang mahasiswa, dan warna titik-titik tersebut membedakan antara mahasiswa yang lulus (Graduate), putus sekolah (Dropout), dan masih terdaftar (Enrolled). Berikut gambarnya:

Gambar 16. Scatter Plot Representasi visual dari hubungan antara nilai mata kuliah pada semester pertama (sumbu x) dan nilai mata kuliah pada semester kedua (sumbu y) untuk setiap mahasiswa.

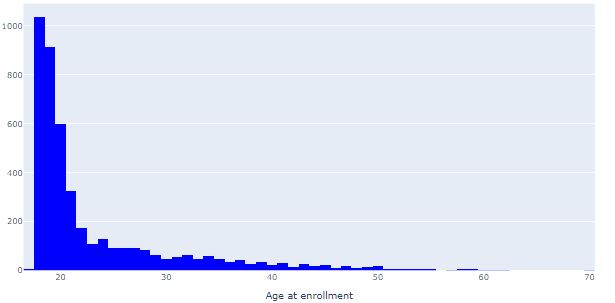
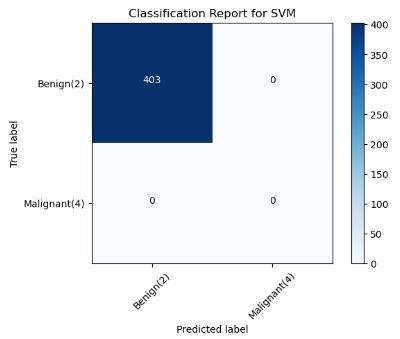
1. Dan dibawah ini merupakan Scatter plot di atas memvisualisasikan hubungan antara jumlah mata kuliah yang diambil pada semester pertama (sumbu x) dengan jumlah mata kuliah yang diambil pada semester kedua (sumbu y) untuk setiap mahasiswa dalam dataset. Setiap titik dalam scatter plot mewakili seorang mahasiswa, dan warna titik-titik tersebut membedakan antara mahasiswa yang lulus (Graduate), putus sekolah (Dropout), dan masih terdaftar (Enrolled).



Gambar 17. Scatter Plot Enrolled (Pendaftar) dari hubungan antara nilai mata kuliah pada semester pertama (sumbu x) dan nilai mata kuliah pada semester kedua (sumbu y) untuk setiap mahasiswa.

1. Selanjutnya peneliti membuat Box plot di bawah ini menggunakan Plotly Express untuk memvisualisasikan distribusi umur mahasiswa saat pendaftaran (Age at enrollment). Box plot ini memberikan informasi mengenai median, kuartil, dan adanya potensial outlier pada distribusi umur mahasiswa. Pada box plot tersebut, kita dapat melihat sebaran umur pada saat pendaftaran untuk seluruh mahasiswa dalam dataset. Median atau garis tengah kotak menunjukkan nilai tengah dari distribusi, sedangkan kotak menggambarkan kuartil pertama hingga kuartil ketiga. Adanya potensial outlier ditandai dengan titik-titik di luar batas-batas whisker. Berikut gambar dibawah ini :

Gambar 18. Box Plot distribusi umur mahasiswa saat pendaftaran

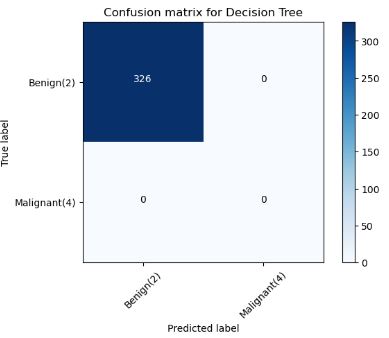
Dan Dibawah ini merupakan Plotly Histogram yang dibuat oleh peneliti supaya gambar menjadi lebih interaktif

Gambar 19. Plotly Histogram distribusi umur mahasiswa saat pendaftaran

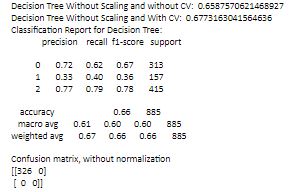
**3.6. Membuat Model**

Setelah melakukan analisis eksplorasi data (EDA) dan mempersiapkan dataset, peneliti melangkah ke tahap pembuatan model. Dalam tahap ini, dua model berbeda diterapkan untuk mengelompokkan data, yaitu Algoritma Pohon Keputusan (Decision Tree) dan Mesin Vektor Pendukung (SVM). Pertama, menggunakan model Decision Tree, peneliti mentraining model tanpa melakukan penskalaan pada fitur-fitur dataset. Hasil akurasi model diukur menggunakan metode 10-fold cross-validation. Confusion matrix, yang merepresentasikan hasil prediksi, dievaluasi untuk melihat performa model dalam mengklasifikasikan kategori target. Selanjutnya, peneliti menerapkan model SVM dengan menggunakan Linear Support Vector Classifier (LinearSVC). Sebagai langkah awal, data numerik dinormalisasi menggunakan MinMaxScaler. Model SVM dilatih dan diuji tanpa melakukan penskalaan serta dengan validasi silang menggunakan metode 10-fold cross-validation. Confusion matrix dan classification report digunakan untuk menganalisis performa model SVM. Kedua model tersebut memungkinkan peneliti untuk mendapatkan wawasan yang mendalam mengenai kemampuan klasifikasi pada dataset yang telah disiapkan. Grafik-grafik confusion matrix juga disertakan untuk mempermudah pemahaman hasil dari masing-masing model.

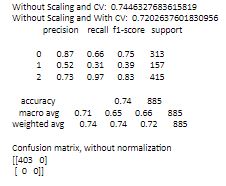
**IV. HASIL DAN PEMBAHASAN**

 Berikut merupakan hasil visualisasi Klasifikasi dari kedua model yang telah dibuat:

Gambar 20. confusion matrix Perbandingan kedua model

Dalam penelitian ini, dua algoritma klasifikasi, yaitu Decision Trees dan Support Vector Machine (SVM), dievaluasi untuk memprediksi keberhasilan akademis dan drop-out mahasiswa. Hasil eksperimen menunjukkan bahwa SVM memberikan performa yang lebih baik dibandingkan Decision Trees dalam konteks ini. Tanpa penskalaan, SVM mencapai akurasi total sekitar 74%, sedangkan Decision Trees hanya sekitar 66%. Evaluasi dengan validasi silang menunjukkan hasil yang lebih stabil, dengan SVM mencapai skor rata-rata sekitar 72%, sedangkan Decision Trees sekitar 68%. Analisis lebih lanjut terhadap kinerja klasifikasi menunjukkan bahwa SVM menghasilkan precision, recall, dan F1-score yang lebih seimbang untuk setiap kelas, sedangkan Decision Trees memiliki variasi yang lebih besar antar kelas. Matriks kebingungan SVM menunjukkan kemampuan yang baik untuk memprediksi baik keberhasilan maupun drop-out mahasiswa, sementara Decision Trees memiliki kinerja rendah terutama pada kelas drop-out. Dibawah ini hasil dari evaluasi tersebut :

Gambar 21. Hasil Evaluasi Model Decision Tree



Gambar 22. Hasil Evaluasi Model SVM

**V. KESIMPULAN**

SVM lebih cocok untuk memodelkan masalah ini dengan akurasi dan keseimbangan yang lebih baik antara kelas-kelas yang dihasilkan. Validasi silang memberikan pandangan yang lebih luas terkait generalisasi model, dan hasil eksperimen ini dapat membantu dalam pemilihan model untuk memprediksi keberhasilan akademis dan potensi drop-out mahasiswa. Setelah itu mengintegrasikan kekuatan masing-masing algoritma dapat memperbaiki kinerja model secara keseluruhan, sementara pendekatan gabungan memiliki potensi untuk mengatasi kekurangan satu algoritma dengan memanfaatkan kelebihan algoritma lainnya dan analisis yang lebih komprehensif memberikan kesempatan untuk mendapatkan wawasan yang lebih mendalam dibandingkan dengan mengevaluasi setiap algoritma secara individu.

REFERENCES

[1] M. P. Agustini, A. A. Supianto, and W. Purnomo, “Aplikasi DataMining untuk Memprediksi Mahasiswa Berpotensi Drop Out menggunakan Algoritme K-Nearest Neighbor (K-NN),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 6, pp. 6230–6237, Jun. 2019, doi: e-ISSN: 2548-964X.

[2] A. I. Sang, E. Sutoyo, and I. Darmawan, “Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine,” *e-Proceeding of Engineering*, vol. 8, no. 5, pp. 8954–8963, Oct. 2021, doi: ISSN : 2355-9365.

[3] M. Maulidah, Windu Gata, Rizki Aulianita, and Cucu Ika Agustyaningrum, “ALGORITMA KLASIFIKASI DECISION TREE UNTUK REKOMENDASI BUKU BERDASARKAN KATEGORI BUKU,” E-Bisnis : Jurnal Ilmiah Ekonomi dan Bisnis, vol. 13, no. 2, pp. 89–96, Dec. 2020, doi: 10.51903/e-bisnis.v13i2.251.

[4] N. Fitriyah, B. Warsito, and D. A. I Maruddani, “ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM) ,” *JURNAL GAUSSIAN,* , vol. 9, no. 3, pp. 376–390, Aug. 2020, doi: ISSN: 2339-2541.

[5] F. Abdusyukur, “PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI PENCEMARAN NAMA BAIK DI MEDIA SOSIAL TWITTER,” *Komputa : Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, pp. 73–82, May 2023, doi: 10.34010/komputa.v12i1.9418.

[19] I. Sutoyo, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik,” J. Pilar Nusa

Mandiri, vol. 14, no. 2, p. 217, 2018, doi: 10.33480/pilar.v14i2.926.

[20] E. T. Susdarwono and A. Setiawan, “PENERAPAN TEORI KEPUTUSAN DALAM MODEL

PENGAMBILAN KEPUTUSAN TERKAIT MASALAH EKONOMI PERTAHANANKONSEP

POHON KEPUTUSAN,” vol. 11, no. November, pp. 243–257, 2020.