

Penggunaan Algoritma *Support Vector Machine* Untuk Penentuan Rekomendasi Penerima Beasiswa Di SMA Negeri 8 Kota Bogor

¹ Muhammad Rafly, ² Ionia Veritawati

^{1,2}Program Studi Teknik Informatika, Universitas Pancasila, Jakarta

E-mail: ¹rafly.muhammad72@gmail.com, ²ionia.veritawati@univpancasila.ac.id

ABSTRAK

Beasiswa adalah program bantuan berupa pendanaan pendidikan yang berupa diberikan untuk membantu biaya siswa dalam mendapatkan pendidikan yang layak. SMA Negeri 8 Kota Bogor sebagai satuan pendidikan yang menyelenggarakan pendidikan formal, turut serta dalam penyediaan beasiswa bagi siswanya. Program beasiswa di SMA Negeri 8 Kota Bogor belum menggunakan metode yang efektif dan efisien dalam menentukan kelayakan siswa dalam menerima beasiswa sehingga sering terjadi kesalahan seperti beasiswa kurang tepat sasaran. Untuk mengatasi hal tersebut, dibuat sebuah aplikasi prediksi beasiswa berbasis website guna memprediksi status kelayakan siswa menerima beasiswa. Prediksi dilakukan menggunakan metode klasifikasi dengan model *Support Vector Machine*. Hasilnya pada komposisi *data training* 70% dan *data testing* 30% mendapatkan hasil evaluasi tertinggi dengan *Accuracy* 89%, *Precision* 91%, *Recall* 95%.

Kata kunci : beasiswa, prediksi, *Support Vector Machine*

ABSTRACT

Scholarship is an assistance program in the form of educational funding in the form of being given to assist students in obtaining proper education. SMA Negeri 8 Bogor City as an educational unit that organizes formal education, participates in providing scholarships for its students. The scholarship program at SMA Negeri 8 Bogor City has not used an effective and efficient method in determining the eligibility of students to receive scholarships so that errors often occur such as scholarships that are not on target. To overcome this, a website-based scholarship prediction application was created to predict the eligibility status of students to receive scholarships. Prediction is done using the classification method with the Support Vector Machine model. The results in the composition of 70% training data and 30% testing data get the highest evaluation results with Accuracy 89%, Precision 91%, Recall 95%.

Keyword : prediction, scholarship, *Support Vector Machine*

1. PENDAHULUAN

Pendidikan adalah suatu proses pembelajaran yang merupakan kebutuhan primer yang harus dipenuhi, dengan pendidikan, suatu negara akan dapat maju dan meningkat dengan pesat karena pendidikan merupakan tonggak kemajuan

suatu bangsa. pesat karena pendidikan merupakan tonggak kemajuan suatu bangsa. Tujuan pendidikan adalah mengembangkan kemampuan dan membentuk watak serta peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa, bertujuan untuk berkembangnya potensi peserta didik agar menjadi manusia yang

beriman dan bertaqwa kepada Tuhan Yang Maha Esa, berakhlak mulia, sehat, berilmu, cakap, dan kompeten. (Razak, & Wiguna, 2022). Hal ini tertuang dalam Undang-Undang Dasar 1945 Bab XIII tentang Pendidikan dan Kebudayaan pasal 31, yaitu “setiap warga negara berhak mendapat pendidikan”.

Hal ini mengindikasikan bahwa setiap warga negara usia sekolah, dari tingkat dasar hingga perguruan tinggi, wajib mendapatkan pendidikan. Hal ini mengindikasikan bahwa setiap warga negara usia sekolah dari tingkat dasar hingga perguruan tinggi wajib mengenyam pendidikan. Pada kenyataannya, banyak warga negara usia sekolah yang tidak mengenyam pendidikan, terutama pendidikan tinggi. Mahalnya biaya pendidikan menjadi masalah yang dihadapi oleh masyarakat yang berada di bawah garis kemiskinan, hal ini disebabkan oleh pendapatan mereka yang tidak mencukupi kebutuhan hidup.

Oleh karena itu, masalah ekonomi juga menjadi faktor utama dalam masalah pendidikan di Indonesia. Beasiswa merupakan dana yang dapat diberikan oleh pemerintah, perusahaan swasta, kedutaan besar, universitas atau dari tempat kerja untuk meningkatkan kualitas sumber daya manusia melalui pendidikan (Hidayatulloh, Suhada, & Yusuf, 2018). Beasiswa dapat diterima oleh siswa dari tingkat sekolah dasar, sekolah menengah pertama, sekolah menengah atas, dan perguruan tinggi. Ada berbagai jenis beasiswa yang tersedia seperti beasiswa prestasi yang diberikan kepada siswa yang memiliki prestasi akademik, beasiswa atletik yang diberikan kepada siswa yang memiliki prestasi atletik dan beasiswa bantuan yang diberikan kepada siswa yang mengalami kesulitan dalam biaya pendidikan.

Untuk mengimplementasikan hal tersebut di atas, pemerintah telah mengupayakan berbagai program, salah

satunya adalah program beasiswa di berbagai jenjang pendidikan. Begitu juga pada pendidikan sekolah menengah atas, pemerintah memberikan beasiswa melalui Program Indonesia Pintar (PIP) (Gunawan, Oktavia, & Borman, 2018). Di Indonesia, pendidikan telah termaktub dalam Undang-Undang Dasar dimana Negara menjamin hak setiap warga negara untuk mendapatkan pendidikan. Namun dalam pelaksanaannya, terdapat banyak kendala seperti kualitas pendidikan yang relatif rendah, fasilitas yang kurang memadai dan pemerataan pendidikan karena kesenjangan ekonomi.

Oleh karena itu, pemerintah mengeluarkan kebijakan terkait hal tersebut berupa Program Indonesia Pintar yang diharapkan dapat membantu anak-anak untuk tetap melanjutkan pendidikan (Mutiah, & Sari, 2023). Sesuai dengan Peraturan Sekretaris Jenderal Kementerian Pendidikan dan Kebudayaan No. 8 Tahun 2020 tentang Petunjuk Pelaksanaan PIP, menyatakan bahwa beasiswa ini diberikan kepada peserta didik dari jenjang pendidikan dasar sampai dengan jenjang pendidikan menengah untuk mendapatkan layanan pendidikan sampai tamat dari satuan pendidikan yang ditempuh sesuai dengan usia peserta didik, yaitu 6 tahun sampai dengan 21 tahun.

SMA Negeri 8 Kota Bogor turut ambil bagian dalam penyaluran program beasiswa PIP. Selama ini sekolah menentukan penerima beasiswa dengan cara menganalisis secara manual aspek-aspek yang menjadi pertimbangan dalam menentukan pemberian beasiswa. Karena belum adanya metode yang pasti dalam proses seleksi calon penerima beasiswa, hal ini menyebabkan terjadinya kesalahan sasaran dan ketidaktepatan sasaran penyaluran beasiswa. Pemberian beasiswa yang tidak tepat sasaran mengurangi tingkat kepercayaan pemohon terhadap pemberi beasiswa. Muhammad Abdul Karim melakukan penelitian tentang kelayakan siswa untuk

mendapatkan beasiswa dengan menggunakan model *Support Vector Machine* dan *K-nearest Neighbors*. Penelitian tersebut menganalisis data seperti indeks prestasi kumulatif, status kemiskinan, status yatim piatu, pendapatan orang tua, dan kelayakan beasiswa. Penelitian tersebut menemukan bahwa kedua model memiliki kinerja terbaik dengan komposisi data pelatihan sebesar 70% dan data pengujian sebesar 30%. Algoritma *K-nearest Neighbors* memiliki nilai K optimal sebesar 3. Hasil ini menunjukkan bahwa kedua model tersebut efektif dalam menentukan kelayakan siswa untuk mendapatkan beasiswa (Karim, 2023). Peneliti lainnya, Apandi Sugianto, mengembangkan sistem klasifikasi untuk memprediksi kepuasan terhadap pelayanan perekaman e-KTP dengan menggunakan algoritma *Naïve Bayes* dan *Decision Tree*. Penelitian tersebut menemukan bahwa algoritma *Naïve Bayes* lebih akurat, dengan tingkat akurasi 91,70% dan nilai uji *f-measure* 93,92%, dibandingkan dengan algoritma *Decision Tree* yang memiliki tingkat akurasi 65,90% dan nilai uji *f-measure* 79,26%. Algoritma *Naïve Bayes* juga memiliki waktu eksekusi yang lebih cepat yaitu 0 detik, sedangkan algoritma *Decision Tree* membutuhkan waktu 57 detik (Apandi, & Sugianto, 2019).

Penelitian ini mengeksplorasi penggunaan *Python*, khususnya algoritma *Support Vector Machine*, untuk memprediksi status penerimaan siswa untuk mendapatkan beasiswa di SMA Negeri 8 Kota Bogor. *Python* dipilih karena memiliki library dan framework yang luas sehingga memudahkan pengembangan aplikasi yang berpusat pada data, termasuk klasifikasi data. Penelitian ini menghasilkan sebuah aplikasi berbasis web yang dapat mengkategorikan secara akurat siswa yang berhak mendapatkan beasiswa di sekolah tersebut. Aplikasi ini bertujuan untuk menyediakan alat yang berguna

bagi penyedia beasiswa untuk mengambil keputusan yang tepat ketika memilih siswa yang akan menerima beasiswa. Dengan menggunakan aplikasi ini, keputusan dapat dibenarkan dan beasiswa dapat ditargetkan secara lebih efektif kepada siswa yang paling layak.

2. LANDASAN TEORI

Prediksi

Prediksi, juga dikenal sebagai peramalan, melibatkan pembuatan tebakan atau prediksi tentang kejadian di masa depan (Ramadhani, & Ardiansyah, 2021). Dalam *machine learning*, prediksi dibuat dengan menggunakan model terlatih yang dapat mengidentifikasi pola atau hubungan antara variabel *input* dan *output*. Ada tiga jenis peramalan: peramalan ekonomi, teknologi, dan permintaan. Peramalan yang baik melibatkan serangkaian langkah persiapan atau prosedur, seperti menentukan tujuan, memilih model peramalan, mendapatkan data, memvalidasi model, membuat ramalan, mengimplementasikan hasil, dan memantau keandalannya. Analisis deret waktu, yang berfokus pada hubungan antara variabel dependen dan independen dari waktu ke waktu, melibatkan metode seperti *smoothing*, *Box Jenkins*, dan proyeksi *trend* dengan regresi. *Smoothing* mengurangi ketidakteraturan pada data masa lalu, sementara *Box Jenkins* menggunakan model matematika untuk peramalan jangka pendek. Metode-metode ini digunakan untuk membuat prediksi yang akurat di berbagai industri

Beasiswa

Beasiswa adalah bentuk bantuan keuangan yang tidak berasal dari dana pribadi atau keluarga. Beasiswa dapat diberikan untuk studi di universitas atau institusi pendidikan baik di dalam maupun luar negeri (Gafur, Yulianti, & Hidayat, 2008). Beasiswa biasanya

mencakup biaya pendidikan untuk siswa yang aktif menghadiri kuliah di universitas. Beasiswa juga dapat dilihat sebagai dana mahasiswa atau kewajiban kepada masyarakat. Beasiswa disediakan untuk mendukung individu, terutama mereka yang masih bersekolah atau kuliah, dalam menyelesaikan tujuan pendidikan mereka. Tujuannya adalah agar mereka dapat mencari pengetahuan dan berhasil menyelesaikan pendidikan mereka. Beasiswa dapat diberikan oleh lembaga pemerintah, perusahaan, atau yayasan. Beasiswa dapat dikategorikan sebagai hadiah gratis atau hadiah dengan kewajiban bekerja, yang dikenal sebagai ikatan dinas, setelah menyelesaikan pendidikan. Lamanya ikatan dinas bervariasi tergantung pada penyedia beasiswa.

Data Science

Data Science adalah bidang multidisiplin yang bertujuan untuk mendapatkan nilai dari data melalui berbagai proses seperti ekstraksi, persiapan, eksplorasi, transformasi, dan presentasi (Haris, Nurlaila, Ratnadewi, & John, 2019). Ilmu ini melibatkan teknik-teknik seperti *machine learning*, *deep learning*, *computer vision*, dan *natural language processing* untuk menganalisis dan mengklasifikasikan data dalam jumlah besar. Dengan menggunakan algoritma pembelajaran ansambel, prediksi dari berbagai sensor dapat digabungkan untuk meminimalkan kesalahan pemilihan model. Siklus hidup *data science* mengikuti tahapan seperti konsumsi data di mana data dikumpulkan dari berbagai sumber, termasuk data terstruktur dan tidak terstruktur. Penyimpanan dan pemrosesan data melibatkan pemilihan sistem penyimpanan yang sesuai untuk berbagai jenis data, serta pembersihan, transformasi, dan penggabungan data. Tahapan-tahapan ini memungkinkan para analis untuk mendapatkan wawasan yang dapat ditindaklanjuti dan memberikan

dasar untuk pengambilan keputusan berdasarkan data.

Support Vector Machine

Support Vector Machine (SVM) adalah metode berbasis statistik yang digunakan dalam *machine learning* yang memiliki landasan teori yang jelas dan tidak dianggap sebagai *black box*. Mengimplementasikan SVM relatif mudah karena proses penentuan *Support Vector* dapat dibingkai sebagai *Quadratic Programming* (QP) *problem* sehingga lebih mudah untuk diotomatisasi jika ada library yang tersedia untuk menyelesaikan masalah tersebut (Yahya, 2022). SVM memiliki kemampuan generalisasi yang baik dengan ukuran sampel yang kecil, namun kesulitan dalam memilih parameter yang paling tepat, yang dapat mempengaruhi kinerjanya. Menemukan nilai parameter yang tepat, seperti koefisien *soft-margin C* dan parameter fungsi kernel, sangat penting untuk meningkatkan akurasi SVM. SVM bekerja dengan mengidentifikasi *hyperplane* dengan margin terbesar yang memisahkan data antar kelas. *Support Vectors*, yang merupakan titik data terdekat dengan *hyperplane* di setiap kelas, memainkan peran penting dalam proses klasifikasi. SVM memiliki prinsip utama untuk menemukan *hyperplane* terbaik yang berfungsi untuk memisahkan dua kelas dalam ruang *input*. *Hyperplane* dapat berupa garis dalam dua dimensi dan dapat berupa bidang datar dalam beberapa bidang. Menemukan lokasi *hyperplane* yang optimal adalah inti dari metode SVM. Diasumsikan terdapat data learning dengan titik data x_i ($i=1,2,\dots,m$) memiliki dua kelas $y_i = \pm 1$, yaitu kelas positif (+1) dan negatif (-1) sehingga akan diperoleh fungsi keputusan seperti pada Persamaan (1).

$$f(x) = \text{sign}(w \cdot x + b) \quad (1)$$

Dengan (\cdot) merupakan skalar sehingga pada Persamaan (2)

$$w \cdot x = w^t x \quad (2)$$

Keterangan:

w = bobot *hyperplane*

x = *vector* data masukan

b = bidang relative

$$w \cdot x_i + b = 0, \text{ untuk semua } i = 1, 2, 3, \dots, n \quad (3)$$

Keterangan :

w = *vector* bobot

x_i = data *input*

b = nilai bias

n = banyak data

Untuk dapat memenuhi Persamaan (3), nilai w dan b dapat diketahui dengan Persamaan (4), di mana data dua kelas dapat bernilai +1 atau -1.

$$y_1((w_1 \cdot x_1) + (w_2 \cdot x_2) + \dots + (w_i \cdot x_i) + b) \geq 1 \quad (4)$$

Keterangan:

y_1 = kelas dari sampel data ke- i

w_i = *vector* bobot ke- i

x_i = data *input* fitur

Normalisasi

Normalisasi dalam konteks *machine learning* mengacu pada proses mengubah nilai fitur ke dalam skala tertentu, biasanya antara 0 dan 1, tanpa mengubah distribusi data. Hal ini dilakukan untuk menangani variasi dalam rentang nilai yang ada dalam data, karena perbedaan yang signifikan dalam rentang tersebut dapat berdampak pada kinerja klasifikasi algoritma. Metode normalisasi yang umum digunakan adalah *Min-Max Scaling*, *Standard Scaling*, dan *Robust Scaling*. *Min-max Scaling*, misalnya, menggunakan nilai minimum dan maksimum dari setiap atribut untuk mengubah nilai dalam rentang 0-1. Pendekatan ini membantu menjaga perbandingan nilai antara variabel *input*

dan menghindari ketidakseimbangan skala dalam data. Rumus yang digunakan dalam *Min-Max Scaling* adalah $(x - \min) / (\max - \min)$. Normalisasi sangat penting dalam *machine learning* karena memastikan bahwa semua fitur memiliki bobot yang sama ketika digunakan dalam algoritma (Rumah Coding, 2024).

K-Fold Cross Validation

Cross Validation merupakan metode yang digunakan dalam penerapan *machine learning* untuk mengevaluasi kemampuan model *machine learning* terhadap data yang digunakan dengan cara membagi data menjadi beberapa subset. Jenis uji *Cross Validation* yang digunakan pada penelitian ini adalah *K-Fold Cross Validation*. *K-Fold Cross Validation* berfungsi untuk mengukur performa model algoritma yang dibangun dengan cara membagi data set secara acak dan mengelompokkan data sebanyak nilai K pada *K-Fold* (Arisandi, Warsito, & Hakim, 2022). Pada skema metode *K-Fold Cross Validation*, data set dibagi menjadi beberapa subset yang lebih kecil secara acak. Subset data diproses sebanyak K kali dengan setiap percobaan menggunakan subset ke- K sebagai *data testing* dan menggunakan subset sisanya sebagai *data training*.

Confusion Matrix

Confusion Matrix adalah metode pengukuran dalam bentuk tabel yang digunakan untuk mengevaluasi performa model *machine learning* yang dibuat dengan membandingkan hasil prediksi model dengan nilai aktualnya (Al Kabir, Basuki, & Wicaksono, 2019). Terdapat empat komponen dalam *confusion matrix*, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). *True Positive* (TP) adalah jumlah data yang diprediksi positif dengan kondisi aktualnya juga positif. *True Negative* (TN) adalah jumlah data yang diprediksi bernilai

negatif dengan kondisi aktualnya juga negatif. *False Positive* (FP) adalah jumlah data yang diprediksi positif namun kondisi aktualnya negatif dan *False Negative* (FN) adalah jumlah data yang diprediksi negatif namun kondisi aktualnya positif. Hasil dari *confusion matrix* tersebut digunakan dalam perhitungan tiga matriks akurasi, yaitu *accuracy*, *precision*, dan *recall*. Ketiga matriks ini kemudian digunakan untuk mengukur performa model *machine learning* yang dibangun. Tabel *confusion matrix* digambarkan pada tabel 1.

Tabel 1 *confusion matrix*

		Aktual	
		Positif	Negatif
Prediksi	Positif	TP	FP
	Negatif	FN	TN

a. *Accuracy* adalah pengukuran kedekatan antara nilai prediksi dengan nilai aktual. Nilai akurasi diketahui dengan menghitung jumlah data yang diprediksi dengan benar dari seluruh prediksi model. Persamaan 5 merupakan persamaan untuk menentukan nilai akurasi.

$$Accuracy = \frac{TN+TP}{TN+FP+TP+FN} \quad (5)$$

b. *Precision* merupakan hasil perbandingan antara jumlah data yang diprediksi *True Positive* dengan jumlah keseluruhan data yang diprediksi positif. Persamaan 6 merupakan Persamaan untuk mengetahui nilai *precision*.

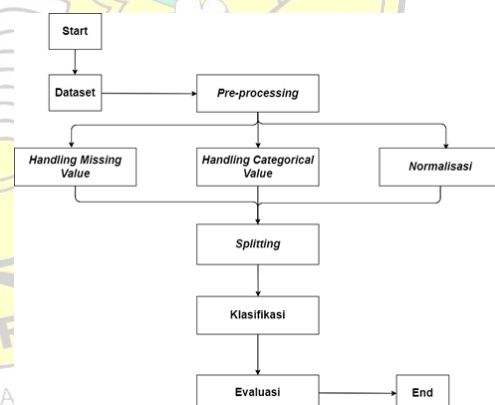
$$Precision = \frac{TP}{FP+TP} \quad (6)$$

c. *Recall* merupakan hasil perbandingan antara jumlah data yang diprediksi *True Positive* dengan jumlah keseluruhan data yang memiliki kondisi aktual positif. Persamaan 2.7 merupakan Persamaan untuk mengetahui nilai *recall*

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

3. METODOLOGI

Penelitian ini mengikuti alur kerja terstruktur yang mencakup sejumlah tahapan dan metode yang terorganisir. Gambar 1 merupakan alur penelitian klasifikasi prediksi berbasis menggunakan *Support Vector Machine*.



Gambar 1. Alur Penelitian

Dataset

Kumpulan data adalah kumpulan data terstruktur yang dapat mencakup berbagai jenis informasi, seperti angka, teks, dan gambar. Kumpulan data ini dikumpulkan oleh para profesional dari berbagai sumber dan diproses melalui tahap-tahap seperti pembersihan dan kategorisasi data. Kumpulan data dikategorikan berdasarkan variabelnya dan digunakan di berbagai industri dan disiplin ilmu. Dalam konteks penelitian tentang prediksi status penerimaan

beasiswa, tahap pengumpulan data sangat penting. Tahap ini melibatkan pengumpulan data dengan tujuan untuk melatih sistem agar dapat membuat prediksi yang akurat tentang penerimaan beasiswa. Tahap ini membutuhkan kuantitas dan kualitas data, dengan fokus pada pengumpulan data yang secara khusus berkaitan dengan prediksi beasiswa. Dengan memastikan tahap pengumpulan data yang menyeluruh dan akurat, penelitian ini dapat membangun fondasi yang kuat untuk mengembangkan model prediksi beasiswa yang sukses.

Pre-processing

Data pre-processing adalah langkah penting dalam analisis data yang melibatkan pengorganisasian data dalam format yang konsisten dan mengelola nilai yang hilang dan variabel kategorikal. Proses ini mencakup *handling missing values*, yang melibatkan penanganan data yang tidak ada atau memiliki kesalahan atau kegagalan. Penanganan nilai yang hilang secara tepat sangat penting untuk analisis data yang andal (Pandey, 2020). *Handling of categorical variables*, yang mengelompokkan data dengan karakteristik yang sama, perlu diubah menjadi nilai numerik untuk diproses oleh algoritma *machine learning* (Sahu, 2021). Normalisasi melibatkan pengorganisasian ulang basis data dengan menghapus data duplikat, membuat format standar untuk penulisan, dan menetapkan korelasi yang relevan di antara data dalam tabel. Normalisasi data menyederhanakan analisis, pengambilan keputusan, dan modifikasi data. Secara keseluruhan, *data pre-processing* sangat penting untuk menyiapkan data dan memastikan integritas dan keandalannya dalam analisis lebih lanjut.

Splitting

Data splitting adalah teknik penting dalam ilmu data untuk mengembangkan model data yang akurat dan dapat

digunakan (JavaTpoint, 2021). Teknik ini melibatkan pembagian data menjadi beberapa subset, biasanya untuk tujuan *training* dan *testing*. *Training dataset* digunakan untuk membangun dan melatih model, sehingga algoritma *machine learning* dapat menyesuaikan parameternya berdasarkan data yang disediakan. Performa model kemudian dievaluasi menggunakan dataset validasi untuk menilai kemampuannya dalam mengenali pola. Di sisi lain, *testing dataset* digunakan untuk menilai kinerja model setelah proses pelatihan selesai. Algoritma *machine learning* yang berkinerja tinggi harus dapat mengkategorikan data baru dengan benar yang bukan merupakan bagian dari proses *training*, yang menunjukkan kemampuannya untuk melakukan generalisasi. Tingkat pengenalan algoritma biasanya lebih tinggi jika algoritme tersebut memiliki tingkat generalisasi yang lebih tinggi.

Klasifikasi

Klasifikasi data adalah proses pengorganisasian data ke dalam kategori yang relevan agar lebih efisien dan mudah dicari (Iskandar, 2024). Hal ini memainkan peran penting dalam manajemen risiko, kepatuhan, dan keamanan data. Hal ini melibatkan penandaan data untuk memudahkan pencarian dan pelacakan. Klasifikasi data melayani berbagai tujuan seperti menjaga kepatuhan terhadap peraturan, memfasilitasi akses, dan memenuhi kebutuhan bisnis atau pribadi. Klasifikasi data membantu organisasi memanipulasi, melacak, dan menganalisis data secara sistematis. Meskipun klasifikasi data biasanya digunakan untuk data terstruktur, klasifikasi data sangat penting untuk data yang tidak terstruktur, yang tidak memiliki label yang jelas. Dengan mengklasifikasikan data yang tidak terstruktur, data tersebut menjadi lebih berguna dan lebih mudah untuk dicari atau ditanyakan.

Evaluasi

Evaluasi merupakan suatu proses data yang deskriptif, informatif dan prediktif. Evaluasi juga merupakan serangkaian kegiatan atau aktivitas yang bertujuan untuk dapat mengukur tingkat keberhasilan dalam suatu program. Evaluasi bertujuan memberikan masukan untuk perencanaan program, memberi masukan untuk memodifikasi program dan memperoleh informasi tentang faktor pendukung dan penghambat program. Informasi yang di kumpulkan dari proses evaluasi dapat meningkatkan tingkat kinerja kegiatan yang sedang berlangsung mendapatkan gangguan interupsi yang terjadi sedari awal hingga evaluasi, dan menyadari apa yang harus dilakukan kedepannya untuk menghindari masalah untuk tetap produktif.

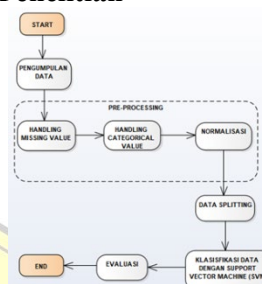
4. HASIL DAN PEMBAHASAN

Data yang digunakan adalah data siswa yang mengajukan beasiswa di SMA Negeri 8 Kota Bogor pada tahun ajaran 22/23 kelas XI IPA 1-6 dan XI IPS 1-3. Data yang diperoleh terdiri dari 314 (tiga ratus empat belas) data siswa dengan 11 (sebelas) atribut atau fitur dengan 8 (delapan) atribut bertipe string dan 3 (tiga) atribut bertipe numerik. Atribut-atribut tersebut adalah nama mahasiswa, tahun masuk, nilai rata-rata per semester, keterangan yatim piatu, alat transportasi, tanggungan orang tua, jarak, tipe rumah, gaji ayah, gaji ibu dan status penerimaan. Jika Anda harus menggunakan satuan campuran, nyatakan dengan jelas satuan untuk setiap kuantitas dalam sebuah persamaan.

Dari 11 atribut yang ada, tidak semua dapat digunakan dalam proses klasifikasi. Terdapat 9 atribut yang digunakan dalam proses klasifikasi, yaitu Nilai rata-rata per semester, Surat Keterangan Yatim Piatu, Sarana Transportasi, Jarak, Gaji Ayah, Gaji Ibu, Tipe Rumah,

Tanggungan Orang Tua, Status Penerimaan yang menentukan kelayakan calon mahasiswa untuk menerima beasiswa sebagai label. Sedangkan 2 atribut lainnya yaitu nama mahasiswa dan tahun masuk tidak digunakan.

Tahapan Penelitian



Gambar 2. Tahapan penelitian

Gambar 2 ini menjelaskan tahapan-tahapan proyek penelitian. Penelitian ini terdiri dari tujuh tahap, dimulai dengan pengumpulan data. Proses pengumpulan data dilakukan dengan cara mengamati siswa-siswi SMA Negeri 8 Bogor yang mengajukan beasiswa pada tahun ajaran 22/23. Data yang dikumpulkan meliputi informasi mengenai siswa kelas XI IPA 1-6 dan XI IPS 1-3 yang berjumlah 314 siswa. Setelah pengumpulan data, para peneliti melanjutkan dengan *data pre-processing*, yang melibatkan tiga proses: *handling missing values*, *handling categorical values*, dan normalisasi. *Handling missing values* itu penting karena sebagian besar model *machine learning* memiliki masalah dengan data *null*. Hal ini dapat diatasi melalui penghapusan atau imputasi. *Handling categorical values* melibatkan perubahan data kategorikal string ke dalam bentuk numerik menggunakan teknik seperti *One Hot Encoding* dan *Ordinal Encoding*. Atribut dalam penelitian ini ditentukan berdasarkan faktor-faktor seperti nilai rata-rata per semester dan status yatim piatu, yang telah diidentifikasi sebagai kriteria penting untuk penerimaan beasiswa.

Skenario Penerapan Metode yang Digunakan

Teks ini menjelaskan sebuah simulasi yang dilakukan dengan menggunakan *Microsoft Excel* untuk memahami tahapan-tahapan dalam proses pengolahan data. Simulasi ini melibatkan penerapan skenario pada sebuah dataset data mahasiswa pelamar beasiswa, dengan total 314 baris. Tahapan dalam skenario tersebut meliputi pengumpulan data dan *data pre-processing*. Dalam *data pre-processing*, nilai yang hilang ditangani dengan memeriksa nilai nol, dan karena tidak ada yang ditemukan, proses dilanjutkan. Nilai-nilai kategorikal kemudian ditangani dengan mengubahnya menjadi bentuk numerik. Kolom informasi yatim piatu diubah menjadi 1 untuk “Ya” dan 0 untuk “Tidak”. Kolom alat transportasi dikonversi ke nilai numerik berdasarkan kategori metode transportasi yang berbeda. Kolom tanggungan orang tua dikonversi menjadi 1 untuk “Bersama orang tua” dan “Panti Asuhan”, dan 0 untuk “Asrama”, “Wali”, dan “Asrama”. Kolom jarak sudah memiliki tipe data numerik, sehingga dimasukkan sesuai dengan itu. Kolom tipe rumah memiliki nilai string bertingkat.

Teknik Evaluasi

Untuk mengetahui performa algoritma *Support Vector Machine* dalam klasifikasi mahasiswa yang berhak menerima beasiswa, dilakukan evaluasi dengan menggunakan *K-Fold Cross Validation* dengan nilai $K=10$. Pada proses ini, data dibagi menjadi 10 subset. Kemudian dilakukan 10 kali uji coba, dimana pada setiap uji coba, subset ke- K digunakan sebagai *data testing* dan subset lainnya digunakan sebagai *data training*. *K-Fold Cross Validation* diimplementasikan dengan menggunakan modul *cross_val_score* pada *sklearn*. Berikut ini adalah hasil dari *K-Fold*

Cross Validation yang dilakukan pada algoritma tersebut.

Dari percobaan yang dilakukan dengan menerapkan *K-Fold Cross Validation*, didapatkan hasil untuk algoritma *Support Vector Machine* (SVM) memiliki permutasi tertinggi pada $K=5$, $K=6$ dan $K=8$ seperti yang dijelaskan pada Tabel 2 .

Tabel 2 *k-fold cross validation*

Fold	Skor <i>Cross Validation</i>
1	0.90625
2	0.875
3	0.75
4	0.83870968
5	0.93548387
6	0.93548387
7	0.90322581
8	0.93548387
9	0.87096774
10	0.87096774

Dari hasil klasifikasi dengan *Support Vector Machine* (SVM) yang didapatkan seperti pada tabel 2, maka didapatkan *confusion matrix* pada tabel 3.

Tabel 3 *confusion matrix*

		<i>Actual</i>	
		<i>Negative</i>	<i>Positive</i>
<i>Prediction</i>	<i>Negative</i>	65	7
	<i>Positive</i>	12	42

Berdasarkan *Confusion Matrix*, nilai *akurasi*, *presisi*, dan *recall* adalah sebagai berikut:

- Accuracy* dihitung berdasarkan Persamaan 5

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} = \frac{65 + 42}{65 + 7 + 42 + 12} = 0.84$$

b. *Precision* dihitung berdasarkan Persamaan 6

$$Precision = \frac{TP}{FP + TP} = \frac{42}{7 + 42} = 0.85$$

c. *Recall* dihitung berdasarkan Persamaan 7

$$Recall = \frac{TP}{TP + FN} = \frac{42}{42 + 12} = 0.77$$

Untuk perbandingan nilai *Accuracy*, *Precision* dan *Recall* yang diperoleh berdasarkan komposisi *data training* dan *data testing* yang berbeda terhadap prediksi yang dilakukan menggunakan algoritma *Support Vector Machine*. Dari tabel 5.5 diketahui bahwa untuk total 314 (tiga ratus empat belas) data mahasiswa, komposisi *data training* dan *data testing* yang menghasilkan nilai terbaik pada setiap metrik *Accuracy* didapatkan pada komposisi 70% *data training* dan 30% *data testing*. Komposisi ini menghasilkan nilai *Accuracy* sebesar 89.36%, *Precision* sebesar 91% dan *Recall* sebesar 95% seperti pada Tabel 4 berikut

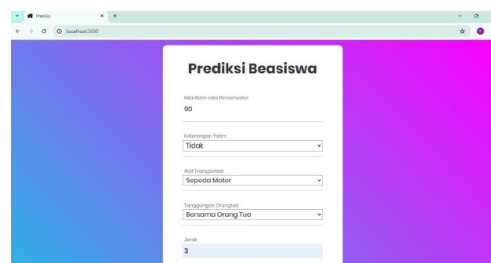
Komposisi dalam persen (%)		Metrik Akurasi		
<i>Data Training</i>	<i>Data Testing</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
90	10	81%	85%	88%
80	20	86%	91%	94%
70	30	89%	91%	95%
60	40	89%	92%	89%
Rata-rata		86%	89%	91%

Tabel 4 Evaluasi Model dengan Komposisi Data yang Berbeda

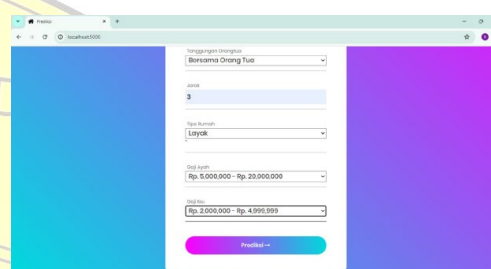
Implementasi

Setelah melakukan implementasi yang sudah dilakukan dari data *pre-processing*, *data splitting*, klasifikasi, evaluasi, dan prediksi data baru. Untuk melakukan prediksi dengan data baru dengan memasukkan ke dalam form dalam bentuk aplikasi berbasis website. Implementasinya terdiri dari implementasi *input* dan implementasi *output*. Dari *input* terdiri dari memasukan data pengaju beasiswa ke dalam form

dari proses prediksi beasiswa. Gambar 3 dan Gambar 4 merupakan *input* dari proses prediksi.

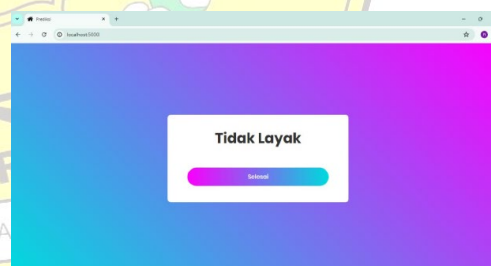


Gambar 3. Implementasi data input



Gambar 4. Implementasi data input

Implementasi *output* terdiri dari output proses prediksi beasiswa. Gambar 5 merupakan *output* dari proses prediksi 5 merupakan *output* dari prediksi beasiswa.



Gambar 5. Implementasi output

5. KESIMPULAN

Berdasarkan hasil yang diperoleh pada penelitian mengenai perbandingan algoritma *Support Vector Machine* dalam memprediksi kelayakan siswa menerima beasiswa di SMA Negeri 8 Kota Bogor, maka dapat disimpulkan bahwa Mengenai atribut ekonomi yang sedang berjalan saat ini, dengan adanya *machine learning* berbasis *web* dengan pemodelan

Support Vector Machine semua atribut yang meliputi nilai rata-rata per semester, informasi panti asuhan, alat transportasi, tanggungan orang tua, jarak, tipe rumah, gaji ayah dan gaji ibu dapat membuat SMA Negeri 8 Kota Bogor menjadi lebih akurat dan tepat sasaran dalam merekomendasikan status kelayakan siswa penerima beasiswa. Berdasarkan hasil dan uji evaluasi diatas, nilai kinerja algoritma pemodelan *Support Vector Machine* dengan nilai tertinggi pada pembagian *data training* sebanyak 70% dan *data testing* 30% dengan nilai *accuracy* 89%, *precision* 91%, *recall* 95% dan *K-Fold Cross Validation* dengan nilai tertinggi pada K=5, K=6 dan K=8 menghasilkan nilai 0.93548387. Dengan nilai evaluasi tersebut dapat membantu SMA Negeri 8 Kota Bogor dalam menentukan kelayakan siswa untuk menerima beasiswa

DAFTAR PUSTAKA

- Razak, A., & Wiguna, S. (2022). Pengaruh Beasiswa Program Indonesia Pintar (PIP) terhadap Minat Belajar Aqidah Akhlak Siswa Kelas VIII di MTS Alwashliyah, Kecamatan Babalan. *Jurnal Keguruan Ilmu Pendidikan*, 1, 249.
- Hidayatulloh, T., Suhada, S., Nursyifa, E., & Yusuf, L. (2018). Pengambilan Keputusan untuk Penerima Beasiswa Sekolah Menengah Atas Menggunakan Model Fuzzy Multiple Attribute Decision Making Weighted Product. *Jurnal Pilar Nusa Mandiri*, 14, 247.
- Gunawan, R. D., Oktavia, T., & Borman, R. I. (2018). Perancangan Sistem Informasi Beasiswa Online untuk Program Indonesia Pintar (PIP) (Studi Kasus: SMA N 1 Kota Bumi). *Jurnal Mikrotik*, 8, 43.
- Mutiah, N., & Sari, R. P. (2023). Penentuan Penerimaan Beasiswa PIP Menggunakan Metode MOORA di SD Negeri 11 Sandai. *Jurnal Khatulistiwa Informatika*, 11, 43.
- Karim, M. A. (2023). *Perbandingan Algoritma Support Vector Machine dan K-Nearest Neighbor dalam Prediksi Kelayakan Siswa untuk Menerima Beasiswa di SMK Bina Vokasi Nusantara* (Skripsi, Universitas Pancasila). Depok, Indonesia: Universitas Pancasila.
- Apandi, T. H., & Sugiarto, C. A. (2019). Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP. *JUITA: Jurnal Informatika*, 7(2). <https://dx.doi.org/10.30595/juita.v7i2.3608>
- Ramadhani, F. D., & Ardiansyah, M. (2021). Definisi Prediksi. Dalam *Sistem Prediksi Penjualan Menggunakan Metode Single Exponential Smoothing dan Tren Parabolik* (hlm. 5). South Tangerang, Indonesia: Pascal Books, PT. Mediatama Digital Cendekia.
- Gafur, A., Yulianti, S., & Hidayat, N. (2008). Cara Mudah Mendapatkan Beasiswa. Dalam *Beasiswa*. Jakarta, Indonesia: Penebar Plus.
- Haris, M. S., Nurlaila, Q., Ratnadewi, & John, F. (2019). *Introduction to Data Science*. Makassar, Indonesia: CV. Tohar Media.
- Yahya, S. (2022). *Data Mining*. Sukabumi, Indonesia: CV Jejak.

- Rumah Coding. (2024, Juni 14). Normalisasi dan Standarisasi Data dengan Scikit-Learn. Rumah Coding. Diakses pada 5 September 2024, dari <https://rumahcoding.id/blog/normalisasi-dan-standarisasi-data-dengan-scikit-learn/>.
- Aplikas. Diakses pada 5 September 2024, dari <https://aplikas.com/blog/data-classification/>.
- Arisandi, R. R. R., Warsito, B., & Hakim, A. R. (2022). Aplikasi Naive Bayes (NBC) untuk Klasifikasi Status Gizi Stunting pada Balita dengan Pengujian *K-Fold Cross Validation*. *Jurnal Gaussian*, 11.
- Al Kabir, A. H., Basuki, S., & Wicaksono, G. W. (2019). Analisis Sentimen Data Umpan Balik dari Pelatihan Teknologi Informasi Menggunakan Algoritma *Support Vector Machine*. Repositor.
- Pandey, P. (2020). *A Guide to Handling Missing Values in Python*. Diakses pada 4 Juni 2023, dari <https://www.kaggle.com/code/parulpandey/a-guide-to-handling-missing-values-in-python>.
- Sahu, A. (2021). *How to Handle Categorical Features*. Diakses pada 4 Juni 2023, dari <https://medium.com/analytics-vidhya/how-to-handle-categorical-features-ab65c3cf498e>.
- JavaTpoint. (2021). *Train and Test Datasets in Machine Learning*. Diakses pada 4 Juni 2023, dari <https://www.javatpoint.com/train-and-test-datasets-in-machine-learning>.
- Iskandar, M. I. (2024, Juli 16). Data Classification: Metode untuk Mempermudah Pengolahan Data.