Skripsi

Implementasi Algoritma *Support Vector Machine* Untuk Prediksi Tingkat Kelulusan Pada UPT Puskom Universitas Muhammadiyah Magelang

Jenis Skripsi:
Penelitian Eksperimental

Disusun Sebagai Salah Satu Syarat Memperoleh Gelar Sarjana Komputer (S.Kom.)

Disusun oleh:
Mita Utami
NIM. 18.0504.0030

Pembimbing:
Endah Ratna Arumi, M.Cs.
NIDN. 0601129001

Pembimbing:
Pristi Sukmasetya, S.Komp., M.Kom.
NIDN. 0618129201

Program Studi Teknik Informatika
Fakultas Teknik
Universitas Muhammadiyah Magelang
Tahun 2025

Bab 1 Pendahuluan

1.1 Latar Belakang Permasalahan

Kemajuan teknologi yang semakin pesat mengharuskan berbagai bidang untuk dapat beradaptasi dalam berbagai aspek digital. Universitas Muhammadiyah Magelang (UNIMMA) merupakan salah satu institusi pendidikan tinggi di Indonesia mewujudkan hal tersebut dengan melakukan pelatihan komputer dasar sebagai salah satu upaya dalam meningkatkan keterampilan mahasiswanya. Pelatihan dilaksanakan oleh Unit Pelaksana Teknis Pusat Komputer (UPT Puskom) yang mana memiliki tugas utama untuk melakukan pelatihan serta sebagai unit pendukung kebutuhan teknologi dan informasi di lingkup UNIMMA. Secara keseluruhan, pelatihan akan dilakukan dalam dua hal, yaitu mencakup materi dasar *Microsoft Office* seperti *Word*, *Excel*, dan *PowerPoint*, serta aplikasi dasar *Google*. Kegiatan ini diperuntukkan bagi mahasiswa baru dan dilaksanakan pada semester 1 dan 2 dengan tujuan agar kemampuan mahasiswa dalam bidang teknologi dan informasi dapat meningkat sebagai salah satu langkah untuk mempersiapkan mahasiswa dalam proses perkuliahan serta mengurus keperluan administratif di kawasan kampus. Kegiatan dilakukan dalam tujuh kali pertemuan yang terdiri dari enam pertemuan berisi penyampaian materi dan satu pertemuan terakhir adalah ujian. Dalam proses untuk mengetahui hasil pembelajaran yang dilakukan oleh mahasiswa maka dilakukan evaluasi. Tujuan dari evaluasi adalah untuk membantu mahasiswa dalam mengetahui kinerja pembelajarannya, keberhasilan pembelajaran mahasiswa diharapkan akan terlihat pada hasil tes maupun tugas-tugasnya (Qoiriah et al., 2021).

Keberhasilan pelatihan yang telah dilakukan dapat dilihat melalui tingkat kelulusan pesertanya, karena perolehan hasil tersebut menunjukkan sejauh mana mahasiswa memahami materi yang diberikan. Kelulusan pada setiap mata kuliah yang diambil oleh mahasiswa pada setiap semester menjadi hal yang perlu diperhatikan dikarenakan dapat mempengaruhi lamanya masa studi yang ditempuh (Ahmad et al., 2024). Berdasarkan data yang diperoleh dari hasil ujian pelatihan pada tahun 2024, terdapat 151 mahasiswa di Fakultas Teknik dan 109 di antaranya berhasil menyelesaikan pelatihan dengan tingkat kelulusan sebesar 72%. Fakultas Psikologi dan Humaniora mencatat angka yang lebih rendah, yaitu 63%, dimana 62 mahasiswa dinyatakan lulus dari total 97 mahasiswa. Hasil ini diperoleh dengan mengumpulkan jawaban ujian peserta pelatihan melalui Google Formulir yang kemudian diperiksa secara manual sehingga diperoleh nilai ujian. Nilai tersebut digunakan sebagai acuan dalam menentukan kelulusan mahasiswa. Tingkat kelulusan yang tergolong rendah mendorong adanya langkah pencegahan yang harus diambil. Salah satu langkah yang dapat digunakan adalah dengan melakukan prediksi terhadap tingkat kelulusan mahasiswa di masa mendatang. Peramalan nilai akhir mahasiswa dapat memberikan keuntungan baik bagi mahasiswa maupun pengajar untuk lebih memahami kinerja mahasiswa sebelum menetapkan nilai dan juga untuk mengetahui faktor yang menyebabkan mahasiswa tidak lulus. Menentukan kelulusan mahasiswa tidak hanya didasarkan pada satu faktor tunggal, tetapi melibatkan berbagai atribut seperti nilai ujian, kehadiran, kelas yang diikuti, dan faktor-faktor lainnya (Suriani, 2023). Hal ini memungkinkan dilakukannya langkah korektif

secara cepat terhadap mahasiswa yang memperoleh hasil di bawah standar. Bagi pengajar atau dosen, faktor ini sangat penting karena akan menjadi dasar untuk menentukan tindakan selanjutnya terhadap mahasiswa yang mungkin tidak lulus, serta menyesuaikan metode pelatihan yang ditetapkan.

Prediksi didefinisikan sebagai proses untuk mengetahui kondisi di masa yang akan datang berdasarkan data historis atau data masa lalu yang sudah ada. Dalam bidang *Data mining* dan *machine learning, Support Vector Machine* (SVM) merupakan salah satu metode yang terbukti menghasilkan nilai peramalan tinggi terutama pada klasifikasi *biner*. Klasifikasi *biner* adalah model statistik yang membagi kumpulan data menjadi dua kelompok (Shedriko, 2021). *Support Vector Machine (SVM)* adalah metode untuk menganalisis data dan mengenali pola yang dapat digunakan untuk pengklasifikasian (Bumbungan et al., 2023). SVM merupakan salah satu metode klasifikasi *supervised learning* yang ditandai dengan adanya kelas/label/target pada himpunan datanya. Dalam penelitian yang dilakukan oleh (Putri et al., 2023) membandingkan metode K-NN, *Naïve Bayes*, SVM untuk memprediksi kelulusan mahasiswa tingkat akhir. Penelitian yang dilakukan bertujuan untuk melihat model pengklasifikasian mana yang meghasilkan nilai akurasi yang lebih baik. Berdasarkan pengujian terhadap 379 data menggunakan *splitting data* 90:10, 80:20, 70:30, 60:40, dan 50:50 menunjukkan bahwa K-NN mencapai nilai akurasi rata-rata tertinggi (87,8%) dibandingkan dengan dua metode lainnya yaitu *Naïve Bayes* (82,6%) dan SVM (73,6%).

Sementara itu penelitian lainnya dilakukan oleh (Junaidi et al., 2024) untuk memprediksi kelulusan tepat waktu menggunakan algoritma *Naïve Bayes*, *Random Forest*, *Support Vector Machine* (SVM), dan *Artificial Neural Network* (ANN). Tujuan dari penelitian ini adalah untuk mengidentifikasi pola yang dapat memprediksi kelulusan tepat waktu di kalangan mahasiswa, sehingga informasi tersebut dapat membantu perguruan tinggi dalam mengambil langkah-langkah pencegahan terhadap mahasiswa yang tidak lulus dalam jangka waktu yang ditentukan. Penelitian dilakukan terhadap 101 mahasiswa calon wisudawan Fakultas Sains dan Teknologi Universitas PGRI Sumatera Barat. Dari *dataset* yang ada, dibagi menjadi 70% data *training* dan 30% data *testing* lalu diuji menggunakan *Confussion Matrix* sehingga diketahui metode yang menghasilkan tingkat akurasi tertinggi adalah SVM (94%), sementara metode lainnya menghasilkan akurasi sebesar *Naïve Bayes* (87%), *Random Forest* (84%), dan ANN (65%).

Berdasarkan uraian sebelumnya, penelitian ini bertujuan untuk memprediksi kelulusan peserta pelatihan di UPT Puskom dan mengidentifikasi faktor-faktor penyebab ketidaklulusan sehingga memungkinkan dilakukannya pencegahan sejak dini. *Support Vector Machine* dipilih sebagai metode dalam penelitian ini karena dianggap mampu memberikan tingkat akurasi yang lebih baik dibandingkan dengan metode lainnya, menjadikannya tepat untuk digunakan dalam meramalkan kelulusan.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah bagaimana implementasi metode *Support Vector Machine* dalam memprediksi tingkat kelulusan pelatihan di UPT Puskom UNIMMA.

1.3 Tujuan Penelitian

Penelitian ini bertujuan untuk menilai sejauh mana metode *Support Vector Machine* dapat menghasilkan tingkat akurasi yang relevan terhadap data kelulusan pelatihan di UPT Puskom UNIMMA.

1.4 Manfaat Penelitian

Hasil dari penelitian yang telah dilakukan diharapkan dapat memberikan pemahaman tentang seberapa efektif metode *Support Vector Machine* dalam memprediksi hasil kelulusan di UPT Puskom UNIMMA, sehinga bisa dimanfaatkan sebagai acuan dalam proses pengambilan keputusan untuk meningkatkan efisiensi program pelatihan.

Bab 2 Studi Literatur

2.1 Tinjauan Penelitian Terdahulu

Penelitian yang dilakukan oleh Wafa et al., (2022) memanfaatkan metode SVM untuk melakukan memprediksi penyakit diabetes. Pemilihan algoritma SVM *Radial Basis Function* (RBF) dan metode *Forward Selection* karena dianggap mampu memberikan tingkat akurasi yang tinggi, sehingga dapat berpotensi mengurangi angka kasus diabetes. Data yang digunakan dalam pelatihan mencakup 8 atribut yaitu *n_pregnant, glucose_conc, bp, skin_len, insulin, bmi, pedigre_funct, age* dimana 80% digunakan sebagai data *training* dan 20% sebagai data *testing*. Setelah dilakukan pengujian menggunakan metode *Support Vector Machine* dengan kernel *Radial Basis Function* dan *Forward Selection* terhadap 204 data uji, menghasilkan *accuracy* 91,2%, *precision* 93,0%, *recall* 94,3%, dan untuk *f1-score* 93,7%. Pengaplikasian algoritma SVM dengan kernel RBF dan metode *Forward Selection* terbukti efektif dan menghasilkan tingkat akurasi yang tinggi dalam memprediksi kemungkinan resiko diabetes sehingga berpotensi membantu profesional medis dalam melakukan identifikasi awal dan pengambilan keputusan yang berhubungan dengan diagnosis diabetes dengan cara yang lebih efisien dan tepat.

Sementara itu Bumbungan et al., (2023) membuktikan bahwa penerapan *Particle swarm optimization* (PSO) untuk memilih parameter *Support Vector Machine* (SVM) mampu meningkatkan tingkat akurasi secara signifikan. SVM memberikan hasil yang baik dalam klasifikasi dan dapat menangani isu seperti *overfitting*, jumlah data *training* yang terbatas, serta lambatnya proses konvergensi, namun SVM masih menghadapi tantangan dalam komputasi dengan *dataset* besar dan dalam memilih parameter yang tepat. Data yang digunakan mencakup 353 mahasiswa yang kemudian dilakukan pengujian menggunakan model SVM, baik dengan maupun tanpa optimasi parameter (Gama, C, Epsilon) melalui PSO dalam *software Rapid Miner* 9.10. Hasil evaluasi performa SVM tanpa optimasi PSO, yang diuji menggunakan *Confussion Matrix* dan validasi *10-fold cross-validation* menunjukkan tingkat akurasi sebesar 93,33%, *recall* 91,04%, *precision* 98,39%, *f1-score* 94,57%. Setelah diterapkannya PSO, hasilnya menunjukkan peningkatan dengan akurasi 98,02%, *recall* 98,55%, *precision* 98,08%, *f1-score* 98,31%. Implementasi PSO sebagai optimasi parameter terbukti mampu meningkatkan akurasi serta efektivitas dalam prediksi, sehingga dapat menjadi solusi untuk memonitor serta meningkatkan kualitas lulusan.

Penelitian Abdullah et al., (2022) menunjukkan bahwa algoritma SVM dan teknik *boosting* mampu melakukan klasifikasi secara efektif dalam meramalkan nilai serta waktu kelulusan mahasiswa. Bukti dari temuan ini terletak pada tingkat akurasi yang diperoleh selama pengujian, yang mencapai 86,36% dan AUC 0,876 dengan kesalahan sebesar 13,64%. Hasil penelitian ini diperoleh melalui analisis terhadap 807 data mahasiswa yang dilakukan dengan *Rapid Miner*, mengaplikasikan pendekatan CRISP-DM dan diuji menggunakan metode *boosting* (Adaboost) serta *k-fold cross validation* dengan nilai k=5,7, dan 10. Variabel yang diuji mencakup jenis kelamin, indeks prestasi semester 1 hingga 4, serta total SKS semester 1 hingga 4. Penerapan algoritma SVM dan teknik *boosting* terbukti efisien dalam meramalkan nilai serta waktu kelulusan siswa

berdasarkan catatan akademis mereka. Sistem ini bermanfaat bagi perguruan tinggi dalam membuat keputusan yang dapat meningkatkan kualitas lulusan serta memberikan dukungan lebih awal kepada siswa yang beresiko tindak memenuhi kriteria kelulusan.

Penelitian lain yang dilakukan oleh Haryatmi et al., (2021) memanfaatkan algoritma SVM untuk meramalkan kelulusan siswa di SMK Adiluhur. Sejumlah 514 data dengan 10 atribut yang terdiri dari 9 variabel prediktor dan 1 hasil digunakan dalam pemodelan dengan algoritma SVM yang dijalankan melalui *Google Colab*. Data yang ada diproses melalui pembersihan hingga tersisa 405 data yang sesuai, kemudian dilakukan seleksi fitur untuk memilih data yang paling relevan sehingga dapat meningkatkan performa model dan efisiensi komputasi. Setelah proses pelatihan data dengan model SVM selesai, model dievaluasi menggunakan *Confussion Matrix* yang menghasilkan nilai akurasi 95,06%, *precision* 97%, dan *recall* 82%. Faktor utama yang berpengaruh adalah nilai akademik, sementara faktor extrakurikuler, non-akademik, dan interpersonal menunjukkan hubungan yang lebih rendah terhadap kelulusan.

Berdasarkan penjelasan dari sejumlah sumber yang telah diuraikan, disimpulkan bahwa prediksi tingkat kelulusan sangat penting untuk mengevaluasi kemampuan mahasiswa dan sebagai dasar untuk mengambil tindakan terhadap mahasiswa yang mungkin akan gagal. Metode *Support Vector Machine* (SVM) dipilih karena dalam penelitian sebelumnya telah terbukti efektif dalam memprediksi kelulusan dan SVM menunjukkan hasil yang lebih unggul dengan tingkat akurasi yang cukup tinggi, yaitu lebih dari 86%. Penelitian ini memiliki perbedaan dengan studi-studi sebelumnya karena tidak hanya memperkirakan tingkat kelulusan, tetapi juga mengenali berbagai faktor yang mempengaruhinya. Dengan cara ini, langkah-langkah pencegahan dapat diambil baik dalam hal materi ajaran maupun metode pengajaran.

2.2 Kajian Teoretis

2.2.1. *Data mining*

Data mining merupakan suatu kumpulan teknik yang digunakan sebagai bagian dari pengetahuan dalam pencarian di basis data (Haryatmi et al., 2021). *Data mining* digunakan untuk menganalisis data secara terstruktur untuk mengidentifikasi pola atau tren yang relevan dengan tujuan penelitian. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan (Nugroho et al., 2023). Penggunaan *data mining* dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang faktor yang mempengaruhi kelulusan (Rahmayanti et al., 2022). Proses ini mencakup berbagai langkah seperti pengumpulan data, pengumpulan dan integrasi data, pengumpulan dan ekstraksi data, transformasi data, dan analisis data. *Data mining* digunakan dengan tujuan mengolah data mentah agar dapat menghasilkan informasi yang bernilai dan mendukung proses pengambilan keputusan. Dalam penelitian ini, data yang digunakan terdiri atas tujuh belas atribut data yaitu, no, npm, nama, absn1, absn2, absn3, absn4, absn5, absn6, nilai_word, nilai_excel, nilai_ppt, nilai_google, jumlah, hasil, asal_sekolah, dan *gender*.

Berdasarkan alur CRISP-DM (*Cross Industry Standard Process for Data mining*), terdapat enam alur dalam *data mining* yaitu pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), pengolahan data (*data preparation*), permodelan (*modeling*), evaluasi (*evaluation*), dan *deployment*.

Teknik dalam *data mining* antara lain:

1. Klasifikasi (*Classification*)
   Klasifikasi merupakan metode dalam *data mining* yang digunakan untuk memetakan data ke dalam kategori atau kelas tertentu berdasarkan pola dari data yang telah dilabeli sebelumnya.
2. Prediksi (*Prediction*)
   Prediksi adalah teknik di dalam *data mining* yang berfungsi untuk memperkirakan nilai atau hasil di masa depan berdasarkan pola masa lalu.
3. Klastering (*Clustering*)
   Klastering merupakan salah satu teknik di dalam *data mining* yang memiliki tujuan untuk mengelompokkan sekumpulan data ke dalam beberapa kelompok berdasarkan kemiripan datanya.
4. Asosiasi (*Association Rule Learning*)
   Asosiasi digunakan untuk menmukanhubungan atau pola keterkaitan antar item dalam kumpulan data.

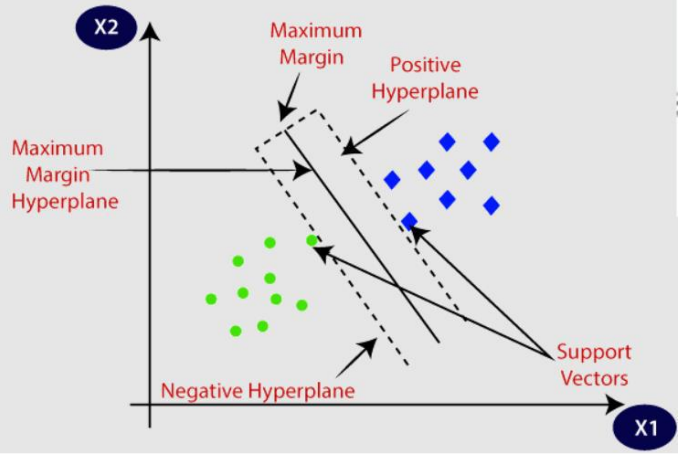
2.2.2. Prediksi Kelulusan

Prediksi kelulusan adalah suatu tahap untuk menilai kemungkinan seorang siswa atau peserta program pelatihan akan berhasil atau gagal, dengan menggunakan informasi masa lalu yang tersedia. Pada hakekatnya prediksi hanya merupakan suatu perkiraan (*guess*), akan tetapi dengan menggunakan teknik-teknik tertentu, maka prediksi menjadi lebih dari sekedar perkiraan (Renyut et al., 2022). Prediksi kelulusan biasanya dilakukan menggunakan pendekatan *data mining* dan pembelajaran mesin (*machine learning*) untuk mengidentifikasi pola-pola tertentu yang berhubungan dengan keberhasilan atau kegagalan peserta. Analisa dan prediksi diharapkan mampu menemukan faktor-faktor yang mempengaruhi dan memprediksi kelulusan tepat waktu (Yatimah, 2021).

Kelulusan mahasiswa dalam sebuah mata kuliah dapat mempengaruhi pengambilan mata kuliah di semester yang akan datang (Sagala, 2021). Hal ini akan meningkatkan beban belajar bagi mahasiswa karena dapat mengganggu mata kuliah di semester berikutnya. Oleh sebab itu, dilakukan peramalan sebagai langkah proaktif untuk mahasiswa dan pengajar supaya dapat melakukan pelatihan dengan optimal dan menurunkan tingkat kegagalan dalam ujian. Salah satu cara untuk mengatasi adanya ketidaktepatan waktu lulus mahasiswa adalah dengan memanfaatkan algoritma *machine learning* yang dapat memproses data dan membuat prediksi berdasarkan pola yang terlihat (Putri et al., 2024).

2.2.3. *Support Vector Machine*

Support Vector Machine (SVM) termasuk dalam salah satu algoritma *supervised learning* yang pada umumnya digunakan dalam klasifikasi dan regresi. SVM dikembangkan oleh Vladimir Vapnik dan koleganya pada awal tahun 1990-an. Dalam pemodelan karakterisasi, SVM mempunyai konsep yang lebih matang dan lebih jelas secara numerik dibandingkan dengan metode pengelompokan lainnya (Tajrin et al., 2024). SVM memiliki fungsi utama untuk memisahkan data secara linier, namun dalam penerapannya terdapat sejumlah data non-linier yang menyebabkan SVM perlu menggunakan kernel. Kernel digunakan untuk memetakan data ke dimensi yang lebih tinggi guna meningkatkan akurasi, sebab akurasi yang dicapai sebelumnya belum optimal. Hasil akurasi data yang dihasilkan algoritma SVM ditentukan oleh parameter dan fungsi kernel yang digunakan (Amelia et al., 2022). Fungsi kernel memungkinkan untuk mengimplementasikan suatu model pada ruang dimensi lebih tinggi (ruang fitur) tanpa harus mendefinisikan fungsi pemetaan dari ruang *input* ke ruang fitur (Wafa et al., 2022). Dengan bantuan kernel SVM, pelatihan *input* dapat dihubungkan ke dalam dimensi ruang yang lebih luas dan menemukan *hyperplane* sebagai area pemisah. Dengan bantuan vektor dan margin, SVM menemukan *hyperplane* (Junaidi et al., 2024). Seperti yang terlihat pada gambar 2.1, *hyperplane* adalah batas keputusan yang memisahkan tipe dari satu kelas dengan kelas yang lain (Permana & Silvanie, 2021).



Gambar 2.1 *Hyperplane* SVM (Bumbungan et al., 2023)

Dalam metode SVM terdapat beberapa kernel yang umum digunakan antara lain:

1. *Linier Kernel*

   Digunakan jika data dapat dipisahkan secara linier. Dirumuskan sebagai berikut:

   $K(x, y) = x \cdot y$

3. Polynomial Kernel
   Menghasilkan permukaan pemisah berbentuk polinomial, digunakan untuk data yang hubungan antar fiturnya polinomial. Dirumuskan sebagai berikut:
   $K(x, y) = (x . y + c)^d$

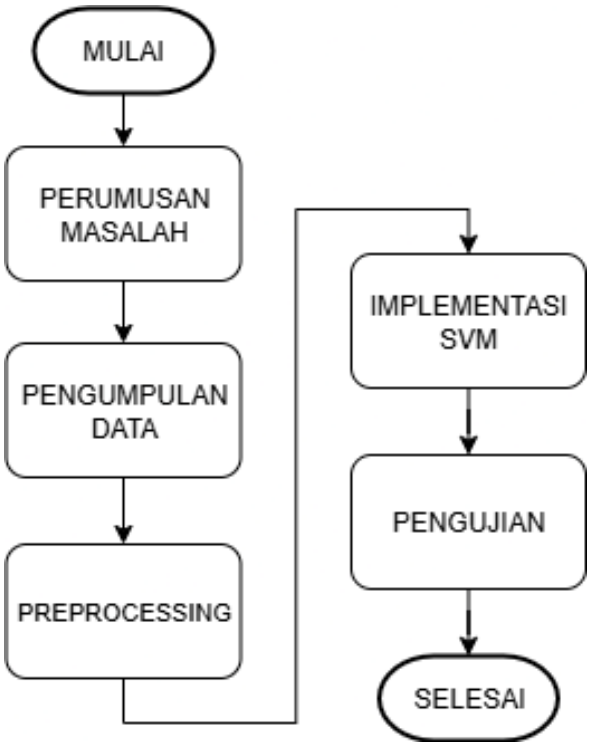
4. *Radial Basis Function* (RBF) / Gaussian Kernel
   Ini adalah kernel yang paling umum, digunakan untuk data yang distribusinya kompleks dan tidak linier. Dirumuskan sebagai berikut:
   $K(x, y) = \exp(-y \mid\mid x - y \mid\mid^2)$

5. Sigmoid Kernel
   Kernel ini dikenal sebagai kernel yang prinsip kerjanya meniru cara kerja neuron dalam jaringan saraf. Dirumuskan sebagai berikut:
   $K(x, y) = \tanh(\alpha x . y + c)$

Bab 3 Metode Penelitian

3.1 Prosedur Penelitian Eksperimental

Data yang digunakan dalam penelitian ini adalah data hasil ujian pelatihan di UPT Puskom Universitas Muhammadiyah Magelang tahun 2024 dari dua fakultas yaitu, Fakultas Teknik, dan Fakultas Psikologi dan Humaniora. Data kemudian diolah dengan metode *Support Vector Machine* sehingga didapat hasil prediksi kelulusan mahasiswa. Tahapan penelitian dapat dilihat pada gambar 3.1



Gambar 3.1 Prosedur Penelitian

3.1.1 Perumusan Masalah

Langkah awal dalam penelitian dimulai dengan perumusan masalah, yakni tingginya presentase mahasiswa yang tidak lulus pada pelatihan di UPT Puskom UNIMMA, yang mengakibatkan mahasiswa harus menjalani ujian ulang. Dengan demikian, sangat penting untuk melakukan analisis prediktif guna mengidentifikasi berbagai faktor yang dapat mempengaruhi, sehingga pelatihan dapat dilakukan dengan lebih efektif.

3.1.2 Pengumpulan Data

Penelitian ini memanfaatkan sumber data primer berupa hasil ujian dari pelatihan yang dilakukan oleh UPT Puskom UNIMMA. Data yang diperoleh mencakup tujuh belas atribut yang nantinya akan diterapkan dalam prediksi dengan memanfaatkan algoritma SVM.

3.1.3 *Preprocessing*

Preprocessing dilakukan untuk membersihkan dan melakukan integrasi terhadap data yang akan diprediksi. *Preprocessing* dilakukan karena data mentah tidak dapat digunakan langsung oleh sistem (Putranto et al., 2023). Terdapat tujuh belas atribut data yang akan digunakan yaitu, no, npm, nama, absn1, absn2, absn3, absn4, absn5, absn6, nilai_word, nilai_excel, nilai_ppt, nilai_google, jumlah, hasil, asal_sekolah, dan *gender*. Proses *preprocessing* dilakukan dalam lima tahap yaitu,

1. Transformasi data
   Transformasi data yang dilakukan adalah dengan menjumlahkan beberapa kolom pada atribut absensi menjadi angka dalam satu kolom baru.
2. *Encoding* data kategorikal
   Mengubah atribut hasil, *gender*, dan asal sekolah dari data yang bersifat kategorikal/teks menjadi angka agar data dapat dibaca oleh model.
3. Pembersihan data
   Dilakukan dengan menghapus atribut no, nama, absn1, absn2, absn3, absn4, absn5, absn6, dan jumlah karena dianggap tidak relevan dan tidak memiliki keterkaitan dengan performa model.
4. Splitting data
   Data yang sudah dibersihkan lalu dibagi menjadi *data training* dan *data testing* dengan perbandingan 80:20 yakni 80% untuk *data training* dan 20% untuk *data testing*.

3.1.4 Implementasi SVM

Support Vector Machine (SVM) digunakan untuk menganalisis pola dari data pelatihan untuk menyusun suatu model prediksi. Model ini selanjutnya akan diterapkan untuk memperkirakan kategori (Lulus atau Tidak Lulus) pada data baru yang tercipta. Tahapan pelatihan data pada metode SVM antara lain:

1. Pemilihan kernel, model ini memanfaatkan fungsi kernel linier dikarenakan data yang digunakan dapat dibedakan secara linier. Kernel linier beroperasi dengan menentukan garis atau *hyperplane* optimal yang memisahkan kelas sasaran secara langsung tanpa harus memetakan ke ruang dengan dimensi lebih tinggi.
2. Penyesuaian parameter, digunakan parameter C (*Cost Function*) dikarenakan parameter ini mampu mengatur keseimbangan antara margin tertinggi dan kesalahan dalam pengklasifikasian. Nilai C yang tinggi akan berupaya memperkecil kesalahan klasifikasi tetapi dapat menyebabkan model menjadi *overfitting*.
3. Inisiasi model dan pelatihan, dalam proses ini digunakan *library scikit-learn* untuk melakukan inisiasi model SVM dengan kernel linier dan parameter C.

4. Identifikasi support vector, SVM akan secara otomatis menentukan data yang paling penting untuk posisi *hyperplane*, yang disebut *support vector*. Data ini memainkan peran yang sangat krusial dalam penentuan model klasifikasi.
5. Setelah semua tahap dilakukan, model SVM yang telah dilatih akan siap digunakan untuk proses prediksi dan pengujian.

3.1.5 Pengujian

Data yang telah dilakukan pemodelan kemudian diuji untuk mengetahui tingkat keberhasilannya. Pengujian dilakukan menggunakan *Confussion Matrix* untuk mencari nilai *accuracy*, *precision*, *recall*, dan *f1-score*. *Confussion Matrix* terdiri dari dua kelas yaitu positif dan negatif. Pada *Confussion Matrix* ini didalamnya berisi suatu informasi tentang kelas yang aktual dan kelas yang diprediksi dari proses klasifikasi (Fajriansyah et al., 2025).

		Prediksi	
		TRUE	FALSE
Aktual	TRUE	True Positif (TP)	False Negative (FN)
	FALSE	False Positif (FP)	True Negatif (TN)

Gambar 3.2 *Confussion Matrix*

1. *Accuracy*, untuk melihat presentase prediksi yang benar dari seluruh data yang diuji. Rumus:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. *Precision*, untuk melihat proporsi hasil prediksi positif yang benar-benar positif. Rumus:

$$Precission = \frac{TP}{TP + FP}$$

3. *Recall*, untuk melihat proporsi data aktual positif yang berhasil dideteksi oleh model. Rumus:

$$Recall = \frac{TP}{TP + FN}$$

4. *F1-score*, rata-rata harmonik antara *precission* dan *recall*, memberikan keseimbangan untuk kedua metrix tersebut. Rumus:

$$F1 = 2 . \frac{Presiccion \, . \, Recall}{Precission + Recall}$$

3.2 Analisa Sistem

Analisis sistem ini digunakan untuk merancang serta memahami kebutuhan sistem, sehingga model prediksi dapat mengelompokkan peserta pelatihan ke dalam dua kategori yaitu "Lulus" dan "Tidak Lulus" berdasarkan atribut yang dimiliki. Data yang digunakan terdiri dari delapan atribut, yang terdiri dari tujuh fitur (*input*) yaitu npm, nilai_word, nilai_excel, nilai_ppt, nilai_google, asal_sekolah, *gender*, absen dan 1 atribut label berupa hasil.

Sebelum diolah menggunakan metode SVM, dilakukan *preprocessing* untuk memastikan bahwa tiap atribut data tidak memiliki variabel yang dapat menggangu proses *modeling*. *Preprocessing* dilakukan dalam lima tahapan yaitu, pembersihan data (*data cleaning*), transformasi data, normalisasi data, *encoding* data kategorikal, dan *splitting data*.

Data yang sudah dilakukan *preprocessing* kemudian diuji dengan metode SVM menggunakan kernel linier. Selanjutnya data tersebut akan diklasifikasi oleh metode SVM terhadap data uji sehingga menghasilkan performa berupa *accuracy*, *precision*, *recall*, dan *f1-score* dengan metode *Confussion Matrix*.

Bab 5 Penutup

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai implementasi algoritma Support Vector Machine (SVM) dalam memprediksi tingkat kelulusan peserta pelatihan di UPT Puskom UNIMMA dapat disimpulkan bahwa prediksi kelulusan berhasil dilakukan menggunakan metode SVM dengan beberapa tahapan seperti preprocessing data (pembersihan dan encoding), pelatihan model dengan algoritma SVM kernel linier, hingga evaluasi menggunakan Confussion Matrix. Model SVM menunjukkan performa yang sangat baik dengan akurasi sebesar 98%, presisi 99%, recall 97%, dan f1-score 98% seperti terlihat dari Confussion Matrix yang hanya menghasilkan satu kesalahan klasifikasi dari total 50 data uji. Ini menunjukkan bahwa algoritma SVM sangat efektif dalam mengklasifikasikan peserta berdasarkan atribut yang digunakan.

Selain itu, penelitian ini juga telah mengimplementasikan antarmuka berbasis web menggunakan Streamlit guna mempermudah penggunaan model prediksi secara interaktif. Melalui antarmuka ini, pengguna dapat memasukkan data baru yang relevan, kemudian sistem secara otomatis melakukan preprocessing, standarisasi, serta prediksi dengan model SVM yang telah dilatih. Hasil prediksi ditampilkan secara langsung dalam bentuk teks maupun visualisasi sederhana sehingga lebih mudah dipahami oleh pengguna non-teknis. Distribusi kelulusan peserta pelatihan menunjukkan bahwa 67,3% peserta dinyatakan lulus dan 32,7% tidak lulus, yang berarti masih ada sepertiga peserta yang gagal. Informasi ini penting sebagai dasar evaluasi terhadap efektivitas pelatihan.

5.2 Saran

Saran yang dapat diberikan dari penelitian ini adalah, meskipun implementasi antarmuka pengguna berbasis web dengan Streamlit telah dilakukan sehingga hasil prediksi dapat diakses dengan lebih mudah tanpa harus memahami bahasa pemrograman Python, pengembangan lanjutan tetap diperlukan agar antarmuka menjadi lebih ramah dan interaktif. Beberapa pengembangan yang dapat dilakukan antara lain menambahkan fitur pelaporan otomatis, memperkaya visualisasi hasil prediksi, serta mengintegrasikan sistem prediksi dengan administrasi pelatihan yang telah berjalan di UPT Puskom UNIMMA.

Hasil prediksi yang diperoleh melalui antarmuka ini diharapkan dapat dimanfaatkan secara langsung oleh pengajar pelatihan untuk melakukan pendampingan terhadap peserta yang diprediksi tidak lulus, maupun menyesuaikan strategi pembelajaran selama proses pelatihan berlangsung. Penelitian selanjutnya disarankan untuk memfokuskan diri pada pengembangan sistem yang lebih adaptif, yakni dengan menyesuaikan materi pelatihan berdasarkan kelemahan peserta yang teridentifikasi dari hasil prediksi. Dengan pendekatan ini, proses pembelajaran diharapkan dapat berjalan lebih efektif dan tingkat kelulusan keseluruhan peserta pelatihan dapat meningkat secara signifikan.

Referensi

Abdullah, M. F., Kusrini, & Arief, M. R. (2022). Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Menggunakan Metode *Support Vector Machine*. Saintekbu, 14(01), 35–44. https://doi.org/10.32764/saintekbu.v14i01.1096

Ahmad, N., Hafizh, S., & Sulthanah, R. (2024). Prediksi Kelulusan Mata Kuliah Mahasiswa Teknologi Informasi Menggunakan Algoritma K-Nearest Neighbor. Jurnal Manajemen Informatika (JAMIKA), 14(2), 135–149. https://doi.org/10.34010/jamika.v14i2.12454

Amelia, U., Indra, J., & Masruriyah, A. F. N. (2022). Implementasi Algoritma *Support Vector Machine* (Svm) Untuk Prediksi Penyakit Stroke Dengan Atribut Berpengaruh. Scientific Student Journal for Information, Technology and Science, III(2), 254–259.

Bumbungan, S., Kusrini, & Kusnawi. (2023). Penerapan *Particle swarm optimization* (PSO) dalam Pemilihan Parameter Secara Otomatis pada *Support Vector Machine* (SVM) untuk Prediksi Kelulusan Mahasiswa Politeknik Amamapare Timika. Jurnal Teknik AMATA, 4(1), 81–93. https://doi.org/10.55334/jtam.v4i1.77

Fajriansyah, A., Yusup, D., & Prihandini, K. (2025). Prediksi Kelulusan Nilai Kalkulus Menggunakan Naïve Bayes. JATI (Jurnal Mahasiswa Teknik Informatika), 9(2), 3256–3260. https://doi.org/10.36040/jati.v9i2.13343

Haryatmi, E., & Hervianti, S. P. (2021). Penerapan Algoritma *Support Vector Machine* Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu. Resti, 5(2), 958–965.

Junaidi, S., Anggela, R. V., & Kariman, D. (2024). Klasifikasi Metode *Data mining* untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes, Random Forest, *Support Vector Machine* (SVM) dan Artificial Neural Nerwork (ANN). Journal of Applied Computer Science and Technology, 5(1), 109–119. https://doi.org/10.52158/jacost.v5i1.489

Lukman, & Herlinda. (2024). Prediksi Kelulusan Siswa dengan Metode *Support Vector Machine* (SVM) di SMK Adiluhur. STRING (Satuan Tulisan Riset Dan Inovasi Teknologi), 9(1), 115. https://doi.org/10.30998/string.v9i1.23355

Nugroho, B. I., Santoso, N. A., & Murtopo, A. A. (2023). Prediksi Kemampuan Akademik Mahasiswa dengan Metode *Support Vector Machine*. Remik, 7(1), 177–188. https://doi.org/10.33395/remik.v7i1.12010

Permana, D. S., & Silvanie, A. (2021). Prediksi Penyakit Jantung Menggunakan *Support Vector Machine* dan Python pada Basis Data Pasien di Cleveland. JUNIF: Jurnal Nasional Informatika, 2(1), 29–34.

Putranto, A., Azizah, N. L., & Ratna Ika, A. I. (2023). Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit. Jurnal Penerapan Sistem Informasi (Komputer & Manajemen), 4(2), 442–452. https://archive.ics.uci.edu/ml/*dataset*s/heart+disease

Putri, A., Hardiana, C. S., Novfuja, E., Siregar, F. T. P., Rahmaddeni, Fatma, Y., & Wahyuni, R. (2023). Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 3(1), 20–26. https://doi.org/10.57152/malcom.v3i1.610

Putri, S. A. H., & Putra, J. A. (2024). Analisis Komparasi Algoritma C4.5, Naive Bayes Dan K-

Nearest Neighbor Untuk Memprediksi Ketepatan Waktu Lulus Mahasiswa. 7(2), 172–184.

Qoiriah, A., & Yamasari, Y. (2021). Prediksi Nilai Akhir Mahasiswa Dengan Metode Regresi (Studi Kasus Mata Kuliah Pemrograman Dasar). Journal of Information Engineering and Educational Technology, 5(1), 40–43. https://doi.org/10.26740/jieet.v5n1.p40-43

Rahmayanti, A., Rusdiana, L., & Suratno, S. (2022). Perbandingan Metode Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa. Walisongo Journal of Information Technology, 4(1), 11–22. https://doi.org/10.21580/wjit.2022.4.1.9654

Renyut, D. H., Yuyun, & Ferdinand. (2022). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C.45 (Studi Kasus: Sekolah Tinggi Ilmu Administrasi Trinitas Ambon). Simtek : Jurnal Sistem Informasi Dan Teknik Komputer, 7(2), 80–86. https://doi.org/10.51876/simtek.v7i2.137

Sagala, R. M. (2021). Prediksi Kelulusan Mahasiswa Menggunakan *Data mining* Algoritma K-Means. Jurnal TeIKa, 11(2), 131–142.

Shedriko, S. (2021). Perbandingan Algoritma SVM dan KNN dalam Mengklasifikasi Kelulusan Mahasiswa pada Suatu Mata Kuliah. STRING (Satuan Tulisan Riset Dan Inovasi Teknologi), 6(2), 115. https://doi.org/10.30998/string.v6i2.9160

Suriani, U. (2023). Penerapan *Data mining* Untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5. Journalcisa, 3(2), 55–66. http://jesik.web.id/index.php/jesik/article/view/91

Tajrin, Sembiring, S. H. S., & Ndruru, S. K. (2024). Penerapan Metode Forecasting dengan Algoritma *Support Vector Machine* untuk Memprediksi Penerimaan Peserta Didik Baru pada SMA ULUN NUHA. 7(2), 931–937. https://doi.org/10.37600/tekinkom.v7i2.1525

Wafa, H. S., Hadiana, A. I., & Umbara, F. R. (2022). Prediksi Penyakit Diabetes Menggunakan Algoritma *Support Vector Machine* (SVM). Informatics and Digital Expert (INDEX), 4(1), 40–45. https://doi.org/10.36423/index.v4i1.895

Yatimah, M. N. (2021). Implementasi *Data mining* untuk Prediksi Kelulusan Tepat Waktu Mahasiswa STIMIK ESQ Menggunakan Decision Tree C4.5. JUMANJI (Jurnal Masyarakat Informatika Unjani), 5(2), 89. https://doi.org/10.26874/jumanji.v5i2.95