

BAB III

METODOLOGI PENELITIAN

3.1 Subjek dan Objek Penelitian

Subjek dari penelitian ini adalah mahasiswa Prodi Teknik informatika Institut Teknologi Telkom Purwokerto. Sedangkan untuk objek penelitian yang akan diamati yaitu kinerja mahasiswa di prodi Teknik Informatika ITTP Angkatan 2017 hingga 2019 dengan menggunakan IPS dari semester satu hingga enam.

3.2 Alat dan Bahan Penelitian

Dalam melakukan penelitian Klasifikasi Kinerja Mahasiswa Menggunakan Metode C4.5 dan Support Vector Machine (SVM) membutuhkan beberapa peralatan dan bahan untuk menunjang proses penelitian.

3.2.1 Alat penelitian

Alat yang akan digunakan pada penelitian ini meliputi perangkat keras dan perangkat lunak serta beberapa teori dan standar. Berikut adalah detail dari alat-alat yang akan digunakan:

1. Perangkat Keras:

- a) Processor 11th Gen Intel(R) Core(TM) i5-1135G7 @ 2.40GHz
(8 CPUs), ~2.4GHz
- b) RAM 8 GB.

2. Perangkat Lunak:

- a) Sistem operasi Windows 10 Home Single Language.
- b) Software jupyter notebook
- c) Google Form
- d) Pythonanywhere
- e) Microsoft Excel

f) RapidMiner

3.2.2 Bahan penelitian

Bahan penelitian yaitu dari data akademik IF ITTP untuk data mahasiswa. Kemudian cara untuk memperoleh bahan penelitian meliputi:

Observasi

Pengamatan dilaksanakan dengan melakukan observasi terhadap objek penelitian, yaitu mahasiswa Program Studi Teknik Informatika di Institut Teknologi Telkom Purwokerto (ITTP). Observasi juga dilakukan dengan menanyakan data mahasiswa Teknik Informatika ITTP kepada Akademik Prodi Informatika.

Studi literatur

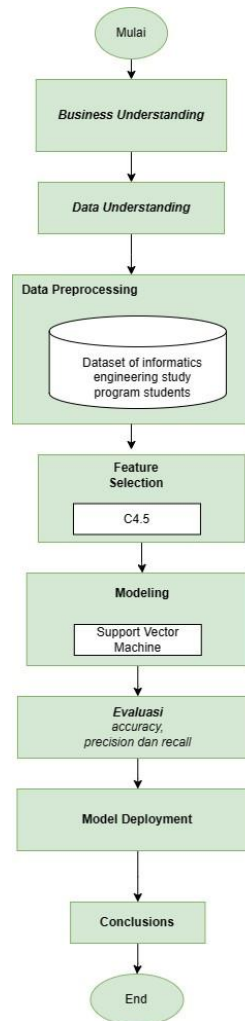
Bahan penelitian berasal dari studi literatur dalam perolehan data dan informasi dengan mengumpulkan dan membaca jurnal, penelitian sebelumnya, dan karangan ilmiah mengenai permasalahan yang berkaitan dengan penelitian.

3.3 Diagram Alir Penelitian

Alur penelitian ini menggunakan *framework Cross-Industry Standard Process for Data Mining* atau CRISP-DM. Pada tahap pertama *business understanding*. Pada tahap ini, identifikasi masalah dilakukan dengan tujuan untuk memahami persoalan yang ingin diselesaikan. Masalah yang diidentifikasi adalah klasifikasi kinerja mahasiswa menjadi dua kategori, yaitu Lulus Tepat Waktu (LTW) dan Lulus Tidak Tepat Waktu (LTTW). Studi literatur juga dilakukan untuk memahami pendekatan yang telah digunakan dalam penelitian sebelumnya terkait topik ini. Kemudian *data understanding*. Tahap ini melibatkan pengumpulan data tentang mahasiswa prodi teknik informatika dari Prodi Informatika ITTP. Data tersebut harus dianalisis untuk memahami karakteristiknya, seperti jenis atribut yang ada, tipe data, dan distribusi

nilai. Tahap selanjutnya adalah *preprocessing data* atau mempersiapkan data untuk proses pemodelan lebih lanjut. Pada tahap ini, algoritma C4.5 digunakan untuk memilih atribut-atribut yang akan diikutsertakan dalam pembentukan model prediksi, sehingga atribut yang kurang relevan atau redundan dapat dihapuskan atau diabaikan. Proses pemilihan atribut ini penting untuk memperoleh model prediksi yang lebih efisien dan akurat. Pada tahap ini juga dilakukan pembersihan data, penghilangan data yang hilang, penanganan outlier, dan pengubahan format data jika diperlukan. Dalam kasus ini, dataset mahasiswa prodi teknik informatika akan diolah untuk memastikan integritasnya dan mempersiapkannya untuk tahap modeling. Lalu, *dataset* yang diperoleh akan dipisahkan menjadi data latih dan data uji menggunakan *10-Fold Cross Validation*. Nilai k diambil 10 fold. Selanjutnya tahap *modelling* dilakukan pemodelan menggunakan algoritma *Support Vector Machine (SVM)*. Setelah model terlatih, evaluasi dilakukan untuk mengevaluasi performa model. Evaluasi dilakukan dengan menggunakan metrik-metrik seperti akurasi (*accuracy*), presisi (*precision*) dan *recall*. Metrik-metrik ini memberikan gambaran tentang seberapa baik model dapat mengklasifikasikan mahasiswa ke dalam kategori yang benar. Evaluasi membantu menentukan keefektifan dan keandalan model yang telah dibangun. Setelah model dikembangkan dan dinilai, langkah selanjutnya adalah *deploy* model untuk digunakan. Pada penelitian ini, model akan diimplementasikan menggunakan *framework Flask* berbasis website. Hal ini memungkinkan penggunaan model yang telah dikembangkan secara praktis dan mudah diakses. Tahap terakhir adalah menyusun kesimpulan berdasarkan hasil penelitian yang telah dilakukan. Kesimpulan ini berisi tentang keberhasilan model dalam mengklasifikasikan kinerja mahasiswa, faktor-faktor yang mempengaruhi kinerja mahasiswa, serta implikasi dari hasil penelitian ini. Selain itu, saran dapat diberikan untuk pengembangan lebih lanjut dalam penelitian klasifikasi kinerja mahasiswa menggunakan metode SVM.

Diagram alir penelitian atau tahapan penelitian pada Gambar 3.1 sebagai berikut:



Gambar 3. 1 Diagram alir penelitian

3.4 Teknik Pengumpulan data

Pengumpulan data dilakukan dengan mengambil dataset langsung ke akademik ITTP secara private dengan menggunakan Surat Permintaan Data Tugas Akhir / Skripsi. Data tersebut diperoleh dari tahun 2017 hingga 2019. Tabel 3.1 berisi deskripsi dari dataset yang akan digunakan.

Tabel 3. 1 Atribut dan Deskripsi *dataset* mahasiswa

Atribut	Deskripsi
Jumlah SKS	Atribut ini mencatat total jumlah SKS (Satuan Kredit Semester)
Mata kuliah Agama	Nilai mata kuliah Agama.
Mata kuliah Bhs Indonesia	Nilai mata kuliah Bhs indonesia.
Mata kuliah Kewarganegaraan dan pancasila	Nilai mata kuliah Kewarganegaraan dan Pancasila.
IP semester 1	Nilai Indeks Prestasi mahasiswa pada semester pertama.
IP semester 2	Nilai Indeks Prestasi mahasiswa pada semester kedua.
IP semester 3	Nilai Indeks Prestasi mahasiswa pada semester ketiga.
IP semester 4	Nilai Indeks Prestasi mahasiswa pada semester keempat.
IP semester 5	Nilai indeks Prestasi mahasiswa pada semester kelima.
IP semester 6	Nilai Indeks Prestasi mahasiswa pada semester keenam.
Nilai D	Total nilai D
Nilai E	Total nilai E
Lulus Tepat Waktu (LTW) dan Lulus Tidak Tepat Waktu (LTTW)	output class [1: LTW, 0: LTTW]

Dalam kasus klasifikasi kinerja mahasiswa IPS dapat digunakan sebagai atribut penting untuk memprediksi dan membedakan kinerja mahasiswa. Mahasiswa dengan IPS tinggi cenderung memiliki kinerja akademik yang baik, sementara mahasiswa dengan IPS rendah dapat mengindikasikan kinerja akademik yang tidak memadai. Jumlah SKS dapat memberikan gambaran tentang seberapa banyak materi yang telah dipelajari oleh mahasiswa. Informasi ini dapat membantu dalam

memprediksi apakah mahasiswa akan memenuhi syarat untuk mengambil Tugas Akhir 1 dan Tugas Akhir 2 berdasarkan persyaratan minimal SKS yang harus dicapai. Kemudian, Nilai dari mata kuliah ini mencerminkan kinerja akademik mahasiswa dalam menguasai dan memahami materi yang berkaitan dengan Pancasila, Agama, Bahasa Indonesia, dan Kewarganegaraan. Selain itu, Syarat untuk lulus mata kuliah Pancasila, Agama, Bahasa Indonesia, dan Kewarganegaraan minimal lulus C. Oleh karena itu, dengan menggunakan atribut nilai mata kuliah ini, penelitian dapat mengevaluasi apakah mahasiswa telah memenuhi syarat untuk mengambil Tugas Akhir 1 dan Tugas Akhir 2. Dataset yang sudah terkumpul masih berupa data mentah mahasiswa prodi Teknik Informatika angkatan 2017 meliputi NIM, Nama, Jenis Kelamin dan IP semester 1 hingga 6 seperti pada Gambar 3.2.

NIM	Nama	Jenis Kelamin	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	Variabel Target
0		PRIA	2.89	2.94	3.10	2.98	3.02	3.69	LTW
1		PRIA	3.11	2.94	2.95	3.45	3.54	3.64	LTW
2		PRIA	3.44	2.94	3.50	3.68	3.29	3.33	LTW
3		PRIA	2.61	2.61	2.98	3.30	3.38	3.77	LTW
4		PRIA	3.06	3.17	3.37	3.33	3.57	3.83	LTW
...	
167		WANITA	2.94	3.00	3.55	3.61	3.60	3.94	LTW
168		PRIA	3.22	3.72	3.80	3.78	3.95	3.52	LTW
169		PRIA	3.11	3.39	3.76	3.65	3.32	3.96	LTW
170		PRIA	2.61	2.44	3.38	3.15	3.48	3.59	LTW
171		PRIA	3.11	2.33	3.62	3.43	3.28	3.37	LTW

Gambar 3. 2 Sample Data mahasiswa prodi informatika

3.5 Preprocessing Data

Preprocessing data adalah langkah-langkah yang diperlukan untuk mempersiapkan data sebelum dilakukan proses klasifikasi menggunakan metode C4.5 dan Support Vector Machine (SVM). Berikut adalah tahapan preprocessing data yang dapat dilakukan untuk kasus klasifikasi kinerja mahasiswa menggunakan metode C4.5 dan SVM:

3.5.1 Pembersihan data (*Data cleaning*)

Apabila terdapat duplikasi atau data ganda, langkah pembersihan data dilakukan dengan menghapus data tersebut.

Proses ini melibatkan pengecekan berdasarkan Nomor Induk Mahasiswa (NIM), dan apabila terdapat entri yang sama, salah satu dari data tersebut dihapus. Selain itu, dilakukan pula pengecekan terhadap tipe data dan adanya nilai null.

3.5.2 Transformation

3.5.2.1 Pengelompokan nilai atau binning

Data asli atau nilai kontinu dipartisi ke dalam beberapa interval atau kelas kategori diskrit berdasarkan kisaran nilai berdasarkan aturan program studi. Atribut mata kuliah Pancasila, Agama, Bhs Indonesia, dan Kewarganegaraan akan dikelompokkan seperti pada Tabel 3.2

Tabel 3.2 Pengelompokan nilai

Nilai	Range
A	$\geq 80,00$
AB	75,00 - 79,9
B	65,00 - 74,99
BC	60,00 - 64,49
C	50,00 – 59,99
D	40,00 – 49,99
E	$\leq 39,99$

3.5.2.2 One-Hot Encoding

Proses jenis akan dikelompokkan menjadi data yang bersifat kategori. Mengubah variabel target "Lulus Tepat Waktu (LTW)" dan "Lulus Tidak Tepat Waktu (LTTW)" menjadi bentuk . Hal ini diperlukan karena metode klasifikasi C4.5 dan SVM membutuhkan label dalam bentuk numerik serta membagi data yang akan diolah menjadi data training dan data testing. Tabel 3.3 merupakan tabel pelabelan Variabel Target.

Tabel 3.3 Tabel Pelabelan Variabel Target

Variabel Target	LTW	1
	LTTW	0

3.5.3 Pemilihan atribut

Menggunakan algoritma C4.5 untuk menghitung information gain pada setiap atribut pada dataset, yang mengukur sejauh mana atribut tersebut memberikan informasi yang bermanfaat dalam memprediksi kelas target. Memilih atribut dengan information gain tertinggi sebagai root node untuk pohon keputusan.

3.5.4 Hasil Dataset Baru

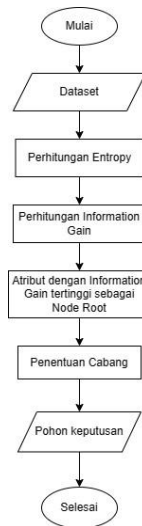
Setelah penyelesaian tahap preprocessing, dataset baru akan dihasilkan dan telah disesuaikan dengan kebutuhan penelitian, sebagaimana terlihat pada Gambar 3.3 Dataset ini akan digunakan sebagai bahan untuk menguji performa algoritme klasifikasi yang dipakai.

NIM	Nama	Jenis Kelamin	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	Variabel Target
0		1	2.89	2.94	3.10	2.98	3.02	3.69	1
1		1	3.11	2.94	2.95	3.45	3.54	3.64	1
2		1	3.44	2.94	3.50	3.68	3.29	3.33	1
3		1	2.61	2.61	2.98	3.30	3.38	3.77	1
4		1	3.06	3.17	3.37	3.33	3.57	3.83	1
...	
167		0	2.94	3.00	3.55	3.61	3.60	3.94	1
168		1	3.22	3.72	3.80	3.78	3.95	3.52	1
169		1	3.11	3.39	3.76	3.65	3.32	3.96	1
170		1	2.61	2.44	3.38	3.15	3.48	3.59	1
171		1	3.11	2.33	3.62	3.43	3.28	3.37	1

Gambar 3. 3 Dataset setelah Melewati Tahap Preprocessing

3.6 Alur Algoritma C4.5

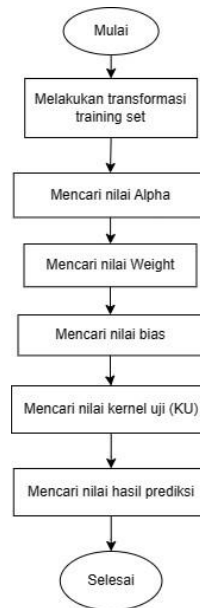
Alur algoritme C4.5 dapat dilihat pada Gambar 3.4



Gambar 3. 4 Alur algoritme C4.5

Dataset yang sudah ada dilakukan perhitungan entropy total dan entropy masing-masing jawaban dari setiap atribut. Selanjutnya menghitung information gain untuk setiap atribut pada dataset, yang mengukur sejauh mana atribut tersebut memberikan informasi yang bermanfaat dalam memprediksi kelas target. Memilih atribut dengan information gain tertinggi sebagai root node untuk pohon keputusan. lalu, dilakukan perhitungan ulang dengan cara yang sama untuk menentukan cabang hingga terbentuklah pohon keputusan.

3.7 Alur Algoritma Support Vector Machine (SVM)



Gambar 3. 5 Alur algoritma SVM


Gambar 3.5 menunjukkan proses awal algoritma SVM, yang diawali dengan mengambil data training set dan melakukan transformasi atau preprocessing jika diperlukan. Ini bisa mencakup pemilihan fitur, normalisasi data, atau pemrosesan lainnya agar data siap digunakan untuk proses pembelajaran. Lalu mencari nilai Alpha. Nilai Alpha penting dalam menentukan vektor-vektor pendukung (support vectors) yang membentuk batas keputusan. Kemudian menghitung bobot (weight) yang digunakan untuk menggabungkan kontribusi setiap sampel data dalam pembentukan batas keputusan. Bobot ini akan mempengaruhi bentuk dan posisi batas keputusan yang dihasilkan oleh model SVM.

Tahap selanjutnya adalah menghitung bias yang merupakan konstanta yang mempengaruhi posisi batas keputusan. Bias membantu menentukan apakah sebuah sampel data akan diklasifikasikan sebagai kelas positif atau kelas negatif. Menghitung kernel antara setiap sampel data uji dengan setiap sampel data latih yang telah dijadikan support vectors. Kernel uji digunakan untuk menghitung kemiripan antara sampel data uji dengan data latih yang penting dalam proses prediksi.

Berdasarkan nilai Alpha, Weight, bias, dan kernel uji, menghitung hasil prediksi untuk setiap sampel data uji. Hasil prediksi akan menentukan kelas atau label yang diberikan untuk setiap sampel data uji. Kemudian Algoritma SVM akan menghasilkan prediksi kinerja mahasiswa berdasarkan model yang telah dibangun.

3.8 Model Deployment

Setelah pengujian model maka, model bisa disimpan dengan format *.Joblib* untuk penggunaan website. Gambar 3.5 menunjukkan rancangan *user interface* ketika model di *deploy*. Untuk membuat website menggunakan *framework* Flask menggunakan bahasa pemrograman Python. Selain Python, juga menggunakan bahasa markup HTML (*Hypertext Markup Language*) untuk membuat struktur dan konten halaman *website*, bahasa stylesheet CSS (*Cascading Style Sheets*) Untuk mengatur tampilan serta gaya halaman *website* dan JavaScript untuk memberikan interaktivitas dan logika tambahan pada website.



The image shows a web form titled "Prediksi Kinerja Mahasiswa". At the top left, there is a label "Jenis Kelamin" above a dropdown menu. Below this, there are six input fields, each with a label "IPS 1" through "IPS 6" to its left. At the bottom center of the form, there is a button labeled "Submit and Predict".

Gambar 3. 6 Tampilan halaman pengisian data