

BAB I PENDAHULUAN

I.1. State Of The Art

Seiring dengan meningkatnya kompleksitas data pendidikan tinggi, kebutuhan akan teknik analitik yang mampu mengolah dan mengekstraksi informasi bermakna dari data mahasiswa semakin meningkat. Berbagai metode data mining telah dikembangkan dan diimplementasikan untuk memprediksi kinerja akademik mahasiswa, seperti Support Vector Machine (SVM), Naïve Bayes, Decision Tree, dan Neural Network. Masing-masing metode memiliki kekuatan dan kelemahan tergantung pada jenis dan kompleksitas data yang dianalisis.

Penelitian terbaru yang dilakukan oleh Bindhia K. Francis dan Suvanam Sasidhar Babu (2019) menandai sebuah kemajuan penting dalam pendekatan prediksi kinerja akademik mahasiswa melalui penggabungan teknik klasifikasi dan klusterisasi dalam satu kerangka hybrid. Dalam studi tersebut, algoritma klasifikasi (SVM, Naïve Bayes, Decision Tree, dan Neural Network) digunakan bersama dengan algoritma K-Means Clustering dan pendekatan majority voting untuk meningkatkan akurasi prediksi. Hasil penelitian menunjukkan bahwa pendekatan hybrid ini secara signifikan meningkatkan akurasi prediksi hingga mencapai 75,47%, yang lebih unggul dibandingkan metode klasifikasi tunggal.

Studi ini juga menekankan pentingnya pemilihan fitur (atribut) yang relevan dari data mahasiswa, yang mencakup fitur demografis, akademik, perilaku, dan ekstra (seperti kepuasan orang tua dan absensi). Analisis menunjukkan bahwa kombinasi fitur perilaku dan ekstra memberikan hasil prediksi yang paling akurat, mengindikasikan bahwa performa akademik tidak hanya dipengaruhi oleh faktor akademik semata, tetapi juga oleh karakteristik perilaku dan dukungan sosial.

Kelebihan dari pendekatan hybrid ini terletak pada kemampuannya mengintegrasikan kekuatan klasifikasi dalam mengenali pola label serta kekuatan klusterisasi dalam mengelompokkan data yang bersifat heterogen. Dengan demikian, pendekatan ini dinilai lebih adaptif dalam menangani data mahasiswa yang multidimensi dan bervariasi. Di samping itu, penelitian ini juga

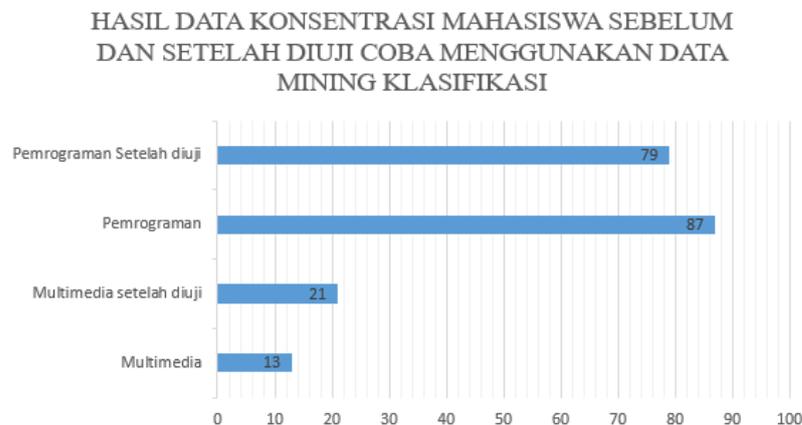
mengimplementasikan metode evaluasi yang komprehensif melalui pengukuran precision, recall, f-score, dan accuracy, yang menjadikan hasil evaluasi lebih objektif dan dapat dibandingkan secara luas.

Sebagai salah satu kontribusi penting dalam Data Mining, penelitian ini membuka peluang bagi pengembangan sistem pendukung keputusan berbasis data yang lebih cerdas dan proaktif dalam konteks pendidikan tinggi. Dengan adanya simulasi hybrid, institusi pendidikan dapat mengidentifikasi mahasiswa berisiko rendah secara lebih awal dan menerapkan intervensi yang lebih tepat sasaran.

I.2. Latar Belakang Masalah

Penentuan konsentrasi jurusan di tingkat Perguruan Tinggi merupakan suatu proses yang kompleks bagi mahasiswa. Dalam menentukan konsentrasi, mahasiswa cenderung mempertimbangkan minat pribadi serta kesesuaian dengan bidang keilmuannya. Ketepatan dalam memilih konsentrasi sangat penting, karena akan memengaruhi prospek karier dan peluang kerja yang diminati setelah mahasiswa menyelesaikan pendidikannya. (Erwandi dkk., 2022).

Dalam menentukan konsentrasi studi, mahasiswa diharapkan mampu menunjukkan performa akademik yang optimal serta memahami kompetensi yang dimilikinya. Pemilihan konsentrasi yang tepat akan berdampak pada meningkatnya minat belajar yang selaras dengan bidang keilmuan, serta mendorong keaktifan dan ketekunan dalam menyelesaikan tugas maupun menghadapi ujian selama perkuliahan, yang pada akhirnya berkontribusi terhadap pencapaian Indeks Prestasi Kumulatif (IPK) yang maksimal.. Seperti pada data yang dalam penelitian. Terlihat bahwa masih banyak mahasiswa yang salah mengambil konsentrasi dalam penjurusan jika dilihat dari nilai akademik mahasiswa tersebut. Berikut adalah gambar yang menunjukkan hasil penelitian yang mengenai mahasiswa yang salah mengambil konsentrasi jurusan (Atiroh, 2022).



Gambar I. 1 Data hasil penelitian mahasiswa yang salah mengambil konsentrasi jurusan

Dalam gambar tersebut dijelaskan bahwa jika dilihat dari nilai mata kuliah pada 100 mahasiswa, selisih yang salah mengambil jurusan multimedia sebelum dan setelah diuji menggunakan Data Mining klasifikasi sekitar 5% dengan jumlah 13 data mahasiswa sebelum diuji menjadi 21 data mahasiswa setelah diuji, begitu juga pada konsentrasi pemrograman sekitar 3% dengan jumlah 87 data mahasiswa sebelum diuji menjadi 79 data mahasiswa setelah diuji. Jika mahasiswa tersebut salah dalam memilih konsentrasi jurusan maka akan menyebabkan dampak pada mahasiswa itu sendiri seperti kehilangan semangat belajar karena tidak paham dengan materi tersebut yang menyebabkan keterlambatan dalam kelulusan mahasiswa tersebut (Ayu & Saryanti, 2019).

Penelitian lain yang mengkaji dibidang Pendidikan khususnya mengenai data mahasiswa telah dilakukan oleh (Nikmatun & Waspada, 2019) dalam penelitian ini pengklasifikasian menggunakan algoritma KNN. Hasil penelitian menunjukkan bahwa hasil klasifikasi masa studi terbaik diperoleh dengan memilih atribut dari semua mata kuliah pilihan dengan nilai akurasi 75,95% dengan jumlah data sebanyak 377 orang dengan 72 atribut nilai mata kuliah. (Hidayanti dkk., 2020) dua algoritma tersebut dibandingkan yaitu C4.5 dan naïve bayes dengan menggunakan 299 data hasil akurasi membuktikan bahwa C4.5 48,06% memiliki Tingkat akurasi paling tinggi daripada naïve bayes yang hanya 42,79%. Selain Teknik klasifikasi terdapat Teknik klastering yang melakukan analisis pada sektor Pendidikan yaitu

(Ramdani, 2023) menjelaskan dalam penelitiannya bahwa Pengklasteran Mahasiswa Pelamar Beasiswa Dengan Cubic Clustering Criterion. Hasil yang di dapat dari penelitian ini adalah dari hasil perhitungan terdapat tiga buah pengelompokan untuk data pengajuan beasiswa, yaitu kelompok yang tidak diterima, kelompok di pertimbangkan menerima beasiswa dan kelompok penerima beasiswa. Kemudian (Ramdani, 2023) melakukan perbandingan antara algoritme k-means dan k-medoids pada klasterisasi data akademik mahasiswa dengan data sejumlah 483 dari angkatan 2017 hingga 2020 yang terdiri dari 11 atribut. Komparasi hasil klasterisasi kedua algoritme tersebut menunjukkan bahwa k-medoids lebih unggul dengan hasil 97%.

Namun menurut (Dewi Eka Putri & Eka Praja Wiyata Mandala, 2022) menjelaskan dalam jurnalnya penggunaan metode Hybrid antara teknik klasterisasi dan klasifikasi memberikan pengaruh yang signifikan dalam menentukan hasil akurasi. Begitupun dengan (Francis & Babu, 2019) menurutnya teknik klasifikasi dan pengelompokan pada penelitiannya diuji secara i dengan kumpulan data mahasiswa dari berbagai disiplin ilmu di institusi pendidikan tinggi di Kerala, India. Hasilnya membuktikan bahwa algoritma Hybrid yang menggabungkan pendekatan klastering dan klasifikasi memberikan hasil yang jauh lebih unggul dalam hal mencapai akurasi dalam prediksi kinerja akademik mahasiswa. Adapun penjelasan (Khosravi dkk., 2020) mengenai Hybrid Data Mining tentang prediksi transportasi pengangkutan muatan dasar di dasar sungai kerikil menunjukkan bahwa dengan penggabungan teknik klasterisasi dan klasifikasi jauh lebih akurat dibandingkan hanya menggunakan satu teknik Data Mining.

Simulasi hybrid data mining memungkinkan para peneliti dan praktisi untuk menguji efektivitas kombinasi metode pada dataset mahasiswa secara empiris, sebelum diterapkan pada sistem pengambilan keputusan yang sesungguhnya. Dengan melakukan simulasi ini, dapat diperoleh pemahaman yang lebih mendalam mengenai sejauh mana model hybrid mampu memberikan prediksi yang lebih akurat dan efisien dibandingkan pendekatan konvensional.

Permasalahan utama yang mendasari penelitian ini adalah bagaimana merancang dan mengevaluasi simulasi model hybrid data mining yang dapat mengolah data mahasiswa secara komprehensif, serta bagaimana mengukur tingkat efektivitasnya dalam mengungkap pola dan melakukan prediksi akademik. Selain itu, penelitian ini juga berusaha menjawab pertanyaan apakah pendekatan hybrid mampu memberikan keunggulan signifikan dibandingkan dengan metode data mining tunggal dalam konteks pendidikan tinggi.

Oleh karena itu, penelitian ini menjadi penting untuk dilakukan guna memberikan kontribusi terhadap pengembangan sistem pendukung keputusan berbasis data mining di lingkungan perguruan tinggi, serta untuk membantu pihak manajemen kampus dalam merancang strategi pembelajaran dan intervensi yang lebih tepat sasaran.

I.3 Rumusan Masalah

Penentuan konsentrasi mahasiswa tidak mudah. Hal ini akan berpengaruh terhadap kegiatan akademik mahasiswa terlebih dalam waktu kelulusan yang sangat berpengaruh. Pemilihan konsentrasi yang asal-asalan tanpa pertimbangan yang matang, menyebabkan dampak negatif pada mahasiswa, yaitu kesulitan dalam penyerapan materi-materi perkuliahan. Tidak jarang mahasiswa yang mengambil konsentrasi karena terpengaruh oleh teman tanpa melihat kemampuannya. Sehingga pada akhirnya tidak jarang yang berpindah konsentrasi jurusan jika tidak sesuai kemampuan. Proses berpindahnya konsentrasi cukup memakan waktu dikarenakan kegiatan belajar mengajar telah dimulai dan proses batal tambah mata kuliah yang telah dipilih memiliki range waktu tidak banyak. Oleh sebab itu perlu metode khusus yang dapat digunakan mahasiswa dalam menentukan konsentrasi mahasiswa.

Metode klasifikasi dengan memanfaatkan data nilai mahasiswa saat ini masih dihadapkan pada tantangan dalam mencapai tingkat akurasi yang optimal. Faktor-faktor seperti kompleksitas data mahasiswa, interaksi antarfitur, dan salah satu faktor yang paling utama adalah penggunaan metode pemrosesan data dapat

mempengaruhi performa algoritma secara signifikan. Pemilihan metode pemrosesan data yang tepat menjadi faktor penting dalam meningkatkan performa algoritma. Penggunaan metode yang tidak sesuai dapat menghasilkan informasi yang tidak relevan atau bahkan mengabaikan faktor-faktor penting dalam prediksi data mahasiswa. Dari penjelasan tersebut maka Penelitian ini menyarankan penggabungan antara metode klasterisasi dan metode klasifikasi. Metode klasterisasi digunakan untuk mengelompokkan nilai mahasiswa menjadi beberapa kelompok. Proses klasterisasi dilakukan dengan menggunakan algoritma k-medoid. Setelah diperoleh kelompok-kelompok nilai mahasiswa tersebut, selanjutnya dilakukan proses klasifikasi untuk menentukan konsentrasi jurusan. Klasifikasi ini dilakukan dengan metode klasifikasi yang ditujukan untuk menghasilkan hubungan yang valid antara variabel input dan variabel target dengan tujuan untuk memberikan rekomendasi terkait membantu proses pengambilan keputusan mahasiswa dalam memilih konsentrasi atau peminatan yang dilihat dari nilai mahasiswa tersebut.

I.4 Tujuan Penelitian

Sejalan dengan rumusan masalah yang telah dipaparkan sebelumnya. Maka tesis ini disusun dengan tujuan khusus pada penelitian ini antara lain

1. Melakukan simulasi dan evaluasi kinerja algoritma *hybrid* dengan membandingkan hasil akurasi, *specificity* dan *sensitivity* dari model *hybrid* terhadap model klasifikasi tunggal seperti *Random Forest*, *k-nearest neighbours (KNN)*, *Naïve Bayes* dan *Decision Tree*.
2. Mengembangkan model *hybrid data mining* yang menggabungkan metode klasifikasi dan klasterisasi untuk menganalisis data mahasiswa guna meningkatkan akurasi prediksi terhadap kinerja akademik.

I.5 Manfaat Penelitian

Berikut ini merupakan manfaat dari penelitian ini berdasarkan tujuan penelitian:

1. Penelitian ini dapat memberikan kontribusi dalam bidang pendidikan dengan pembangunan model data mining.

2. Pengembangan model berbasis data mining untuk klasifikasi data mahasiswa dapat mengarah pada pengembangan teknologi deteksi yang lebih efektif.
3. mahasiswa akan diuntungkan melalui pemberian informasi yang lebih akurat dan dapat dipercaya. Dengan berkurangnya jumlah mahasiswa yang salah pilih konsentrasi, mahasiswa dapat membuat keputusan yang lebih baik.

I.6 Pertanyaan Penelitian

Berdasarkan tujuan penelitian yang telah disampaikan pada bagian sebelumnya, maka dapat ditentukan pertanyaan dari penelitian ini yaitu :

1. Apakah kombinasi algoritma klasifikasi dalam pendekatan simulasi *hybrid data mining* dapat meningkatkan akurasi prediksi prestasi akademik mahasiswa dibandingkan dengan penggunaan algoritma tunggal?
2. Apakah terdapat hubungan yang signifikan antara atribut-atribut mahasiswa dengan hasil klasifikasi yang dihasilkan oleh model simulasi *hybrid data mining*?

I.7 Kontribusi Penelitian

Penelitian ini diharapkan dapat berkontribusi memberikan pemahaman yang lebih mendalam tentang bagaimana memutuskan dalam memilih konsentrasi. Selain itu bagi mahasiswa dapat meningkatkan kepercayaan mengenai pemilihan konsentrasi sehingga akan semakin yakin dalam mengambil keputusan pemilihan konsentrasi multimedia atau pemrograman. Kontribusi ini dapat memberikan wawasan baru terhadap faktor-faktor yang memotivasi mahasiswa dalam membuat keputusan. Kontribusi ini dapat menjadi landasan untuk pengembangan lebih lanjut dalam teknologi deteksi yang lebih efisien dan adaptif.

I.8 Lingkup Penelitian

Adapun lingkup dari penelitian ini yaitu mencakup pada batasan dan jangkauan penelitian yang terdiri dari aspek ruang lingkup masalah, lokasi dan objek penelitian, serta waktu dan periode. Berikut merupakan penjelasan masing-masing aspek dari Lingkup Penelitian.

I.8.1 Ruang Lingkup Masalah

Pada aspek Ruang Lingkup Masalah dalam penelitian ini berfokus pada bagaimana pengaruh nilai akurasi jika menggabungkan teknik klastering dan teknik klasifikasi. Selain itu penelitian ini mengacu bagaimana cara melakukan Hybrid Data Mining untuk menganalisis data supaya penggunaan Hybrid Data Mining dapat menghasilkan tingkat ketepatan dalam klasifikasi yang tinggi sehingga metode Hybrid Data Mining yang digunakan sudah bisa menjadi salah satu alternatif solusi dalam melakukan klasifikasi data. Dalam proses analisa nilai mahasiswa untuk menentukan konsentrasi jurusan dengan algoritma k-medoids dan model algoritma klasifikasi.

I.8.2 Lokasi dan Objek Penelitian

Lokasi pada penelitian ini mengacu pada tempat penelitian yang dilakukan di Universitas yang berada di Serang, Banten dengan nilai mahasiswa sebagai objek penelitian yang dipilih.

I.8.3 Waktu dan Periode

Penelitian ini dilakukan pada kurun waktu dua tahun lamanya sejak tahun 2022 hingga 2025 dengan memperhatikan konsep analisis Hybrid terhadap pengembangan teknik klastering dan teknik klasifikasi

I.9 Signifikansi Penelitian

Dari tinjauan literatur sistematis yang dilakukan, diamati bahwa ada penelitian yang menggabungkan teknik klastering dan teknik klasifikasi Tujuan utamanya adalah untuk membantu pengambilan keputusan dan menambah analitik cerdas. Penelitian ini akan memberi signifikansi mengenai pengetahuan dalam melakukan teknik penggabungan pengelompokan dan klasifikasi pada fitur akademik, perilaku, dan tambahan kumpulan data siswa dan terbukti lebih unggul dibandingkan algoritme lain yang ada. Model ini dapat membantu pendidik dalam mempersepsikan peserta didik mengenali peserta didik yang lemah untuk mengembangkan proses pembelajaran dan mengurangi tingkat kegagalan akademik dan juga membantu administrator untuk mengelola sistem berdasarkan hasil pembelajaran dengan lebih baik.

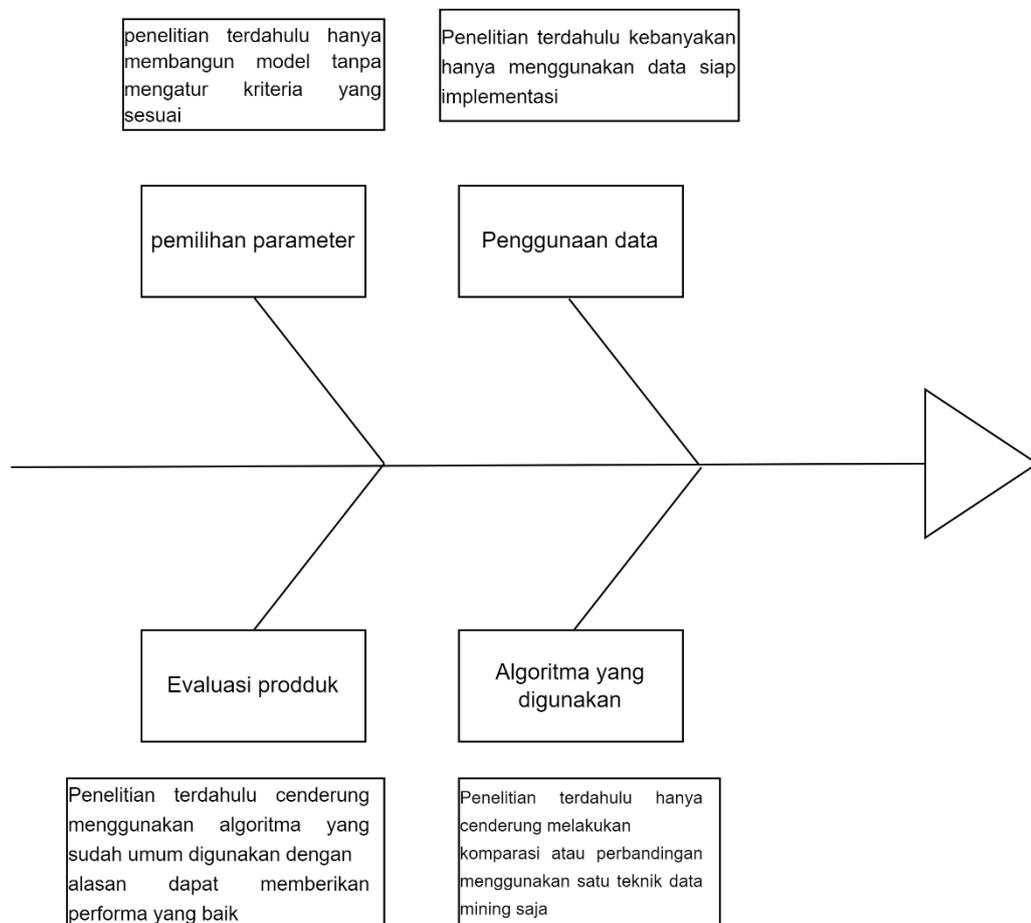
I.10 Peran Penelitian

Dalam penelitian ini, peneliti memiliki peran dalam mengidentifikasi, menganalisis, dan merumuskan penyelesaian masalah pada proses pengembangan model dengan metode klasifikasi yang dipilih. Sebagai peneliti pada bidang Data Mining, peran peneliti sangat penting dalam melakukan upgrading dan eksplorasi terhadap metode yang digunakan untuk memberikan hasil akurasi model yang lebih baik dan pemetaan terhadap data yang tepat. Komitmen dalam menganalisis potensi, tantangan, dan kebutuhan yang ada pada proses Data Mining, serta menerapkan solusi dari permasalahan yang berkelanjutan untuk meningkatkan kualitas hasil yang diperoleh baik model maupun data yang diperoleh dari kedua metode yang digunakan tersebut. Sebagai langkah dasar dalam penelitian, peneliti melakukan pengumpulan data dan informasi terkait isu dan masalah dengan mewawancarai ketua prodi yang menjadi fokus penelitian. Selanjutnya, tahap analisis terhadap data yang telah dikumpulkan guna mengidentifikasi data mana yang layak untuk masuk kedalam proses selanjutnya, melakukan pra proses data guna mengidentifikasi nilai pada mata kuliah mana yang berhubungan dengan konsentrasi multimedia dan pemrograman.

Dengan peran topik penelitian Hybrid Data Mining berdasarkan klasterisasi nilai mahasiswa untuk klasifikasi konsentrasi jurusan, diharapkan dapat memberikan insight dalam pengembangan metode yang digunakan dalam proses Data Mining berlangsung sehingga dapat memberikan performa model yang maksimal. Maka dari itu, peneliti merupakan kunci utama dari aktivitas penelitian ini.

I.11 Kesenjangan Penelitian

Pada tahap ini menggunakan gap analysis dimana pertanyaan atau masalah yang belum dijawab dengan tepat atau sama sekali belum terjawab. metode dari gap analysis pada penelitian ini menggunakan fish bone diagram sebagai metode untuk menganalisis kesenjangan dalam penelitian.



Gambar I. 2 Fish Bone Diagram Gap Analysis

Dari gambar diatas dapat diambil informasi bahwa terdapat empat poin utama yang bisa diidentifikasi sebagai gap yang berkaitan dengan topik dalam penelitian ini. Yang pertama adalah kebanyakan peneliti terdahulu menggunakan data yang siap proses yang diperoleh dari open source data sehingga tidak menjelaskan bagaimana proses extracting seperti yang telah dijelaskan pada metode CRISP-DM. Kedua, para peneliti terdahulu cenderung melakukan komparasi atau perbandingan menggunakan satu teknik Data Mining saja, padahal dengan menggabungkan dua teknik lebih efisien. Yang ketiga, peneliti terdahulu cenderung hanya membangun model tanpa mengatur parameter yang sesuai, hal ini dapat memberikan dampak yang signifikan pada bagaimana model belajar dari data dan bagaimana ia melakukan hasil prediksi yang tidak akurat. Dan yang terakhir, Penelitian terdahulu cenderung menggunakan algoritma yang sudah umum digunakan dengan

alasan dapat memberikan performa terbaik, nyatanya semua algoritma dapat memberikan hasil yang baik tergantung apakah sudah benar dalam prosesnya.

I.12 Rasionalisasi Penelitian

Rasionalisasi penelitian ini dilakukan bertujuan untuk memberikan solusi kritis dan mendalam dalam meningkatkan performa kinerja klasifikasi dalam uji akurasi. Dengan memperhatikan pendekatan dan konsep yang tepat, maka penelitian ini bertujuan untuk mengetahui pengaruh terhadap nilai akurasi dengan menggunakan model klasifikasi ketika ada model algoritma yang berbeda membantu dalam memprediksi secara akurat kinerja mahasiswa.

Namun beberapa tantangan dan kendala yang mungkin terjadi seperti banyaknya kesalahan pada data seperti noise data dapat mempengaruhi performa model yang dihasilkan. Oleh karena itu, penelitian ini menjadi relevan karena menawarkan potensi yang dapat mengatasi permasalahan tersebut. Dengan memanfaatkan pendekatan pengembangan Hybrid dapat memberikan hasil akurasi model yang cukup signifikan.

I.13 Sistematika Penulisan

Penyusunan tesis ini dibuat agar mendapatkan gambaran yang jelas secara menyeluruh tentang pembahasan yang diungkapkan, maka menerapkan keterangan menjadi beberapa bab yaitu:

BAB I PENDAHULUAN

Bab ini menggambarkan isi keseluruhan tesis yang terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan.

BAB II KAJIAN PUSTAKA

Merupakan bab yang berisikan uraian teori-teori secara lingkup mengenai dari tinjauan penelitian, dasar teori, dan kerangka pikiran.

BAB III METODOLOGI PENELITIAN

Merupakan bab berisi uraian tentang tahapan-tahapan yang dilalui dalam pembuatan aplikasi ini, mulai dari studi literature dan mendesain animasi yang akan dibuat.

BAB IV USULAN PENELITIAN

Merupakan bab yang berisi uraian tentang langkah analisa penulisan yang dibahas dalam pengertian, deskripsi usulan penelitian, rancangan usulan penelitian, rancangan pengujian, peancangan sistem yang diajukan, implementasi sistem, serta analisa implementasi yang sudah dirancang.

BAB V HASIL DAN PEMBAHASAN

Merupakan bab yang berisi uraian tentang langkah analisa penulisan yang dibahas dalam pengertian, hasil penelitian dan pembahasan penelitian.

BAB VI PENUTUP

Merupakan bab yang berisi uraian tentang kesimpulan dan saran. Kesimpulan berupa pokok dari hasil penelitian dan pembahasan untuk membuktikan dari temuan pustaka yang diperoleh.