TUGAS AKHIR

ANALISIS DATA UNTUK MENGETAHUI HUBUNGAN ANTARA IPK ATAU LAMA STUDI, DAN JALUR MASUK UNPAR

Cevas Bungaran

NPM: 6181801070

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN

2024

FINAL PROJECT

DATA ANALYSIS TO DETERMINE THE RELATIONSHIP BETWEEN GPA OR DURATION OF STUDY, AND ADMISSION PATH AT UNPAR

Cevas Bungaran

NPM: 6181801070

DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY

2024

LEMBAR PENGESAHAN

ANALISIS DATA UNTUK MENGETAHUI HUBUNGAN ANTARA IPK ATAU LAMA STUDI, DAN JALUR MASUK UNPAR

Cevas Bungaran

NPM: 6181801070

Bandung, 25 Juni 2024

Menyetujui,

Pembimbing

Digitally signed by Mariskha Tri Adithia

Mariskha Tri Adithia, P.D.Eng

Ketua Tim Penguji

Digitally signed by Natalia

Natalia, M.Si.

Anggota Tim Penguji

Digitally signed by Rosa de Lima E. Padmowati

Rosa De Lima, M.T.

Mengetahui,

Ketua Program Studi

Digitally signed by Lionov

Lionov, Ph.D.

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa tugas akhir dengan judul:

ANALISIS DATA UNTUK MENGETAHUI HUBUNGAN ANTARA IPK ATAU LAMA STUDI, DAN JALUR MASUK UNPAR

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 25 Juni 2024

Cevas Bungaran
NPM: 6181801070

ABSTRAK

Keberhasilan studi dari seorang mahasiswa menjadi alat ukur untuk mengetahui kemampuan mahasiswa dalam menguasai materi kuliah. Faktor awal yang bisa menjadi pengukur kemampuan akademik dan minat adalah jalur masuk, seperti USM 1, USM 2, USM 3, PMDK, dan Seleksi Khusus. Seringkali mahasiswa dengan jalur masuk tertentu memiliki nilai IPK dan lama studi yang berbeda, contohnya yaitu jalur masuk PMDK memiliki nilai IPK yang lebih tinggi daripada USM 1. Berdasarkan masalah tersebut dapat diasumsikan bahwa IPK atau lama studi memiliki hubungan dengan jalur masuk.

Untuk membuktikan asumsi tersebut, penelitian ini melakukan analisis deskriptif untuk menganalisis hubungan antara IPK, lama studi, dan jalur masuk UNPAR, menganalisis faktor yang berpengaruh terhadap nilai IPK, lama studi, dan jalur masuk, serta melakukan analisis prediktif untuk melakukan prediksi IPK dan prediksi lama studi.

Sebelum melakukan analisis, langkah yang dilakukan pertama adalah melakukan analisis penyelesaian masalah. Terdapat dua teknik yang digunakan untuk melakukan analisis, yaitu teknik statistika dan teknik *data mining*. Teknik statistika menggunakan metode korelasi seperti *Pearson* dan *Chi-Square* bertujuan untuk menganalisis hubungan antara IPK atau lama studi dengan jalur masuk. Korelasi *Pearson* merupakan teknik untuk mengukur korelasi antara dua atribut numerik, korelasi *Chi-Square* merupakan teknik untuk mengukur korelasi antara dua atribut kategorikal. Teknik *data mining* menggunakan algoritma *Clustering* dan Klasifikasi. Kemudian eksplorasi *library* pada *Python* dilakukan untuk mempelajari *library* yang dapat digunakan pada saat melakukan analisis.

Hasil analisis yang diperoleh adalah IPK dan lama studi memiliki hubungan terhadap jalur masuk. Setiap jalur masuk memiliki pola IPK dan lama studi yang negatif, yaitu semakin tinggi nilai IPK maka lama studi akan semakin rendah, dan sebaliknya, semakin rendah nilai IPK maka lama studi akan semakin tinggi. Analisis variabel yang berpengaruh terhadap IPK dan lama studi dilakukan menggunakan korelasi *Chi-Square*. Variabel yang berpengaruh terhadap nilai IPK adalah lama studi, program studi, dan jalur masuk. Variabel yang berpengaruh terhadap lama studi adalah IPK, program studi, provinsi asal SMA, dan jalur masuk.

Model prediksi dibuat menggunakan algoritma *Naive Bayes* dan *Decision Tree*. Algoritma *Naive Bayes* menghasilkan model prediksi lama studi paling baik, dengan nilai RMSE 1.43. Algoritma *Decisioin Tree* menghasilkan model prediksi IPK paling baik, dengan nilai RMSE 0.637. Kedua model diluncurkan menggunakan GUI.

Kesimpulan dari tugas akhir ini adalah IPK dan lama studi memiliki hubungan korelasi yang negatif. Variabel yang memengaruhi nilai IPK adalah lama studi, program studi, dan jalur masuk. Variabel yang memengaruhi lama studi adalah IPK, provinsi asal SMA, program studi, dan jalur masuk. Model yang dihasilkan untuk memprediksi IPK dan lama studi cukup baik, dan dapat diluncurkan menggunakan GUI.

Kata-kata kunci: IPK, lama studi, jalur masuk, statistika, *data mining*.

ABSTRACT

The success of a student's studies serves as a measure to understand their ability to master course material. An initial factor that can gauge academic ability and interest is the admission path, such as USM 1, USM 2, USM 3, PMDK, and Special Selection. Often, students entering through certain pathways have different GPAs and lengths of study; for instance, students admitted through the PMDK pathway tend to have higher GPAs compared to those entering through USM 1. Based on this issue, it can be assumed that GPA or length of study is related to the admission path.

To validate this assumption, this study conducts descriptive analysis to examine the relationship between GPA, length of study, and the admission path to UNPAR. It also analyzes factors influencing GPA, length of study, and admission path, and performs predictive analysis to forecast GPA and length of study.

Before conducting the analysis, the first step is problem-solving analysis. Two techniques are used for this analysis: statistical techniques and data mining techniques. Statistical techniques involve correlation methods such as Pearson and Chi-Square, aimed at analyzing the relationship between GPA or length of study and admission path. Pearson correlation measures the correlation between two numerical attributes, while Chi-Square correlation measures the correlation between two categorical attributes. Data mining techniques use Clustering and Classification algorithms. Additionally, an exploration of Python libraries is conducted to identify libraries that can be used during the analysis.

The analysis results indicate that GPA and length of study are related to the admission path. Each admission path shows a negative pattern between GPA and length of study: the higher the GPA, the shorter the length of study, and vice versa. The analysis of variables affecting GPA and length of study is performed using Chi-Square correlation. The variables affecting GPA are length of study, study program, and admission path. The variables affecting the length of study are GPA, study program, high school province, and admission path.

Predictive models are created using Naive Bayes and Decision Tree algorithms. The Naive Bayes algorithm produces the best predictive model for the length of study, with an RMSE of 1.43. The Decision Tree algorithm produces the best predictive model for GPA, with an RMSE of 0.637. Both models are deployed using a GUI.

The conclusion of this thesis is that GPA and length of study have a negative correlation. Variables affecting GPA include length of study, study program, and admission path. Variables affecting length of study include GPA, high school province, study program, and admission path. The resulting models for predicting GPA and length of study are quite good and can be deployed using a GUI.

Keywords: GPA, length of study, admission selection path, stastitical, data mining.

Saya persembahkan tugas akhir ini untuk Papa dan Mama...

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa karena atas berkat dan anugerah yang diberikan-Nya, penulis dapat menyelesaikan penyusunan tugas akhir yang berjudul "Analisis Data Untuk Mengetahui Hubungan Antara IPK atau Lama Studi, dan Jalur Masuk UNPAR". Selama mengerjakan tugas akhir, penulis menyadari bahwa banyak sekali doa dan dukungan yang penulis dapatkan dari orang tua, saudara, dan teman-teman. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

- Papa dan Mama yang selalu mendoakan, mendukung, dan memberikan kasih sayang kepada penulis.
- Abiezer Tumpal Sahattua Tobing dan Benaya Jonatan Tobing selaku Kakak penulis yang telah memberikan bantuan dalam bentuk doa dan moral.
- Selina Khodry yang selalu mendampingi serta memberikan dukungan dan doa kepada penulis.
- Ibu Mariskha Tri Adithia, P.D.Eng yang telah memberikan bimbingan dan arahan selama proses penyusunan tugas akhir.
- Syahdan Riyantyo Putro, Alfonsus Oktario Sutomo, Edward Tjahyadi, dan Fachri Mohamad Soetisna yang telah meluangkan waktu dan tempat untuk memberikan bantuan kepada penulis.
- Edward Octovianus Dikarianto dan Joseph Immanuel Kasehung selaku sahabat penulis yang telah memberikan tempat untuk istirahat selama proses penyusunan tugas akhir.
- Seluruh teman dan saudara penulis yang tidak dapat disebutkan satu per satu yang sudah memberikan dukungan dan doa kepada penulis.

Akhir kata, penulis memohon maaf apabila terdapat kekurangan dan kesalahan dalam hasil penyusunan tugas akhir ini.

Bandung, Juni 2024

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Universitas Katolik Parahyangan (UNPAR) merupakan sebuah univertitas yang berada di Bandung berpusat di Jalan Ciumbuleuit. UNPAR sudah berdiri sejak tahun 1955 hingga sekarang dan sudah memiliki lebih dari 10400 mahasiswa aktif dari berbagai fakultas yang ada di UNPAR[1].

UNPAR memiliki lima seleksi jalur masuk yaitu Penelusuran Minat dan Kemampuan (PMDK), Ujian Saringan Masuk (USM) 1, USM 2, USM 3, dan Jalur Kemitraan Sekolah. PMDK merupakan jalur masuk UNPAR dengan sistem penyaringan berdasarkan nilai rapor pada saat mahasiswa masih berstatus pelajar di SMA. Nilai rapor seorang calon mahasiswa harus di atas nilai Kriteria Ketuntasan Minimum (KKM) baik teori maupun praktek. Sebagai contoh jika ada seorang calon mahasiswa yang ingin mengambil jurusan Informatika melalui jalur PMDK, maka calon tersebut harus memiliki nilai Matematika dan Bahasa Inggris di atas KKM. Jalur USM merupakan sistem penyaringan dengan ujian tertulis. Jalur Kemitraan Sekolah merupakan jalur penerimaan melalui seleksi khusus yang disediakan bagi siswa dari sekolah-sekolah yang memiliki kerja sama kemitraan pendidikan dengan UNPAR.

Bagaimana keberhasilan studi adalah kemampuan akademik seorang mahasiswa mengenai sejauh mana mahasiswa mampu menguasai materi selama kuliah. Faktor-faktor yang menjadi penentu keberhasilan studi mahasiswa bisa dilihat dari Indeks Prestasi Kumulatif (IPK) dan lama studi dari seorang mahasiswa. IPK merupakan nilai yang diraih oleh seorang mahasiswa selama berkuliah di sebuah perguruan tinggi. IPK memiliki rentang nilai dari 0.00–4.00. Lama studi adalah total waktu yang ditempuh oleh seorang mahasiswa selama berkuliah dari awal hingga lulus. Lama studi diukur dalam satuan semester dan tergantung jenjang pendidikan apa yang diambil oleh seorang mahasiswa. Jenjang pendidikan tinggi adalah program yang diambil oleh mahasiswa dalam menempuh pendidikan. Jenjang pendidikan tinggi di Indonesia ada empat yaitu:

- Diploma: berfokus kepada sifat praktik untuk menunjang kebutuhan keahlian khusus untuk siap dipekerjakan ketika sudah lulus. Diploma terdiri dari empat tingkatan dan memiliki perbedaan lama studi
  1. Diploma 1 (D1): dua semester
  2. Diploma 2 (D2): empat semester
  3. Diploma 3 (D3): enam semester
  4. Diploma 4 (D4): delapan semester
- Sarjana: berfokus kepada sifat teori untuk menunjang kebutuhan dalam penelitian dan bidang akademis. Mahasiswa dengan lulusan sarjana akan mendapatkan gelar Strata 1 (S1) dan dapat ditempuh selama delapan semester atau empat tahun.
- Magister: jenjang pendidikan tinggi ini dapat ditempuh setelah menyelesaikan S1. Magister biasanya ditempuh selama empat semester atau dua tahun, dan akan mendapatkan gelar Strata 2 (S2).
- Doktor: ini adalah jenjang pendidikan tertinggi yang dapat ditempuh setelah menyelesaikan S2. Jenjang pendidikan Doktor memerlukan waktu sekitar tiga hingga lima tahun dan akan

[1] https://unpar.ac.id/unpar-dalam-angka/

mendapatkan gelar Strata 3 (S3).

Mahasiswa yang berkuliah di UNPAR berasal dari berbagai jalur masuk dan memiliki nilai IPK yang beragam dan lama studi yang beragam. Sebagai contoh terdapat seorang mahasiswa berasal dari jalur masuk PMDK, memiliki nilai IPK di atas 3.00 dengan lama studi 8 semester. Kemudian terdapat mahasiswa lain berasal dari jalur masuk USM 1, memiliki nilai IPK di bawah 3.00 dengan lama studi 11 semester. Mahasiswa dengan IPK tinggi mungkin berasal dari berbagai jalur masuk, baik jalur tertulis seperti USM maupun jalur tidak tertulis seperti PMDK dan Kemitraan. Kemudian mahasiswa dengan lama studi 7 semester mungkin memiliki IPK yang lebih baik daripada mahasiswa dengan lama studi 11 semester.

Dari keberagaman nilai IPK yang ada, mungkin saja jalur masuk dan lama studi berpengaruh terhadap distribusi nilai IPK pada mahasiswa. Mungkin saja jalur masuk PMDK memiliki distribusi nilai IPK yang lebih tinggi dari pada jalur masuk USM 1. Permasalahan yang muncul mencakup variabilitas IPK pada mahasiswa dari berbagai jalur masuk, dampak lama studi terhadap IPK, perbandingan IPK di antara setiap jalur masuk, dan potensi pengaruh faktor eksternal seperti asal SMA seorang mahasiswa.

Permasalahan yang serupa sebelumnya sudah pernah diteliti dan terdapat di jurnal yang berjudul Analisis Hubungan Antara Hubungan Antara Lama Studi, Jalur Masuk dan Indeks Prestasi Kumulatif (IPK) Menggunkan Model Log Linier [11]. Pada jurnal tersebut, penulis melakukan penelitian yang bertujuan untuk menganalisis hubungan antara tiga variabel penting dalam pendidikan tinggi, yaitu lama studi, jalur masuk, dan Indeks Prestasi Kumulatif (IPK) mahasiswa. Analisis menunjukkan adanya hubungan signifikan antara lama studi dengan IPK, serta antara jalur masuk dengan IPK.

Terdapat paling tidak dua teknik yang dapat digunakan untuk melihat hubungan antara IPK atau lama studi dan jalur masuk UNPAR yaitu teknik statistika dan teknik *data mining*. Statistika merupakan sebuah ilmu atau metode ilmiah yang mempelajari mengenai bagaimana merencanakan, mengumpulkan, mengelola, menginterpretasi, dan menganalisa data kemudian hasilnya dipresentasikan [1]. Statistika memiliki properti-properti untuk melihat dan mengukur karakteristik dari sebuah data. Properti-properti ini dapat mengetahui pola, hubungan, dan distribusi dari sebuah data. Beberapa properti statistik yang dapat digunakan adalah

- Rata-rata: menunjukkan titik tengah atau pusat dari sekumpulan data [1].
- Nilai tengah: nilai tengah dari sekumpulan data [1].
- Modus: nilai dari sekumpulan data yang paling sering muncul [1].

Ketiga properti di atas berfungsi untuk mengukur *Central Tendency* sebuah data. *Central Tendency* adalah nilai yang merepresentasikan pusat dari sebuah kumpulan data [1].

Statistika juga memiliki properti untuk mengukur korelasi antar fitur. Korelasi adalah metode analisis untuk mengetahui apakah terdapat fitur yang saling berpengaruh jika terjadi perubahan [1]. *Output* yang dihasilkan dari korelasi berupa rentang nilai antara -1 hingga 1. Jika nilai korelasi mendekati -1 maka kedua fitur memiliki hubungan negatif. Hubungan negatif adalah jika ada perubahan nilai pada satu fitur, maka fitur yang lain akan berubah ke arah yang berlawanan. Jika nilai korelasi mendekati 1 maka kedua fitur memiliki hubungan positif. Hubungan positif adalah jika ada perubahan nilai pada satu fitur, maka fitur yang lain akan berubah ke arah yang sama. Jika nilai korelasi adalah 0 maka kedua fitur tidak memiliki hubungan.

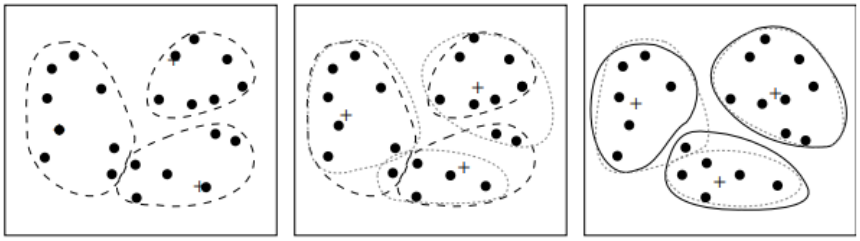
Terdapat dua metode korelasi yang dapat digunakan, yaitu

- Korelasi *Pearson*: korelasi untuk mengetahui hubungan antara dua atribut numerik [1]. Atribut numerik adalah atribut yang bersifat kuantitatif dan direpresentasikan dalam nilai bilangan bulat atau riil [1].
- Korelasi *Chi-Square*: korelasi untuk mengetahui hubungan antara dua atribut kategorikal [1]. Atribut kategorikal adalah atribut yang merepresentasikan sebuah kategori [1].

Selain menggunakan teknik statistika untuk mengetahui hubungan antar fitur, teknik lain yang dapat digunakan untuk mengetahui hubungan antar fitur adalah teknik *data mining*. *Data mining* merupakan proses untuk menemukan pola yang menarik dan pengetahuan dari sebuah data yang

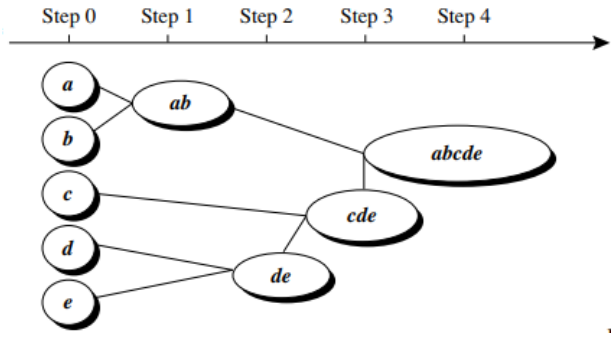
berukuran besar [1]. Teknik *data mining* memiliki algoritma yang digunakan untuk melakukan analisis dan eksperimen yaitu klasterisasi dan klasifikasi. Algoritma klasterisasi merupakan sebuah algoritma dengan cara kerja yaitu membagi populasi atau titik-titik pada data menjadi dua kelompok atau lebih [1]. Beberapa contoh algoritma klasterisasi yang dapat digunakan paling tidak ada dua yaitu *K-Means* dan *Agglomerative*.

- *K-Means*: algoritma dengan pengelompokkan berbasis pada *Centroid* [1]. *Centroid* adalah pusat klaster atau titik rata-rata di dalam sebuah klaster [1]. Lihat Gambar 1.1, cara kerja dari *K-Means* adalah menentukan titik tengah atau *Centroid* lalu menghitung jarak tiap titik data terhadap *Centroid*. Titik data akan dikelompokkan dengan *Centroid* yang terdekat.



Gambar 1.1: Algoritma K-*Means* [1]

- *Agglomerative*: algoritma dengan pengelompokkan yang membentuk hierarki atau bagan seperti pohon. Bagan pohon yang terbentuk dapat disebut sebagai *Dendrogram*. Gambar 1.2 merupakan ilustrasi dari *Agglomerative*, di mana setiap objek akan dimasukkan ke dalam sebuah klaster sendiri dan kemudian setiap klaster tersebut akan digabungkan menjadi klaster yang lebih besar hingga semua objek berada di dalam satu klaster.

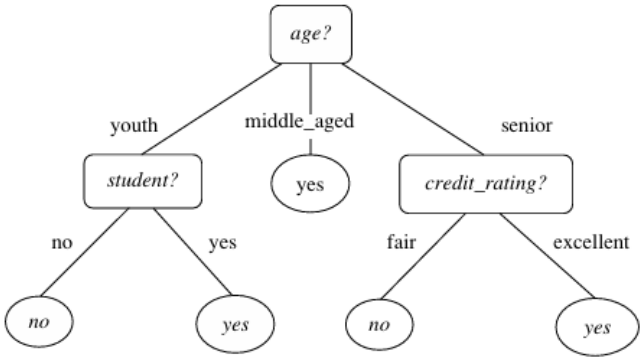


Gambar 1.2: Ilustrasi *Agglomerative* [1]

Apabila sebuah data sudah memiliki label, maka tidak perlu dilakukan klasterisasi tetapi bisa langsung melakukan klasifikasi. Algoritma klasifikasi digunakan dalam melatih sebuah program untuk mengidentifikasi data kemudian mengklasifikasi data tersebut ke dalam sejumlah kelas atau kelompok berdasarkan data yang sudah dilatih sebelumnya [1]. Tujuan dari klasifikasi adalah memprediksi sebuah data baru untuk ditetapkan ke dalam sebuah kelas atau kategori yang sudah ditentukan sebelumnya berdasarkan karakteristiknya.

Algortima klasifikasi yang dapat digunakan paling tidak ada dua yaitu

- *Decision tree*: algoritma klasifikasi yang membentuk struktur seperti pohon untuk menentukan keputusan [1]. Gambar 1.3 merupakan contoh pohon keputusan. Keputusan yang ditentukan oleh *decisoin tree* berasal dari kategori-kategori yang terdapat pada pohon.
- *Naive Bayes* merupakan teknik klasifikasi yang mengasumsikan bahwa nilai suatu atribut prediktor tidak dipengaruhi oleh nilai atribut prediktor yang lain [1]. Atribut prediktor adalah atribut yang digunakan untuk prediksi nilai dari sebuah target. Target adalah atribut yang akan diprediksi.



Gambar 1.3: Contoh Pohon Dari Decision Tree[1]

Pada tugas akhir ini telah dilakukan analisis apakah terdapat hubungan antara IPK atau lama studi, dan jalur masuk UNPAR. Analisis yang telah dilakukan adalah analisis deskriptif menggunakan algoritma klasterisasi dan teknik statistika untuk mengetahui pola, hubungan, dan kemiripan data. Tugas akhir ini menggunakan algoritma Klasifikasi untuk melakukan analisis prediktif terhadap data baru untuk ditetapkan ke dalam sebuah kelas atau kategori yang sudah ditentukan sebelumnya. Eksplorasi *library* pada *Python* bertujuan untuk mempelajari berbagai *library* yang dapat digunakan pada saat melakukan analisis deskriptif dan analisis prediktif.

Data yang digunakan berasal dari Biro Administrasi Akademik (BAA) UNPAR. Data tersebut memiliki informasi mengenai lulusan UNPAR dari tahun 2018–2022. Kolom yang dimiliki adalah IPK, total SKS IPK, lama studi, jalur masuk, program studi, provinsi asal SMA, dan kota asal SMA. Pada *dataset* yang digunakan, terdapat informasi mengenai lulusan UNPAR yang berasal dari Indonesia dan dari luar negeri.

Model prediksi yang telah dibuat diluncurkan menggunakan perangkat lunak yang sudah tersedia di *Python*, yaitu *Graphical User Interfaces* (GUI). GUI tersebut menerima *input* yang merupakan atribut dari data yang dimiliki, dan menghasilkan *output* berupa IPK atau lama studi.

1.2 Rumusan Masalah

Rumusan masalah yang akan dibahas pada tugas akhir ini adalah:

1. Bagaimana menentukan variabel atau fitur yang berpengaruh terhadap IPK, lama studi, dan jalur masuk UNPAR ?
2. Bagaimana melakukan analisis statistika dan membangun model untuk mencari hubungan antara IPK atau lama studi, dan jalur masuk UNPAR ?
3. Bagaimana cara meluncurkan model analisis ?

1.3 Tujuan

Berdasarkan rumusan masalah yang ada, tujuan yang ingin dicapai dari tugas akhir ini adalah:

1. Mempelajari teknik statistika seperti ekstraksi fitur dan mengimplementasikan ke data.
2. Melakukan analisis statistika dan membangun model untuk mencari hubungan antara IPK atau lama studi, dan jalur masuk UNPAR.
3. Membangun aplikasi GUI untuk meluncurkan model.

1.4 Batasan Masalah

Batasan masalah dari tugas akhir ini adalah:

1. Jenjang pendidikan yang digunakan hanya program Sarjana saja.
2. Data lulusan UNPAR tahun 2018–2022 yang berasal dari luar Indonesia tidak digunakan.
3. Cuti mahasiswa akan diabaikan, karena data cuti tidak tersedia.

1.5 Metodologi

Metodologi yang dilakukan pada tugas akhir ini adalah:

1. Studi literatur mengenai statistika, *data science*, algoritma *k-means*, algoritma *agglomerative*, algoritma *naive bayes*, algoritma *decision tree*, *library* pada *python*, dan GUI *python*.
2. Melakukan penyiapan dan eksplorasi data.
3. Melakukan analisis hubungan antara IPK, lama studi, dan jalur masuk UNPAR.
4. Melakukan pembuatan model prediksi.
5. Melakukan pengujian model yang sudah dibuat.
6. Membuat perangkat lunak untuk meluncurkan model.

1.6 Sistematika Pembahasan

Sistematika penulisan pada tugas akhir ini adalah:

1. Bab 1: Pendahuluan
   Membahas latar belakang masalah yang ada, rumusan masalah, tujuan, batasan masalah, dan metodologi tugas akhir.
2. Bab 2: Landasan Teori
   Membahas landasan teori mengenai statistika, properti statistika, *data mining*, *Pearson*, *Chi-Square*, *K-Means*, *Agglomerative*, *Naive Bayes*, *Decision Tree*, *library* pada *Python*.
3. Bab 3: Analisis Penyelesaian Masalah
   Membahas analisis masalah serta solusi dari permasalahan tersebut, membahas contoh kasus untuk mempelajari korelasi *Pearson*, korelasi *Chi-Square*, algoritm *K-Means*, algoritma *Agglomerative*, algoritma *Naive Bayes*, algoritma *Decision Tree*, evaluasi model MSE dan RMSE, mebahas eksplorasi *library Python*.
4. Bab 4: Penambangan Data
   Membahas eksplorasi data menggunakan data *real*, analisis korelasi antara IPK atau lama studi dengan jalur masuk UNPAR, analisis fitur yang berpengaruh terhadap IPK, lama studi, dan jalur masuk UNPAR, pemilihan fitur untuk pembuatan model prediksi IPK dan model prediksi lama studi, pembuatan model prediksi IPK dan model prediksi lama studi.
5. Bab 5: Pembuatan GUI Peluncuran Model
   Membahas pembuatan GUI untuk meluncurkan model prediksi IPK dan prediksi lama studi.
6. Bab 6: Kesimpulan dan Saran
   Membahas kesimpulan yang didapatkan dari hasil penelitian tugas akhir yang telah dilakukan dan saran yang dapat membuat peneletian tugas akhir lebih baik.