

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Universitas Pelita Harapan Medan (UPH Medan) adalah sebuah perguruan tinggi swasta Kristen yang ada di Medan yang diresmikan pada hari Sabtu tanggal 26 April 2014 (Peresmian UPH Medan - UPH | Universitas Pelita Harapan - UPH | Universitas Pelita Harapan n.d.). Dalam dunia pendidikan, perkuliahan dikatakan penting karena menjadi jalan untuk menuju pendidikan yang lebih tinggi dan berkesempatan untuk berpendidikan di luar negeri serta dapat menambah wawasan dalam dunia luar seperti magang. Saat ini, kebanyakan perusahaan menjadikan gelar sarjana atau setidaknya diploma sebagai persyaratan standar untuk melamar pekerjaan yang dimana ini tentunya membuat peluang bagi lulusan SMA untuk mendapatkan pekerjaan menjadi lebih kecil.

Dalam pemilihan program studi yang berfokus pada bidang teknologi akan dihadapkan pada pemilihan yang umum yaitu informatika dan sistem informasi. Keduanya merupakan dua bidang yang fokus pada pengembangan dan pemanfaatan teknologi, tetapi keduanya memiliki perbedaan latar belakang dan fokus yang lebih mendalam sehingga sering disamakan. Umumnya informatika lebih mengfokuskan pada pengembangan perangkat lunak, sistem komputer, dan teknologi informasi yang menyeluruh dimana pendidikannya biasanya sering dijumpai seperti

algoritma, struktur data, basis data, jaringan, dan komputasi teori, Informatika juga tentunya lebih membutuhkan ilmu matematika yang lebih mendalam daripada sistem informasi. Sedangkan sistem informasi lebih mengfokuskan pada kegunaan teknologi informasi yang mengelola informasi dan memecahkan masalah bisnis di dalam organisasi, pendidikannya biasanya mencakup manajemen bisnis, analisis sistem, keuangan dan strategi.

Untuk setiap calon siswa-siswi yang sudah lulus SMA/SMK akan dihadapkan untuk memutuskan melanjutkan pendidikannya atau tidak, dan hal tersebut sering terjadi masalah. Biasanya sebagian dari mereka sudah merencanakan dan menentukan program studi apa yang mereka inginkan untuk difokuskan sedangkan sebagiannya lagi mengalami kebingungan dalam memilih karena mereka tidak tahu *passion* mereka dimana sehingga mereka akan mengikuti pilihan teman atau pilihan orangtua mereka dengan buru-buru tanpa berfikir kedepannya. Pemilihan program studi yang salah dapat berdampak kepada calon mahasiswa baru karena adanya ketidakcocokkan antara pilihannya dengan dirinya sendiri. Untuk itu hendaknya calon mahasiswa baru memilih program studi yang memang sesuai dengan keinginan dan kemampuannya baik secara akademik atau non akademik. Pemilihan ini apabila didukung oleh adanya teknologi yang dapat mendukung dalam pemilihan.

Machine learning merupakan sebuah teknologi baru yang menjadi bagian dari AI (*Artificial Intelligence*). *Machine learning* bekerja layaknya

manusia karena mampu mengolah data yang besar, semakin besar datanya maka semakin besar tingkat ke akurasi. *Machine learning* juga sangat berperan dalam meringankan pekerjaan penggunaannya dalam kehidupan sehari-harinya. Algoritma *machine learning* yang akan dibandingkan yaitu *Naïve Bayes* dan *K-Nearest Neighbor* (KNN). Ini disebabkan karena *Naïve Bayes* merupakan algoritma klasifikasi yang menggunakan dasar *teorema Bayes* dalam menentukan kelas mana yang paling optimal berdasarkan perhitungan peluangnya. Sedangkan *K-Nearest Neighbor* (KNN) merupakan algoritma yang menggunakan data berdasarkan *train dataset* dan menggunakan K sebagai bantuan klasifikasian dengan K sebagai tetangga terdekat

Berdasarkan penelitian terdahulu yang melakukan perbandingan algoritma berdasarkan studi kasus yang berbeda-beda, maka muncullah penelitian yang dilakukan oleh (Krisna dan Salamah 2022) tentang “Perbandingan Algoritma *Naïve Bayes* Dan *K-Nearest Neighbor* Untuk Klasifikasi Berita Hoax Kesehatan Di Media Sosial Twitter” menyatakan bahwa algoritma *K-Nearest Neighbor* (KNN) memberikan hasil akurasi yang lebih baik jika dibandingkan dengan algoritma *Naïve Bayes*. Ada juga penelitian lainnya yang dilakukan oleh (Wisdayani, Nur, dan Wasono 2019) tentang “Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas Di Kabupaten Pati Jawa Tengah” menyatakan bahwa nilai rata-rata akurasi pada metode *Naïve Bayes* sebesar 86,32% sedangkan pada metode *K-Nearest Neighbor* (KNN) sebesar 88,82%. Terakhir penelitian yang

dilakukan oleh (Putry dan Sari 2022) tentang “Komparasi Algoritma KNN Dan *Naïve Bayes* Untuk Klasifikasi Diagnosis Penyakit Diabetes Melitus” menyatakan bahwa nilai akurasi dari metode *Naïve Bayes* mendapatkan nilai 80% dan untuk metode *K-Nearest Neighbor* (KNN) mendapatkan nilai 75%. Dapat dilihat bahwa penelitian terdahulu yang telah dilakukan, tidak dapat dipastikan bahwa baik metode *Naïve Bayes* ataupun *K-Nearest Neighbor* (KNN) lebih tinggi tingkat akurasinya tergantung pada studi kasusnya.

Oleh karena itu, perancangan sistem perbandingan algoritma ini dibuat untuk membandingkan kembali dua metode tersebut untuk melihat tingkat akurasinya dalam bidang pemilihan program studi dan muncullah penelitian yang berjudul “Perbandingan Algoritma Dalam Pemilihan Program Studi Menggunakan Metode *Naïve Bayes* dan *K-Nearest Neighbor* (KNN) Pada Universitas Pelita Harapan Medan” dengan harapan dapat melihat apakah minat yang dipilih calon mahasiswa sesuai dengan nilai akademiknya atau tidak.

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, maka diperoleh rumusan masalahnya sebagai berikut ini:

1. Bagaimana mengimplementasikan perbandingan algoritma dalam pemilihan program studi menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor* (KNN)?

2. Bagaimana melihat hasil keakuratan perbandingan algoritma dalam pemilihan program studi menurut nilai akademik dalam menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor* (KNN)?

1.3. Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka didapatkan tujuan penelitian sebagai berikut ini:

1. Melihat cara mengimplementasikan perbandingan algoritma dalam pemilihan program studi menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor* (KNN).
2. Melihat tingkat keakuratan perbandingan algoritma dalam pemilihan program studi kuliah menurut nilai akademik dalam menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor* (KNN).

1.4. Batasan Masalah

Untuk menghindari semakin melebarnya pokok permasalahan agar tetap fokus pada pokok permasalahan. Maka ditemukan batasan masalah tersebut untuk membatasi ruang lingkup pembahasan, yaitu:

1. Penggunaan dataset diambil dari data Mahasiswa UPH Medan intake 2019 dan 2020 yang mengambil program studi Informatika atau Sistem Informasi
2. Tingkat keakuratan dalam perbandingan algoritma ini hanya berdasarkan nilai akademik Matematika, Bahasa Inggris, Bahasa

Indonesia dan Komputer Ujian Akhir Semester (UAS) pada masa pendidikan SMA 3.

3. Perbandingan algoritma ini menggunakan bantuan Google colab dengan memanfaatkan bahasa Pemrograman Python
4. Perbandingan algoritma ini menggunakan dua metode yaitu *Naïve Bayes* dan *K-Nearest Neighbor* (KNN).

1.5. Manfaat Penelitian

1. Perbandingan algoritma yang dibangun ini membantu untuk melihat apakah minat yang dipilih oleh calon mahasiswa sesuai dengan nilai akademiknya atau tidak.
2. Perbandingan algoritma yang dibangun ini membantu untuk mengetahui keakuratan menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor* (KNN).
3. Perbandingan algoritma ini dibuat untuk membantu meningkatkan mutu dan kualitas tiap mahasiswa yang memang pemilihannya sesuai dengan kemampuan akademik dan minat mereka yang ada di Universitas Pelita Harapan Medan (UPH Medan).

1.6. Sistematika Penulisan

Penelitian ini terdiri dari enam bab dengan sistematika penulisan sebagai berikut ini:

1. BAB I: PENDAHULUAN

Bab I dalam penelitian ini membahas tentang latar belakang masalah, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian dan sistematika penulisan.

2. BAB II: LANDASAN TEORI

Bab II dalam penelitian ini membahas dasar-dasar teori yang berkaitan dengan topik penelitian berdasarkan studi literatur untuk memahami dasar-dasar mengenai program studi, *machine learning*, *Naïve Bayes*, *K-Nearest Neighbor*, bahasa pemrograman *Python* dan kajian penelitian yang memang berhubungan.

3. BAB III METODOLOGI

Bab III dalam penelitian ini membahas tentang tahapan-tahapan yang dilakukan dalam penelitian ini, pada tiap tahapan akan dijelaskan dan dideskripsikan dengan rinci beserta dengan data-data yang digunakan mulai dari mengidentifikasi masalah, pengumpulan data, pelatihan hingga melakukan pengujian berdasarkan algoritma yang akan digunakan.

4. BAB IV HASIL DAN PEMBAHASAN

Bab IV dalam penelitian ini membahas tentang perbandingan hasil penelitian dan pengujian berupa performa masing-masing algoritma

5. BAB V KESIMPULAN & SARAN

Bab V dalam penelitian ini membahas tentang kesimpulan dari semua uraian pada tiap-tiap bab sebelumnya, membahas tentang algoritma

mana yang lebih baik berdasarkan hasil keakuratannya dan berisi saran yang diharapkan dapat berguna untuk pengembangan penelitian ini kedepannya.

