BAB I
PENDAHULUAN

1.1 Latar Belakang Penelitian

Institut Teknologi dan Bisnis STIKOM Bali dahulu bernama STMIK STIKOM Bali adalah bentuk Perguruan Tinggi IT swasta sejak tahun 2002. Menjadi perguruan tinggi IT swasta pertama di Bali, ITB STIKOM Bali yang kini berada di tengah era globalisasi di bidang pendidikan, menghadapi persaingan dan perbandingan dengan perguruan tinggi serta universitas lain, terutama di wilayah Bali.. ITB STIKOM Bali saat ini mempunyai 6 Program Studi yaitu ; S1-Sistem Informasi, S1-Sistem Komputer, S1-Teknologi Informasi, S1-Bisnis Digital, D3-Manajemen Informatika dan S1-*Dual Degree International*.

Setiap tahun, ITB STIKOM Bali melaksanakan penerimaan mahasiswa baru dengan target peningkatan sebesar 10% dari tahun sebelumnya. Strategi pemasaran yang diterapkan oleh ITB STIKOM Bali telah berperan besar dalam pencapaian target tersebut. Proses seleksi mahasiswa baru merupakan upaya untuk menyaring atau memilih calon mahasiswa terbaik dari segi akademis untuk belajar di institusi pendidikan, yang melibatkan berbagai komponen atau kriteria penilaian (multi kriteria). Untuk mencapai strategi pemasaran yang efektif, ITB STIKOM Bali menganalisis data historis guna menemukan pola dan menggali informasi tersembunyi dari data tersebut. Proses analisis ini dikenal dengan istilah data mining. Pemanfaatan teknologi semakin disesuaikan dengan metode *machine learning.* Dengan cara pengelompokkan untuk meningkatkan strategi promosi.

Berdasarkan data historis oleh pihak Penerimaan Mahasiswa Baru ITB STIKOM Bali, melakukan klasterisasi dapat mempengaruhi pengambilan keputusan dalam menentukan wilayah promosi yang tepat sasaran sehingga tidak terjadi penurunan jumlah mahasiswa pada tahun mendatang. Penelitian ini akan melakukan pengelompokkan berdasarkan pada data mahasiswa baru ITB STIKOM Bali.

Terdapat metode pengelompokkan yang digunakan dalam klasterisasi melibatkan algoritma seperti K-Medoids, X-Means Clustering, Mean Shift Clustering dan berbagai metode klaster lainnya. Ada beberapa algoritma yang diusulkan namun dalam penelitian ini, peneliti akan menerapkan algoritma K-Means Clustering untuk menentukan jumlah cluster yang optimal. Algoritma K-Means Clustering sering digunakan karena efektivitas dan efisiensinya. Algoritma ini mudah dipahami serta memiliki waktu komputerisasi yang lebih cepat. Selain itu, K-Means Clustering memiliki capaian akurasi yang tinggi dalam mengukur ukuran objek, sehingga cocok digunakan dalam pemrosesan data berukuran besar karena terukur dan efisien. (K. Arai, 2007).

Sebuah penelitian dilakukan untuk membandingkan algoritma K-Means Clustering, K-Medoids, dan X-Means dengan mengukur akurasinya menggunakan Davies Bouldin Index (DBI). Hasilnya menunjukkan bahwa K-Means mencatat nilai DBI sebesar 0,377, K-Medoids 0,930, dan X-Means 0,497. Dari data ini, K-Means dianggap lebih superior karena memiliki nilai DBI terendah. Nilai DBI yang lebih kecil menandakan bahwa algoritma tersebut lebih efektif dalam mengelompokkan data, yang berarti kualitas pengelompokan juga lebih optimal. (Gideon Bartolomeus, 2022).

Dalam mengoptimasi metode K-Means Clustering, diperlukan perbandingan metode perhitungan jarak, secara umum dalam penelitian dengan metode *K-Means* hanya menggunakan satu perhitungan jarak yaitu *Euclidean*, namun dalam peneltian ini menggunakan 3 perbandingan metode perhitungan jarak.

Penelitian ini menggunakan perbandingan 3 metode perhitungan jarak, yaitu *Euclidean*, *Manhattan* dan *Chebyshev Distance*. *Euclidean, Manhattan* maupun *Chebyshev* mempunyai kelebihan masing-masing. *Chebyshev Distance* adalah metrik jarak yang digunakan dalam berbagai algoritma pembelajaran mesin dan analisis data. Ini mengukur jarak antara dua titik dalam ruang n-dimensi berdasarkan dimensi yang memiliki perbedaan terbesar.

Secara umum, ruang jarak yang paling umum digunakan adalah Manhattan dan Euclidean. Euclidean sering menjadi pilihan karena mengukur jarak paling pendek antara dua titik, sedangkan Manhattan lebih disukai dalam situasi tertentu karena kemampuannya yang lebih efektif dalam menangani kondisi khusus, seperti outliers. (Yudi Agusta, 2007)

Penelitian oleh (Rahma Isparani et al, 2019) menghasilkam perhitungan jarak *Euclidean* lebih kecil dibandingkan *Manhattan* oleh sebab itu *distance measure* paling optimal adalah *Euclidean*. Algoritma *Manhattan* dalam proses iterasi lebih cepat dan efisisen daripada *Euclidean* yang bisa sampai proses iterasi 15 kali berulang, sedangkan *Manhattan* hanya 7 kali proses Iterasi (Faisal Nur Cahya et al, 2023). Dalam penelitian (Siti Hotijah et al, 2023) nilai DBI yg

diperoleh untuk metode *Chebysev Distance* lebih baik dibandingkan *Euclidean Distance* dengan nilai DBI lebih kecil dibandingkan *Euclidean.*

Sehingga hasil dari klasterisasi masih kurang baik dan presentase yang dihasilkan belum maksimal jika hanya menggunakan satu metode yaitu *Euclidean Distance*, maka dari itu penulis mencoba membandingkan dengan metode perhitungan jarak sebagai pembanding yang sesuai yaitu algoritma *Manhattan Distance* dan *Chebysev Distance* dalam optimasi metode *K-Means Clustering*.

Membandingkan ketiga *formula distance* ini menggunakan evaluasi Davies Bouldin-index dan Silhouette Score, penulis dapat menentukan algoritma yang paling sesuai dengan karakteristik data dan tujuan analisis data profil mahasiswa ITB Stikom Bali, nantinya akan meningkatkan akurasi dan relevansi hasil yang diperoleh dari analisis atau model *machine learning* yang diterapkan. Dengan adanya penelitian terkait komparasi ini diharapkan penulis dapat menentukan perhitungan jarak yang paling baik, serta nilai akurasi jarak paling dekat.

Berdasarkan penjabaran permasalahan diatas, maka akan dilakukan Penelitian "Studi Perbandingan Algoritma *Euclidean, Manhattan dan Chebyshev Distance* Dalam Optimasi Metode *K-Means Clustering* Untuk Pengelompokkan Data Penerimaan Mahasiswa Baru ITB STIKOM Bali*"* pengelompokkan berdasarkan data mahasiswa pada tahun ajaran 2021/2022 & 2022/2023 Mendapatkan nilai akurasi dalam metode perhitungan jarak klaster terbaik menggunakan Davies Bouldin Index dan Silhouette Score. Sehingga akan menjadi bahan pertimbangan dalam menentukan promosi di tahun berikutnya.

1.2 Identifikasi Masalah

Berdasarkan penjelasan mengenai latar belakang diatas untuk permasalahan yang akan diteliti yaitu, melakukan klasterisasi data prodi pilihan calon mahasiswa baru menggunakan *K-Means Clustering* serta melakukan optimasi metode K-means Clustering dengan cara komparasi 3 metode perhitungan jarak yaitu *Euiclidean Distance*, *Manhattan Distance dan Chebyshev Distance* dalam pengelompokkan data pilihan program studi calon mahasiswa baru ITB STIKOM Bali tahun ajaran 2021/2022 & 2022/2023. Didapatkan hasil yang akan dievaluasi menggunakan metode *Davies Bouldin Index* (DBI) dan Sillhouette Score yang menghasilkan untuk nilai maksimal jarak antar *cluster*. Meskipun *K-Means Clustering* terbaik tapi belum ada algoritma metode jarak yg tepat dan optimal dalam studi kasus data promosi ini.

1.3 Batasan Masalah

Untuk mempermudah dalam penelitian, batasan masalah dirancang dan yang akan dibahas sebagai berikut :

1. Perhitungan dilakukan dengan metode data mining yaitu Algoritma *K-Means Clustering*.
2. Perhitungan jarak menggunakan perbandingan metode *Euclidean Distance*, *Manhattan Distance* dan *Chebyshev Distance*
3. Data yang gunakan untuk studi penelitian ini yaitu data penerimaan mahasiswa baru ITB STIKOM Bali pada tahun ajaran 2021/2022 dan 2022/2023 sebanyak 2.689 data

4. Untuk perhitungan Algoritma secara manual dilakukan dengan menggunakan Microsoft office Excel.
5. Perhitungan secara sistem menggunakan tools RapidMiner
6. Perbandingan evaluasi akurasi *formula distance* menggunakan metode Davies-Bouldin Index (DBI) dan nilai Sillhouette Score

1.4 Rumusan Masalah

Berdasarkan uraian pada bagian latar belakang, penulis merumuskan masalah sebagai berikut:

1. Manakah metode jarak terbaik dalam perhitungan jarak K-Means Clustering?
2. Bagaimana mengetahui persebaran pendaftar mahasiswa baru ITB STIKOM Bali pada tahun 2022 dan 2023?

1.5 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Menghasilkan satu metode perhitungan jarak terbaik dari perhitungan jarak cluster optimal dari perhitungan *Euclidean*, *Manhattan* dan *Chebysev* berdasarkan data mahasiswa baru ITB STIKOM Bali.
2. Mendeskripsikan persebaran pendaftar berdasarkan data mahasiswa baru tahun 2022 dan 2023 berdasarkan profil mahasiswa pendaftar untuk menentukan srategi promosi dalam setiap klaster yang tebentuk.

1.6 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi yang bermanfaat, baik dari segi akademis maupun praktis, dengan rincian sebagai berikut.

1.6.1 Manfaat Akademik

Manfaat akademik dari penelitian ini agar dapat berkontribusi pada pengembangan ilmu pengetahuan, khususnya dalam bidang ilmu komputer terkait data mining, yang berhubungan dengan klasterisasi K-Means Clustering serta metode perhitungan jarak, seperti Euclidean Distance, Manhattan Distance, dan Chebyshev Distance.

1.6.2 Manfaat praktis

Manfaat praktis dari penelitian ini perushaan dibantu dalam membuat keputusan yang tepat untuk meningkatkan pelayanan operasional secara optimal serta mengurangi kerugian, dengan mempertimbangkan prediksi jumlah calon mahasiswa yang akan mendaftar di tahun-tahun mendatang.