

BAB I PENDAHULUAN

1.1 Latar Belakang

Saat ini terdapat kebutuhan mendesak akan sistem klasifikasi untuk memudahkan tugas manusia. Mekanisme sistem klasifikasi ini secara umum dinilai sangat berguna karena dapat menggantikan peran para ahli. Proses klasifikasi dinilai secara subyektif oleh para ahli sangat memakan waktu. Klasifikasi melibatkan prediksi objek dari kelas yang belum memiliki kelas. Banyak ilmuwan mulai memperkenalkan algoritma klasifikasi seperti Jaringan saraf tiruan, mesin vektor pendukung, dan *K-Nearest Neighbor*.

Ada banyak algoritma klasifikasi, namun *K-Nearest Neighbor* adalah salah satu teknik klasifikasi yang paling populer. Cara ini dinilai sangat sederhana dan efektif. Metode ini menggunakan data pembelajaran untuk menentukan kelas suatu objek yang ingin diketahui kelasnya. Berdasarkan jarak terdekat antara data latih dengan objek. Jarak ini digunakan sebagai nilai kemiripan atau kedekatan antar data uji. Faktanya, penggunaan metode *K-Nearest Neighbor* seringkali dibarengi dengan penggunaan metrik jarak Euclidean, padahal akurasi klasifikasi untuk semua data pelatihan yang menggunakan metrik *Euclidean distance* tidak selalu unggul. Masalah muncul jika salah memilih metrik jarak untuk kasus yang ingin di klasifikasikan. Selain itu, karakteristik data latih yang digunakan juga tidak sesuai. Akibatnya keakuratan klasifikasi yang peneliti lakukan menurun (Basid & Sari, 2017). Oleh karena itu, perlu mempelajari cara memilih metrik jarak terbaik untuk studi kasus dan data pelatihan yang akan di gunakan. Tidak terbatas pada metrik jarak *Euclidean*, banyak metrik jarak alternatif telah ditemukan. Oleh karena itu, penulis berharap dapat memilih metrik jarak yang memberikan hasil akurasi terbaik untuk data pelatihan yang dipilih.

Dalam penelitian sebelumnya, Kurniawan (2018), ``Implementasi Algoritma *K-Nearest Neighbor* dengan Metode Klasifikasi dan Pengukuran jarak *Manhattan Distance* untuk Prediksi Kelulusan UN Berdasarkan Hasil Nilai *Tryout* Berbasis Java Desktop Pada SMA Harapan Jaya 2''. Pada penelitian ini digunakan Metode *K-Nearest Neighbor* (KNN) dengan *Manhattan distance* untuk mengukur

tingkat akurasi prediksi keberhasilan ujian nasional berdasarkan hasil tes, dan menyimpulkan bahwa terdapat perbedaan tingkat akurasi. bahwa Perbedaan tingkat akurasi yang dihasilkan antar parameter nilai K terlalu signifikan dan masih dalam kategori akurasi yang tinggi.

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka pada penelitian yang dilakukan kali ini akan berjudul ``Perbandingan Metode *Euclidean Distance* dan *Manhattan Distance* Pada Pengukuran Jarak Dalam Metode KNN``. Dengan tujuan untuk mengukur tingkat akurasi terbaik pada *Euclidean distance* dan *Manhattan distance* dalam studi kasus klasifikasi pengukuran angka kelulusan pada perguruan tinggi.

1.2 Perumusan masalah

Perumusan masalah dalam penelitian ini yaitu Metode manakah yang memiliki akurasi terbaik dalam pengukuran *eucliden distance* dan *manhattan distance*?

1.3 Batasan Penelitian

Agar penelitian ini bersifat lebih spesifik dan terfokus maka permasalahannya dibatasi sebagai berikut:

1. Metode yang digunakan adalah *K-Nearest Neighbor*(KNN) menggunakan *eucliden distance* dan *Manhattan distance*
2. Data yang didapat berasal dari UC Irvine (*University of California, Irvine*) dengan jumlah 3630 data kelulusan tahun 12 Desember 2021.
3. Variabel yang digunakan terdiri dari 33 variabel, yaitu 1 variabel terikat (*Target*) dan 32 variabel bebas (*Marital status, Application mode, Application order, Course, Daytime/evening attendance, Previous qualification, Previous qualification (grade), Nationality, Mother qualification, Father qualification, Mother occupation, Father occupation, Admission grade, Displaced, Educational special needs, Debtor, Tuition fees up to date, Gender, Scholarship holder, Age at enrollment, International, Curricular units 1st sem (Enrolled), Curricular units 1st sem (evaluations), Curricular units 1st sem (approved), Curricular units 1st sem (grade), Curricular units 2nd sem (enrolled), Curricular units 2nd sem (evaluations), Curricular units 2nd sem*

(*approved*), *Curricular units 2nd sem (grade)*, *Unemployment rate*, *Inflation rate*, *GDP*).

4. Hasil dari pengujian berupa informasi akurasi algoritma *K_Nearest Neighbor*(KNN) menggunakan *Eucliden distance* dan *Manhattan distance*.
5. Bahasa pemrograman yang digunakan *python*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui akurasi terbaik dalam perhitungan jarak *Euclidean* dan jarak *Manhattan*.

1.5 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini antara lain:

1. Dapat mengetahui jarak mana yang tingkat akurasi nya lebih baik dalam studi klasifikasi pengukuran lulus dan tidak lulus di perguruan tinggi.
2. Dapat menjadi bahan referensi bagi peneliti lain dengan topik pembahasan yang sama.
3. Sebagai alternatif metode prediksi pengukuran angka kelulusan pada perguruan tinggi.