Nama : Fatwa Sifaun Nahar

NPM : 20081010073

Kelas : Riset Informatika

**Review Paper**

**Judul Paper:**

Perbandingan Penggunaan Jarak Manhattan, Jarak Euclid, dan Jarak Minkowski dalam Klasifikasi Menggunakan Metode KNN pada Data Iris

**Abstraksi paper**

*Dalam klasifikasi, ukuran kebaikan suatu metode ditentukan oleh akurasi dan skor F1. Penelitian ini bertujuan untuk menentukan rata-rata akurasi dan skor F1 pada klasifikasi menggunakan metode KNN pada data Iris dengan bantuan program Python. Rata-rata akurasi maupun skor F1 ditentukan dengan membuat variasi parameter yang digunakan yaitu random state dari 1 sampai dengan 10000, k yang digunakan dalam metode KNN dari 1 sampai dengan 31 (tergantung prosentase data uji yang digunakan), q (definisi jarak yang digunakan yaitu jarak Manhattan, jarak Euclid, dan jarak Minkowski) serta persentase data uji yang digunakan. Dalam penelitian ini, digunakan persentase data uji 10%, 20 %, 30 %, dan 40 %. Jika digunakan k = 5 dalam metode KNN, variasi random state 1 sampai dengan 10000 dan digunakan data uji 20 % maka diperoleh akurasi rata-rata 96,6667 %. Demikian juga dapat diperoleh histogram dari akurasi tersebut. Histogram tersebut cenderung menceng ke kiri. Dalam penelitian ini, juga diperoleh hasil bahwa dengan melakukan variasi k dalam metode KNN juga dapat diperoleh akurasi maupun skor F1 yaitu bahwa makin besar data uji yang digunakan maka makin kecil nilai k untuk mencapai akurasi ataupun skor F1 maksimum. Lebih lanjut, dalam perubahan nilai p pada jarak Minkowski, apabila p membesar maka nilai maksimum rata-rata akurasi akan cenderung membesar namun nilai k cenderung mengecil.*

**Persoalan Praktis**

Berdasarkan paper, beberapa persoalan praktis yang mungkin perlu diselesaikan melalui penelitian adalah:

1. Pemilihan Jenis Jarak yang Optimal: Bagaimana memilih jenis jarak yang paling sesuai untuk klasifikasi data iris dengan metode KNN? Pemilihan jenis jarak dapat mempengaruhi hasil klasifikasi, dan identifikasi jenis jarak yang optimal adalah permasalahan praktis yang relevan.
2. Optimasi Parameter KNN: Bagaimana memilih nilai k (banyaknya tetangga terdekat) yang optimal dalam metode KNN untuk data iris? Ini mencakup pertimbangan antara akurasi dan kompleksitas komputasi.
3. Pengaruh Jumlah Data Pelatihan Terhadap Hasil : Penelitian dapat lebih mendalam mengenai bagaimana kinerja KNN dipengaruhi oleh jumlah data pelatihan. Sebagai yang disebutkan dalam paper, semakin banyak data pelatihan, semakin lama waktu perhitungan.
4. Pengaruh Persentase Data Uji: Bagaimana pengaruh variasi persentase data uji terhadap kinerja metode KNN? Apakah ada persentase data uji yang lebih sesuai untuk dataset Iris?
5. Waktu Perhitungan: Bagaimana mengatasi masalah waktu perhitungan dalam penggunaan KNN, terutama saat jumlah data pelatihan menjadi besar?

**Research Question**

berikut adalah pernyataan-pernyataan yang menggambarkan research question berdasarkan persoalan praktis yang telah diidentifikasi:

1. Bagaimana dampak pemilihan jenis jarak (Manhattan, Euclidean, Minkowski) pada kinerja metode KNN dalam klasifikasi data iris?
2. Bagaimana nilai k yang optimal dapat dipilih dalam metode KNN untuk klasifikasi data iris, mempertimbangkan akurasi dan waktu komputasi?
3. Bagaimana variasi persentase data uji memengaruhi kinerja metode KNN dalam klasifikasi data iris?
4. Bagaimana penelitian ini dapat membantu mengatasi tantangan waktu perhitungan yang muncul ketika jumlah data pelatihan menjadi besar dalam penggunaan metode KNN pada data iris?
5. Bagaimana metode KNN dapat digunakan untuk mengidentifikasi fitur-fitur yang paling relevan dalam data iris untuk meningkatkan kinerjanya?
6. Bagaimana hasil penelitian ini dapat diterapkan pada masalah klasifikasi lainnya atau dataset yang berbeda di luar data iris?

**Teori Sesuai Paper**

- Teori Jarak dalam Pengolahan Data

- Teori Metode K-Nearest Neighbors (KNN)

- Teori Pengukuran Kinerja Klasifikasi

- Teori Optimasi Parameter dalam Pembelajaran Mesin