## BAB I

## PENDAHULUAN

### Latar Belakang

Berita adalah suatu peristiwa yang dituliskan. Pada zaman sekarang banyak berita yang secara mudah tersebar luas di dunia maya dan secara mudah juga dapat kita akses dan kita baca, tetapi ada beberapa berita yang memuat berita yang tidak sesuai dengan fakta yang terjadi. Oleh karena itu dibuatlah sistem yang mampu membantu membedakan berita yang sesuai dengan fakta dan berita yang hoax (tidak sebenernya / bohong).

## BAB II

## LANDASAN TEORI

### KNN

KNN (*K-Nearest Neighbor)* termasuk dari tahap learning dalam *Artificial Intelligence* dan suatu algoritma pembelajaran yang diterapkan pada mesin, sehingga mesin dapat mempelajari dari data yang ada. Algoritma KNN biasa digunakan untuk permasalahan pemberian *label* pada suatu data yang sangat banyak. Pada tugas sebelumnya *Fuzzy* dapat digunakan juga untuk melakukan pemberian *label* pada data tetapi jika data yang akan diberi label banyak dan parameter yang diberikan pada data banyak juga, maka akan sulit dan membutuhkan waktu yang cukup banyak. Oleh sebab itu KNN dapat mengatasi masalah tersebut. KNN membutuhkan data train yaitu terdapat sejumlah data yang lebih banyak dari data yang akan diteskan dan setiap data memiliki *label-*nya sendiri. Dan yang terpenting adalah nilai k, dan nilai k harus ganjil karena pada saat melakukan pengurutan pada array jarak kemiripan maka diambil sejumlah k yang terkecil lalu akan dicek *label* terbanyak yaitu yang akan di-*label-*kan pada data yang dites.

## BAB III

## PERANCANGAN SISTEM

### Perancangan Fungsi dan Prosedur

|  |  |  |
| --- | --- | --- |
| No. | Nama fungsi atau prosedur | Keterangan |
| 1. | importdata(file) | Fungsi yang digunakan untuk mengambil data dari file excel dengan format CSV |
| 2. | eksportdata(text) | Prosedur yang digunakan untuk me-*write* data dari hasil knn ke file excel dengan format csv |
| 3. | euclidDist(a, b) | Fungsi yang digunakan untuk melakukan perhitungan jarak dengan rumus |
| 4. | manhattan(a,b) | Fungsi yang digunakan untuk melakukan perhitungan jarak dengan rumus ( |
| 5. | chebyshev(a,b) | Fungsi yang digunakan untuk melakukan perhitungan jarak dengan rumus ( |
| 6. | winner(a) | Fungsi yang digunakan untuk membantu mengoutputkan *label* untuk data yang dibandingkan dari hasil perhitungan jarak (Euclidean,manhattan, dll) |
| 7. | KNNz(train, test, k) | Fungsi yang digunakan untuk melakukan klasifikasi pada data testing berdasarkan data train |

Gambar 1 chart akurasi untuk setiap k

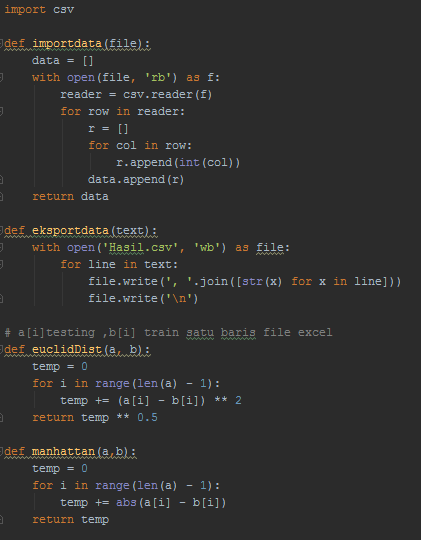
Pada gambar diatas dapat disimpulkan bahwa pada titik 31 dan 32 memiliki nilai akurasi tertinggi sehingga pada kasus ini lebih baik menggunakan k = 31 atau pun 32.

## BAB IV

## PENGUJIAN DAN ANALISA DATA

### Pengujian Data

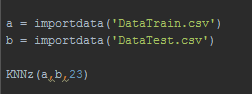
Pengujian data dilakukan menggunakan K-fold validation untuk mengetahui akurasi dari fungsi knn yang dibuat. Penggunaan k-fold ini yaitu membagi data train sebanyak 4 dengan 1 filenya berisi 1000 data train setelah dibagi menjadi 4 yaitu a, b, c dan d setelah itu ambil satu file untuk dijadikan testing dan dites ke file b, c dan d setelah itu lihat hasilnya dan sesuaikan dengan yang asli. Maka akan diketahui akurasi dari data yang benar.



Gambar 2 code



Gambar 3 code



Gambar 4 code

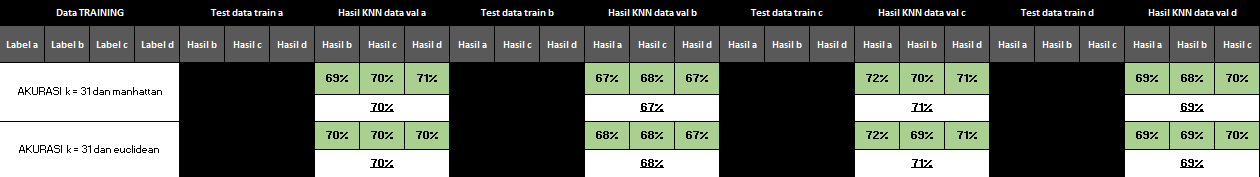
### Hasil Pengujian

## 

Gambar 5 hasil KNN yang di-write pada file hasil.csv

## BAB V

## KESIMPULAN



Gambar 6 K-fold validation dengan nilai k = 31 (best)

Pada hasil pengujian k-fold diatas dapat dilihat bahwa akurasi dari hasil validasi berbeda dan terlihat dengan menggunakan rumus manhattan menghasilkan beberapa yang lebih tinggi dari pada Euclidean distance. Tetapi setelah dirata2kan hasil Euclidean dan manhattan menghasilkan akurasi berbeda pada data test yang ke 2 (B1001 -B2000). Sehingga dapat disimpulkan dengan mengunakan rumus Euclidean distance lebih baik dari rumus manhattan dalam kasus ini.