**LAPORAN TUGAS BESAR**

**PEMBELAJARAN MESIN**

**SUPERVISED LEARNING**

Disusun untuk memenuhi Tugas Mata Kuliah Pembelajaran Mesin

****

Disusun Oleh :

 Aldiyan Farhan N

M Rifqi Wiliatama

PROGRAM STUDI S1 INFORMATIKA

FAKULTAS INFORMATIKA

UNIVERSITAS TELKOM

2021

1. **Formulasi Masalah**

Diberikan beberapa data set kemudian dilakukan pemilihan data set yang ada untuk melakukan pembelajaranan

1. **Explorasi dan Persiapan Data**

Pada bagian explorasi, dipilih dataset kendaraan\_train dan kendaraan\_test. Dataset kendaraan\_train memiliki 285.831 baris data dan kendaraan\_test memiliki 47.639 baris data. Pada data training, kolom id, kode\_daerah, umur kendaraan, kanal penjualan, dan sim tidak digunakan karena tidak berpengaruh pada proses learning.

Sebelum melakukan normalisasi, data diproses terlebih dahulu agar tidak terjadi kesalahan pada learning karena berbedanya tipe data, maka pada kolom “Jenis\_kelamin” apabila ‘Pria’ maka akan diberikan nilai 1, ‘Wanita’ diberikan nilai 0. Untuk data “Kendaraan\_rusak” apabila ‘Pernah’ diberikan nilai 1, ‘Tidak Pernah’ diberikan nilai 0. Kemudian dilakukan pengisian data yang memiliki kolom kosong dengan rata rata dari seluruh nilai pada kolom tersebut

1. **Pemodelan**

Pada pemodelan digunakan Google Colaboratory, Algoritma yang digunakan adakan K-Nearest Neighbor (K-NN). Pada tahap ini dilakukan beberapa tahap dalam proses pre-learning :

* 1. Tahap Pertama : pengambilan nilai k, nilai K yang digunakan adalah dengan
  2. Tahap Kedua : Perhitungan jarak dengan menggunakan Euclidian distance. Setelah ditemukan jarak setiap data test pada setiap data train, dilakukan sorting berdasalkan nilai paling kecil
  3. Tahap Ketiga : Penentuan hasil sorting kemudian data diolah menggunakan algoritma K-NN. Data test akan di tentukan hasil prediksinya berdasarkan tetangga yang paling besar nilai positif/negatifnya.
  4. Tahap Keempat : pengecekan hasil test dengan nilai sesungguhnya. Hasil test akan dicocokan dengan hasil pada data training. Jika data test sama dengan hasil data training maka nilai prediksi benar akan bertambah
  5. Tahap Kelima : penghitungan akurasi dengan totol nilai prediksi benar dan salah akan digabungkan untuk menghitung akurasi dari hasil learining KNN

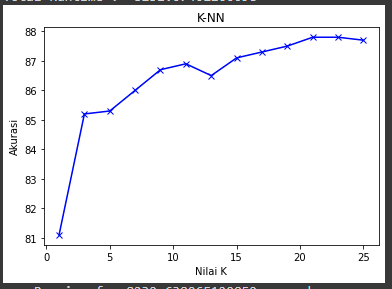
1. **Evaluasi**

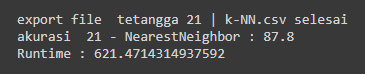
Bagian yang menjadi penilaian adalah tingkat akurasi learning yang menentukan seberapa akurat hasil data train terhadap testing

1. **Experimen**

Pada metode klasifikasi dengan algoritma KNN, banyak data yang digunakan adalah 5000 data training dan 1000 data test. Banyaknya tetangga yang menjadi acuan adalah 12 kelompok tetangga (1,3,5,7,…,25). Perhitungan jarak dengan menggunakan rumus Euclidian distance.

Dari aturan dan batasan experiment, didapatkan hasil sebagai berikut





Pada algorima dan batasan yang digunakan bisa terlihat bahwa akurasi terbaik berada pada k = 21 dengan nilai akurasi 87,8 yang menjadikan hasil learning terbaik adalah KNN dengan banyak tetangga = 21

1. **Kesimpulan**

Dari hasil experiment yang telah dilakukan, didapat bahwa akurasi yang paling efektif adalah dengan nilai 87,8 dengan banyak tetangga (K-neighbor) = 21. Banyaknya data dan banyaknya tetangga mempengaruhi hasil learning KNN. Semakin banyak data yang digunakan maka akan semakin lama waktu exekusi program (Walau dengan supercomputer sekalipun). Untuk perhitungan jarak apabila digunakan perhitungan diluar Euclidian distance juga dapat merubah hasil dari learning.