

Nama : Shafa Auliya

NPM : 2117051042

Kelas : CD

Analisis Jurnal dengan Dataset Stroke

Jurnal 1 :

Sutomo, F., Muaafii, D. A., Al Rasyid, D. N., Kurniawan, Y. I., Afuan, L., Cahyono, T., Maryanto, E., & Iskandar, D. (2023). Optimization of the K-nearest neighbors algorithm using the elbow method on stroke prediction. *Jurnal Teknik Informatika (Jutif)*, 4(1), 125–130. <https://doi.org/10.52436/1.jutif.2023.4.1.839>

Analisis :

- Pada jurnal 1 menggunakan model algoritma *K-Nearest Neighbors (KNN)* menggunakan Elbow Method untuk menentukan nilai k optimum pada dataset *stroke prediction*.
- Pada penelitian ini menggunakan dataset yang diambil dari Kaggle, dengan data yang terdiri dari 5110 dengan 12 kolom/atribut. Dengan atribut seperti ID, Gender, Age, Hypertension, Heart disae, Martial status, Work type, Residence, area, Avg glucose, BMI, Smoking, status, Stroke status.
- Dataset ini memiliki enam atribut binary, delapan atribut dalam kategorikal dan tiga atribut dalam bentuk numerik.
- Proses untuk mengelola data sebelum dilakukan klasifikasi menggunakan KNN :

a) Data Cleaning

Dengan cara membuang duplikasi data, memeriksa data yang inkosisten, dan memperbaiki kesalahan pada data.

b) Feature Selection

Memilih feature yang berpengaruh dan mengesampingkan feature yang tidak berpengaruh dalam pemodelan atau penganalisis data.

c) Feature Encoding

Proses data yang telah dipilih akan diubah kedalam bentuk dimana data dapat diproses dalam data mining.

- Untuk menentukan nilai K dengan cara mengambil jumlah tetangga paling dekat.
- Pada penelitian ini juga mengambil Langkah awal dengan membagi data di setiap kasus menjadi 2 yaitu data training dan data testing.

- Untuk preprocessing data pada dataset ini menggunakan data cleaning, feature selection, feature encoding.
- Hasil pengujian KNN dengan menggunakan nilai $k = 3$, dengan didapatkan nilai akurasi tertinggi sebesar 78% yang dimiliki oleh model splitting data 20-80 dan model splitting data untuk 30-70.
- Dengan menggunakan Elbow Method, untuk nilai K dengan cara melakukan perbandingan nilai MSE, dengan nilai MSE yang minimum akan ditentukan nilai K nya sebagai nilai K optimum. Pada dataset ini dengan split data 20-80 mendapatkan akurasi 84%, split data 25-75 mendapatkan nilai akurasi 81% dan split data 30-70 mendapatkan nilai akurasi 82%.

Paper 2 :

Akbar, F., Saputra, H. W., Maulaya, A. K., Hidayat, M. F., & Rahmaddeni, R. (2022). Implementasi Algoritma decision Tree C4.5 Dan Support Vector Regression Untuk Prediksi Penyakit stroke. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 2(2), 61–67. <https://doi.org/10.57152/malcom.v2i2.426>

Analisis :

- Pada penelitian ini menggunakan model Algoritma Support Vector Regression dan Decision Tree C4.5.
- Pada dataset ini bersumber dari Kaggle. Dengan memiliki 12 fitur, yaitu id, gender, age, hypertension, heart_disease, ever_married, work_type, Residence_type, avg_glucose_level, bmi, smoking_status, dan stroke dan dengan memiliki 5110 dataset.
- Untuk preprocessing seperti :
 - a) Cleaning data

Proses menghilangkan noise dan data yang tidak relevan atau tidak konsisten.
 - b) Splitting data

Membagi data menjadi 2 bagian seperti data latih dan data uji, dengan menggunakan data split, percobaan pelatihan dilakukan berdasarkan rasio split 70:30, dan rasio split dari data pelatihan yang tersisa dihitung sebagai data uji.
- Untuk evaluasi data dilakukan yaitu membandingkan nilai eror masing-masing algoritma dengan menggunakan data yang sama dan splitting data yang sama yaitu 70:30
- Untuk metode Decision Tree C4.5 menghasilkan MAE yaitu 0.078 ± 0.221 dan MSE yaitu 0.055 ± 0.209
- Untuk metode SVR menghasilkan MAE yaitu 0.396 ± 0.045 dan MSE yaitu 0.159 ± 0.044
- Dengan dataset stroke prediction lebih bagus menggunakan model Decision Tree C4.5.

Paper 3 :

Karim, F., Nurcahyo, G. W., & Sumijan, S. (2021). Sistem Pakar Dalam mengidentifikasi gejala stroke menggunakan metode naive Bayes. *Jurnal Sistim Informasi Dan Teknologi*, 3(4), 221–226. <https://doi.org/10.37034/jsisfotek.v3i4.69>

Analisis :

- Metode yang digunakan adalah naivebayes dengan cara mengklasifikasikan data, menentukan variable, membuat rule dan melakukan perhitungan metode naïve bayes.
- Klasifikasi data pada penelitian ini menggunakan data yang bersumber dari Rumah Sakit Otak Dr. Drs. M. Hatta Bukittinggi. Data tersebut akan diklasifikasi mejadi beberapa bagian yaitu data jenis stroke, data gejala penyakit dan data rule dari seorang pakar.
- Menentukan variable dengan membuat 2 bagian yang akan diberikan kode yaitu data jenis penyakit stroke dan data gejala penyakit stroke.
- Menentukan role berisi nilai-nilai gejala terhadap jenis penyakitnya sehingga menjadi kaidah acuan.
- Dengan menghitung dataset mengguakan rumus-rumus dari naïve bayes didapatkan akurasi sebesar 80%.

Link Colab Untuk Deteksi Penyakit Stroke menggunakan Model KNN:

https://colab.research.google.com/drive/1TmL_Kr0sJ8TZS4biIAM4W4eoHL4y3kW?usp=sharig