

DATA MINING MACHINE LEARNING

GROUP : 10



Anggota

- ▶ A. Afif Alhaq H071221064
- ▶ Adrian Hidayat H071221091
- ▶ Evan Pandu Nata H071221057
- ▶ Muhammad Rifky Aqid H071211077



Deskripsi

- Dataset ini disediakan pada Kaggle yang berisi data untuk memprediksi risiko stroke juga dapat dilihat dari data-data konkret
- Data ini terdiri atas 5110 data yang diperoleh dari id, gender, age, hypertension, heart_disease, ever_married, work_type, Residence_type, avg_glucose_level, bmi, smoking_status, stroke.
- Tujuan analisis adalah untuk membantu para profesional medis dalam memberikan peringatan dini kepada pasien dengan risiko tinggi, sehingga tindakan pencegahan dapat diambil lebih awal

Proses Pekan Pertama

Belum dikerjakan

- laporan Akhir

Sedang Dikerjakan

- Menguji model training
- Membuat model deployment

Selesai dikerjakan

- Brainstorming Ide
- Mencari dan Menentukan dataset
- Membuat repository github
- Melakukan eksplorasi data
- Melakukan data preprocessing

Proses Pekan Kedua

Belum dikerjakan

- membuat laporan akhir

Sedang Dikerjakan

- Menentukan model training
- Memperbaiki model deployment

Selesai dikerjakan

- Brainstorming Ide
- Mencari dan Menentukan dataset
- Membuat repository github
- Melakukan eksplorasi data
- Melakukan data preprocessing
- Menguji model training
- Membuat model deployment



Proses Pekan Ketiga

Belum dikerjakan

Sedang Dikerjakan

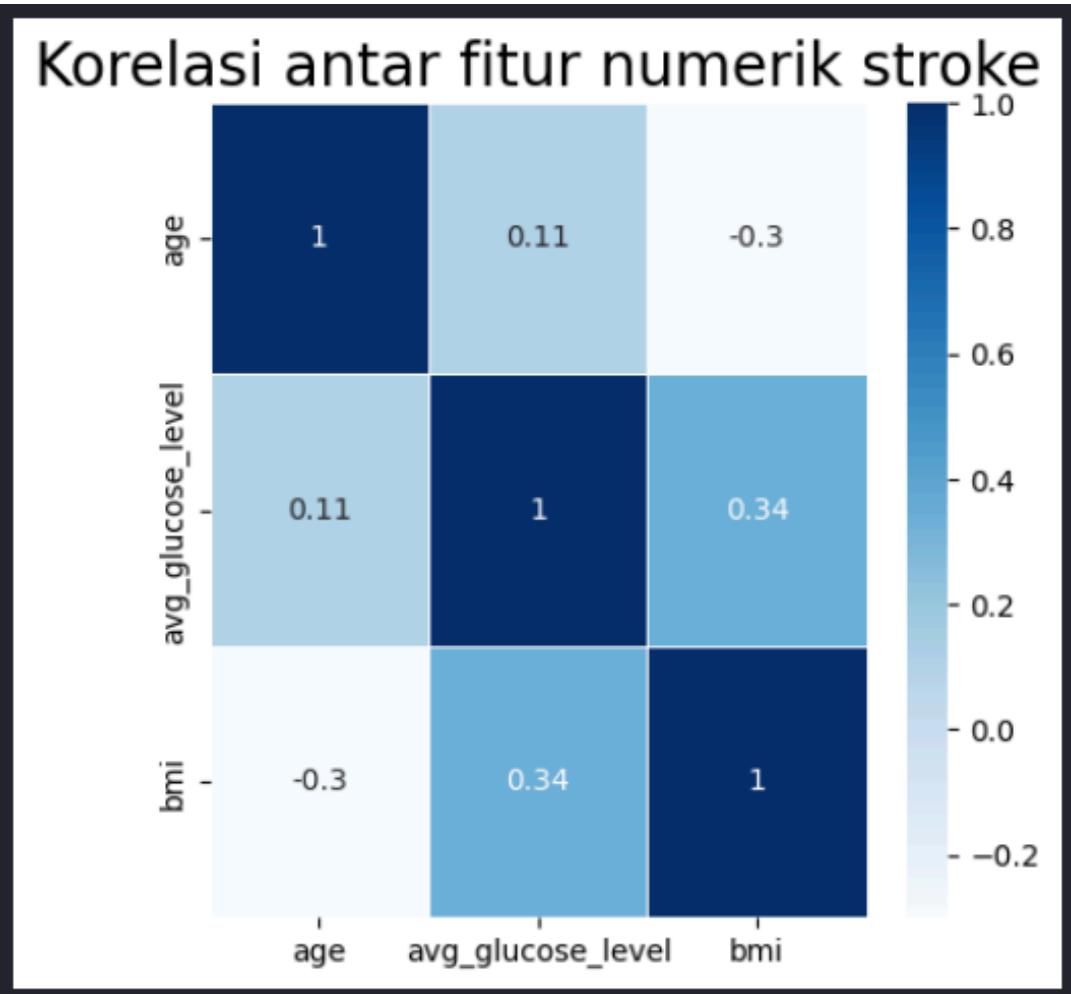
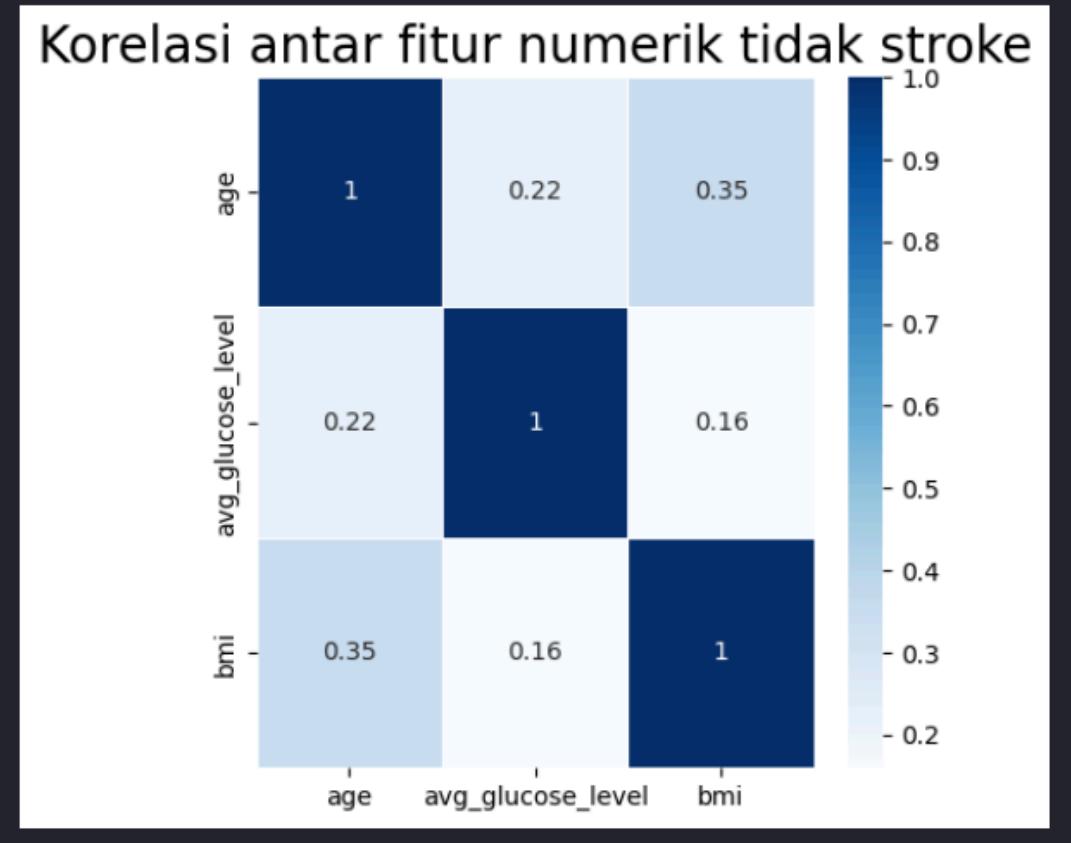
- Membuat laporan akhir

Selesai dikerjakan

- Brainstorming Ide
- Mencari dan Menentukan dataset
- Membuat repository github
- Melakukan eksplorasi data
- Melakukan data preprocessing
- Menguji model training
- Membuat model deployment
- Manajemen directory
- Perbaikan dokumentasi kode

Korelasi

Melihat korelasi antar fitur numerikal untuk pasien stroke.



Modelling

```
Model: KNN
Best score: 0.9739819004524886
Best params: {'knn_leaf_size': 10, 'knn_n_neighbors': 5, 'knn_p': 1, 'knn_weights': 'distance'}

Model: RandomForest
Best score: 0.9787330316742082
Best params: {'rf_max_depth': 30, 'rf_min_samples_leaf': 1, 'rf_min_samples_split': 5, 'rf_n_estimators': 400}

Model: Boosting
Best score: 0.9678733031674207
Best params: {'boost_learning_rate': 0.3, 'boost_n_estimators': 400}
```

Melihat parameter terbaik pada tiap model machine learning yang digunakan dari hasil grid search

DEMO APLIKASI



THANK YOU!