

TUGAS KELOMPOK 6



DAFTAR ANGGOTA



Meinisa



Sella Dianka Fitri



Anissa Luthfi Alifia



Nadia Fitri Yani



Angelica Noviana



Ibnu Farhan Al-Ghifari

LATAR BELAKANG

Mammalian Target of Rapamycin (mTOR) adalah enzim serin/treonin kinase yang memiliki peran penting dalam regulasi pertumbuhan, metabolisme sel, dan fungsi otak. Jalur sinyal mTOR menjadi pusat integrasi antara sel dengan lingkungannya dan berkontribusi dalam perkembangan abnormal sel yang berkaitan dengan berbagai penyakit, termasuk kanker. Karena perannya yang signifikan, mTOR menjadi target utama dalam pengembangan terapi berbasis molekuler. Untuk mendukung terapi ini, diperlukan metode prediktif untuk mengevaluasi bioaktivitas senyawa inhibitor mTOR. Eksperimen biologis yang konvensional membutuhkan banyak waktu dan biaya, sehingga pendekatan komputasional, seperti Artificial Neural Network (ANN) dan Random Forest Regressor (RFR), menjadi solusi yang efisien.

TUJUAN PENELITIAN

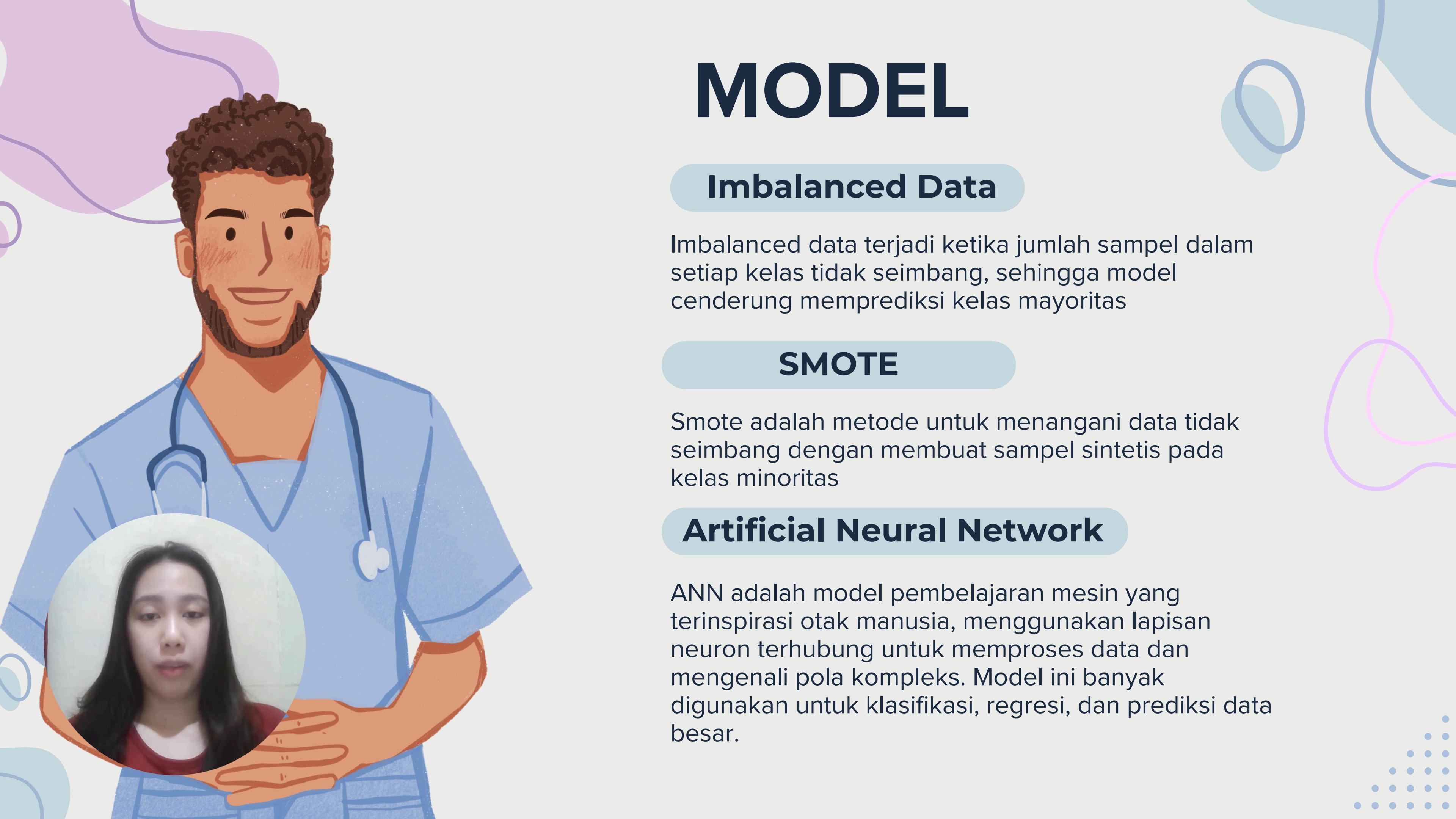
Penelitian ini bertujuan untuk mengevaluasi kemampuan ANN dalam memprediksi bioaktivitas senyawa terhadap protein mTOR menggunakan dataset yang mencakup deskriptor kimia dan nilai bioaktivitas (IC₅₀) yang diseimbangkan dengan teknik SMOTE. Dengan penelitian ini, diharapkan dapat memberikan kontribusi pada pengembangan metode komputasional untuk prediksi bioaktivitas senyawa.

METODE



DATA

Dataset ini diperoleh dari Database ChEMBL, dengan fokus pada target MTOR. Bioaktivitas diukur melalui nilai IC50 (konsentrasi yang menghambat 50% aktivitas target). Proses pengumpulan melibatkan pencarian target (target.search('mTOR')), penyaringan data bioaktivitas (activity.filter), dan penggabungan subset data menjadi satu DataFrame. Dataset akhir mencakup 4185 senyawa,

A stylized illustration of a doctor with curly brown hair and a beard, wearing blue scrubs and a stethoscope. He is holding a young girl's hand, which is shown in a circular inset. The background features abstract, colorful wavy shapes.

MODEL

Imbalanced Data

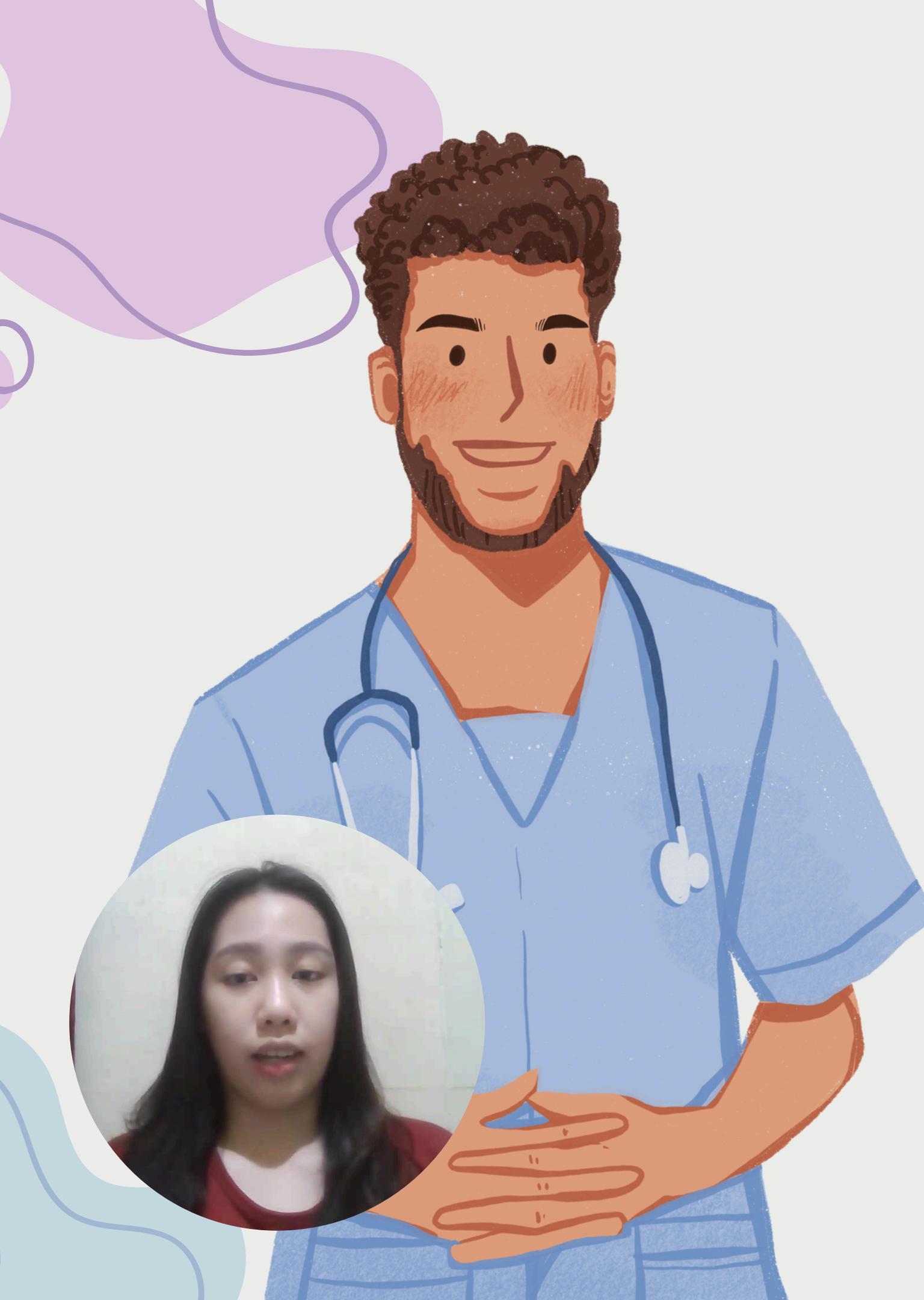
Imbalanced data terjadi ketika jumlah sampel dalam setiap kelas tidak seimbang, sehingga model cenderung memprediksi kelas mayoritas

SMOTE

Smote adalah metode untuk menangani data tidak seimbang dengan membuat sampel sintetis pada kelas minoritas

Artificial Neural Network

ANN adalah model pembelajaran mesin yang terinspirasi otak manusia, menggunakan lapisan neuron terhubung untuk memproses data dan mengenali pola kompleks. Model ini banyak digunakan untuk klasifikasi, regresi, dan prediksi data besar.



MODEL

Random Forest Regressor

Algoritma ini membangun sejumlah decision tree dari berbagai subset data dan menggabungkan prediksinya untuk menghasilkan estimasi akhir yang lebih stabil dan akurat.

XGBoost

XGBoost adalah metode canggih yang berbasis pada Gradient Tree Boosting, yang dirancang untuk bekerja secara efisien dalam menangani masalah skala besar meskipun dengan sumber daya komputasi yang terbatas

Pelatihan Model dan Evaluasi

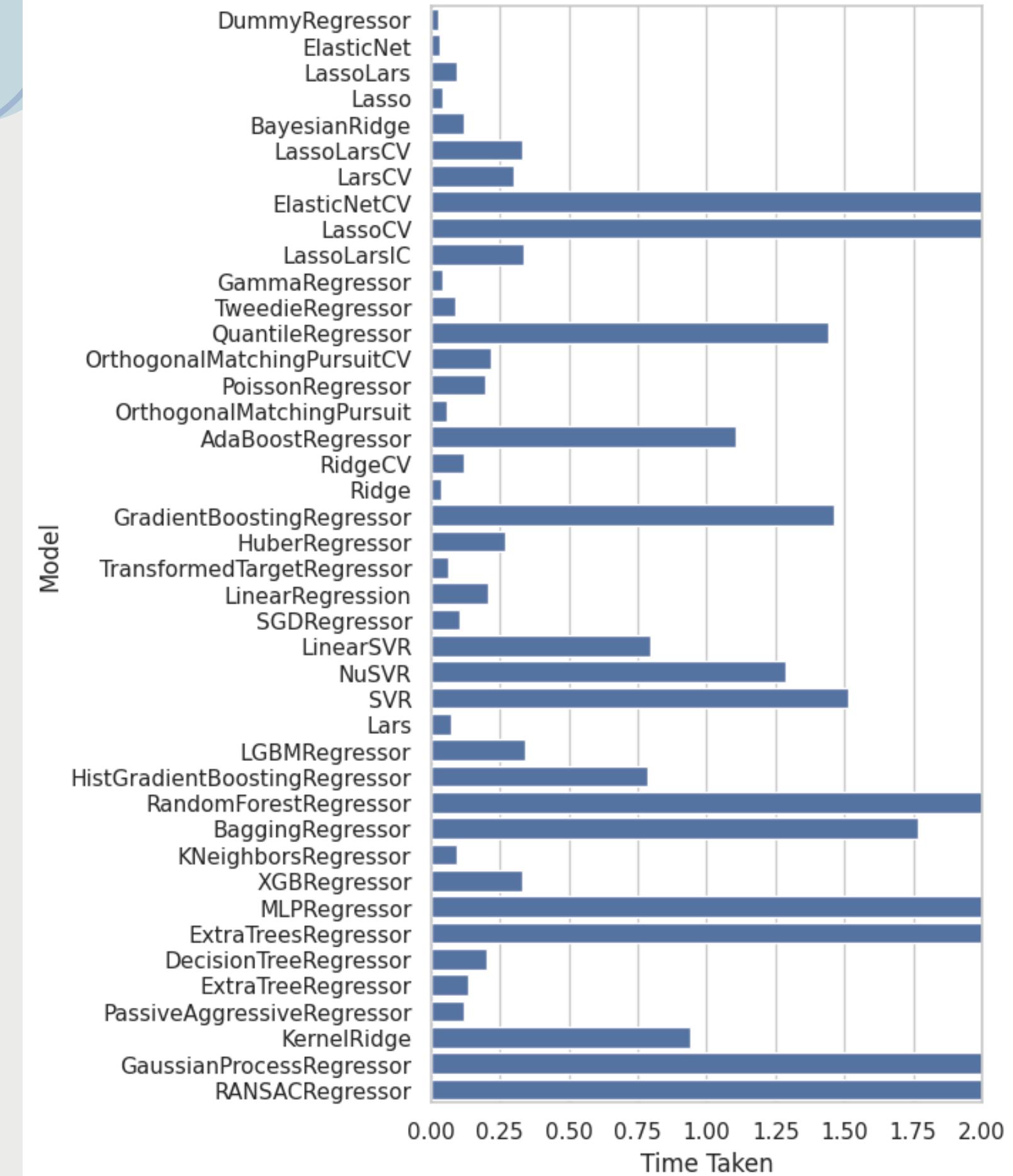
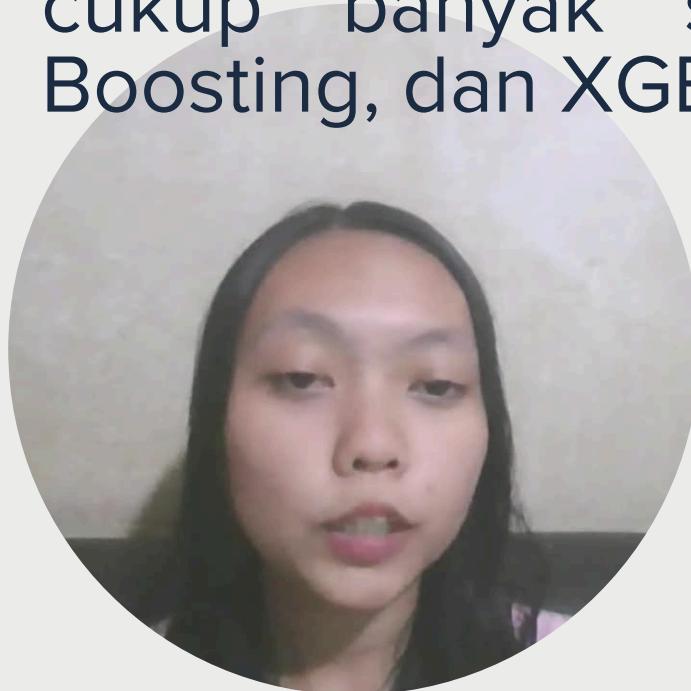
Untuk mengevaluasi kinerja model, metrik yang digunakan adalah classification report yang mencakup precision, recall, F1-score, dan support

HASIL DAN PEMBAHASAN

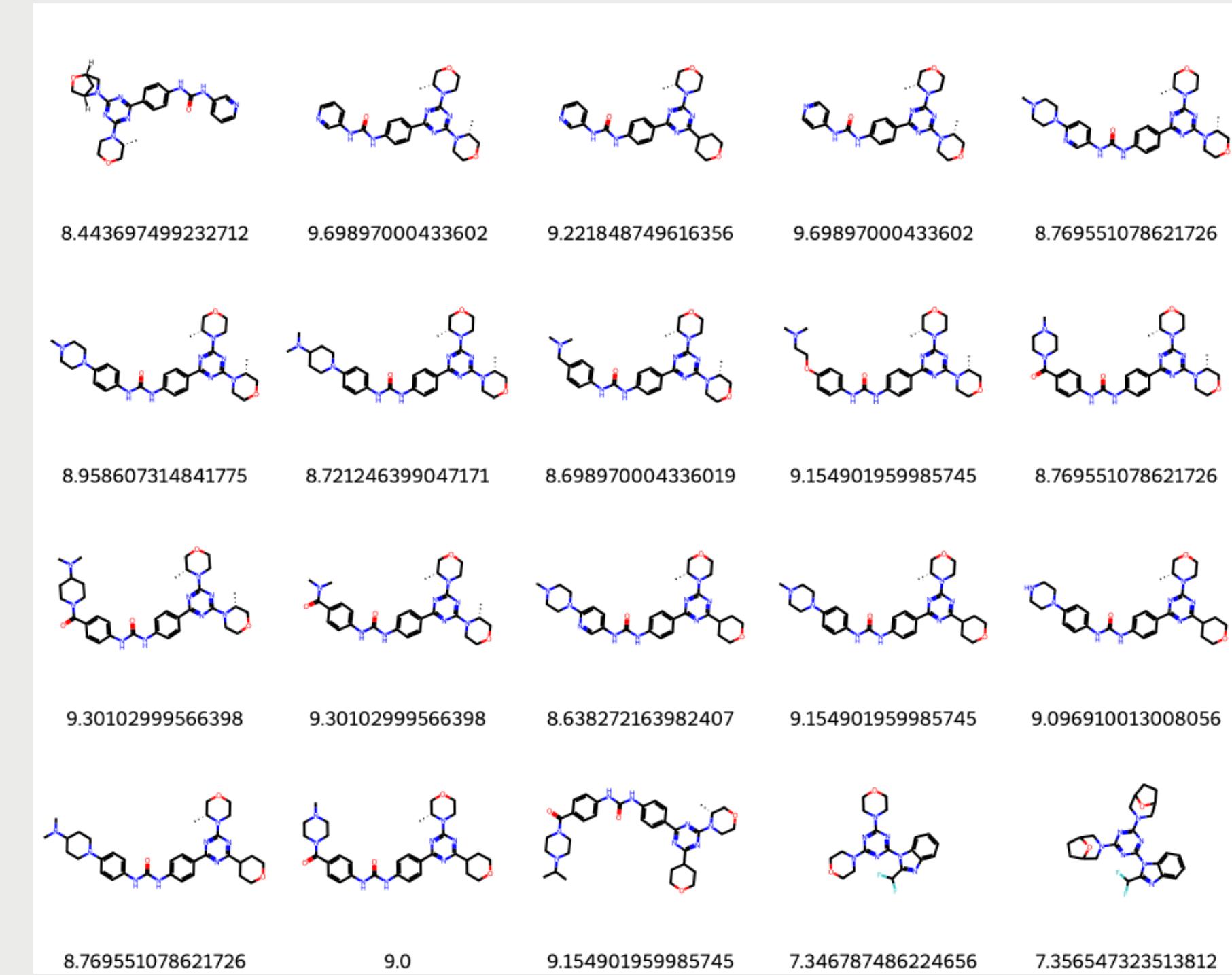


PLOT CALCULATION TIME

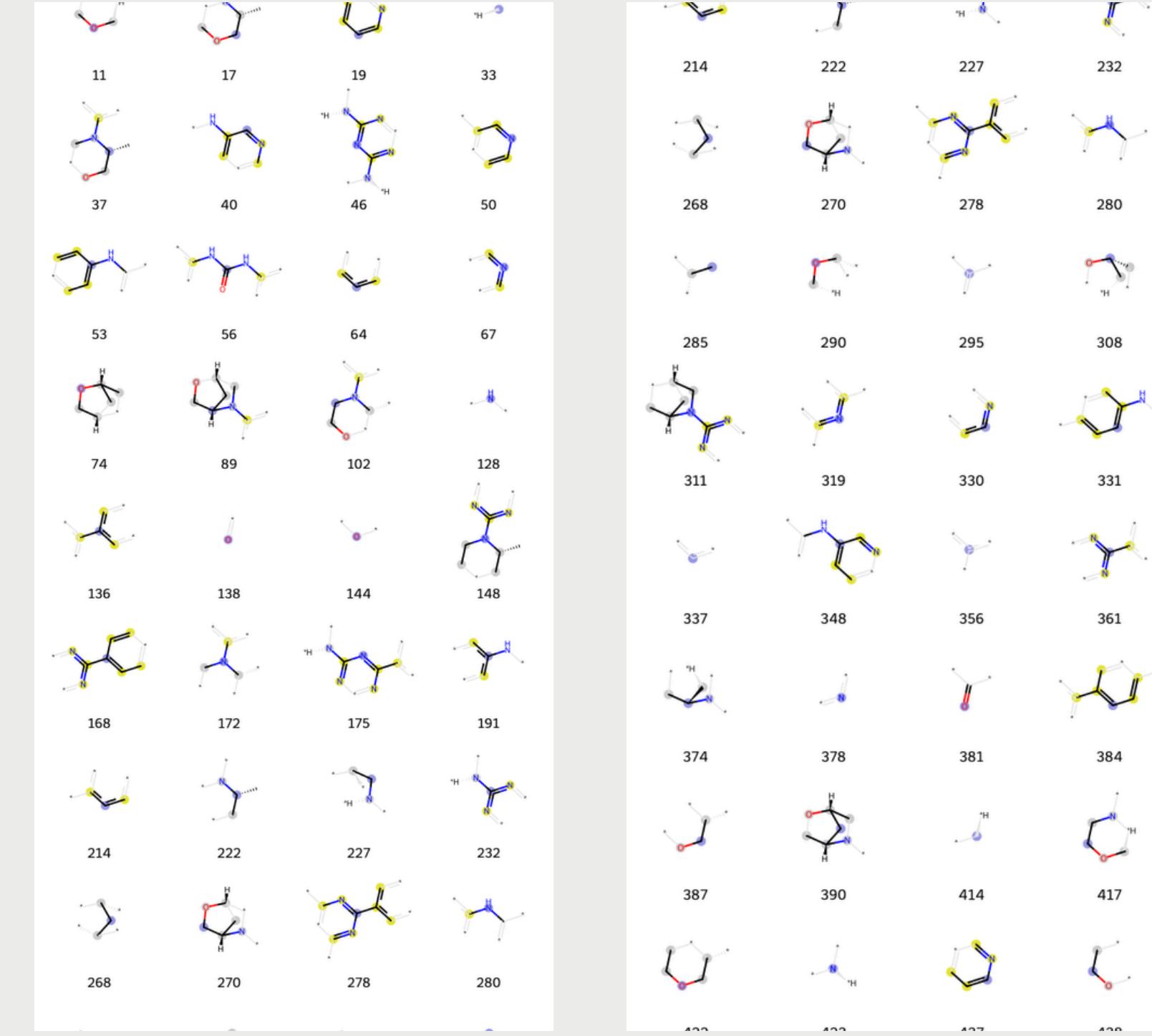
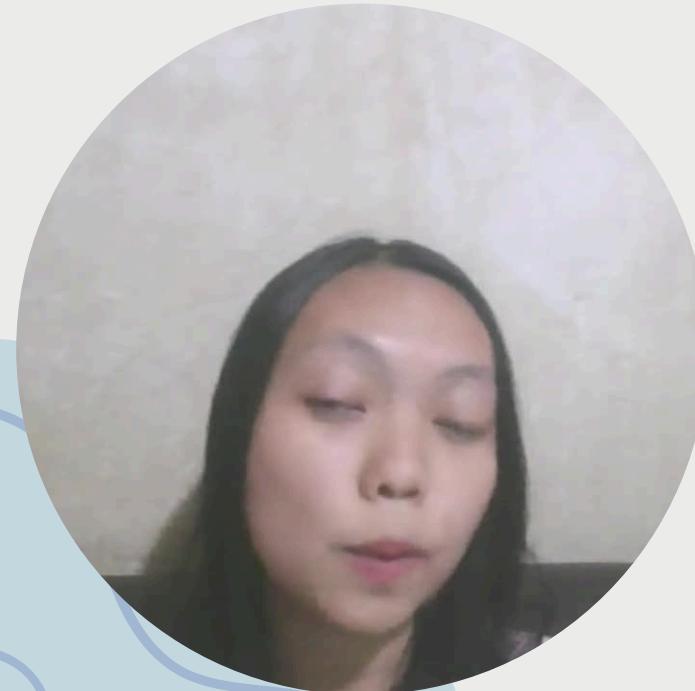
Dapat dilihat perbedaan kalkulasi waktu untuk berbagai model, di mana model dengan waktu komputasi tercepat adalah model Dummy Regressor, hal ini terjadi karena DummyRegressor bekerja dengan cara yang sangat sederhana, yaitu dengan memberikan prediksi yang sama untuk semua data. Untuk model dengan waktu komputasi yang lama ada cukup banyak seperti Random Forest, Gradient Boosting, dan XGBoost.



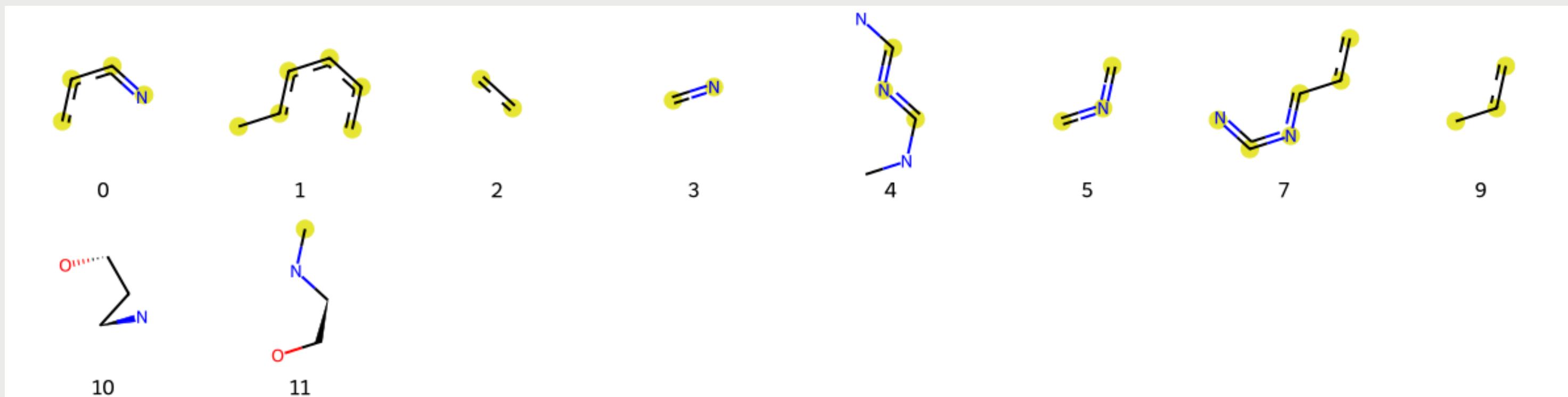
VISUALISASI STRUKTUR MOLEKUL DALAM GRID



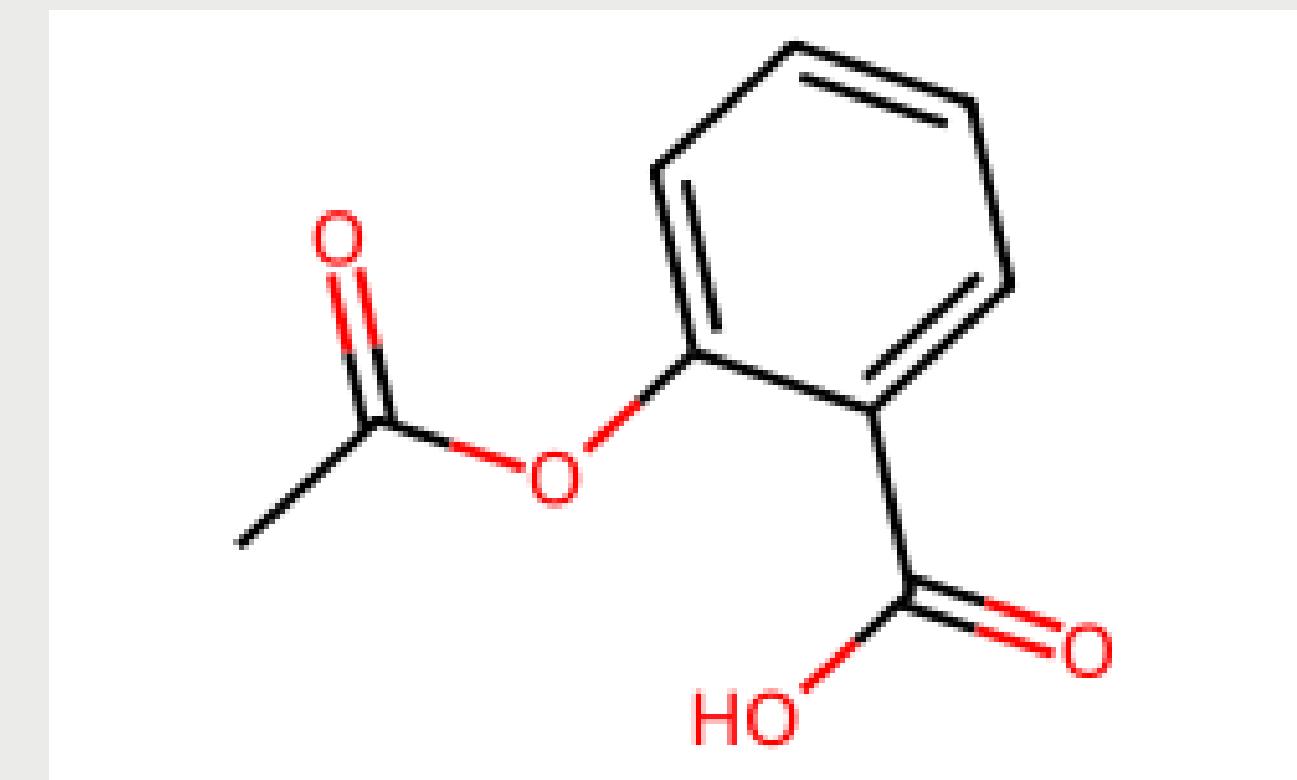
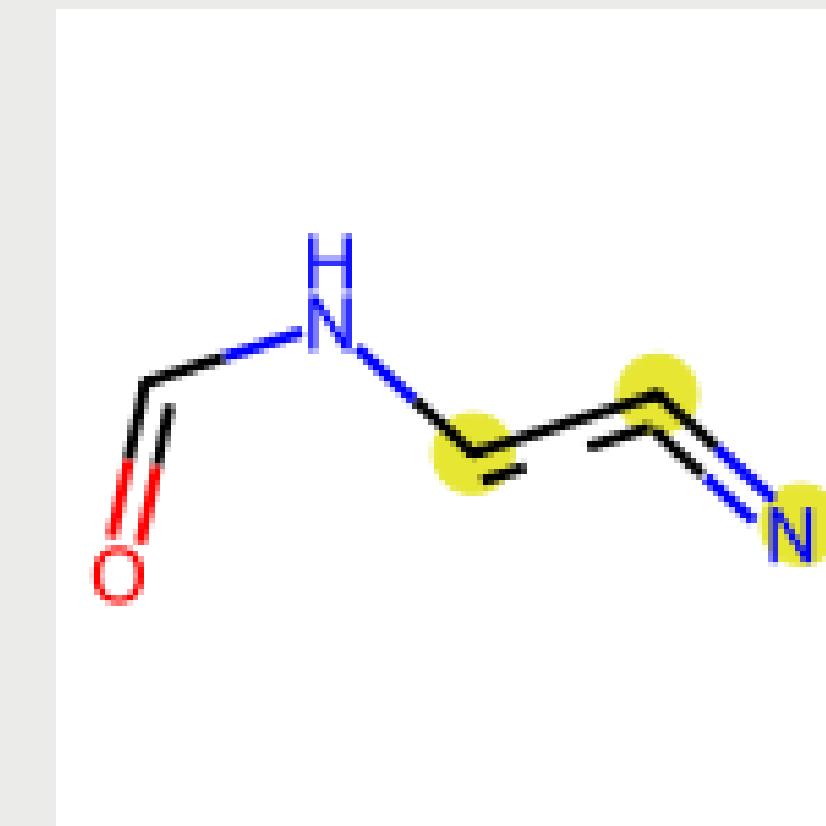
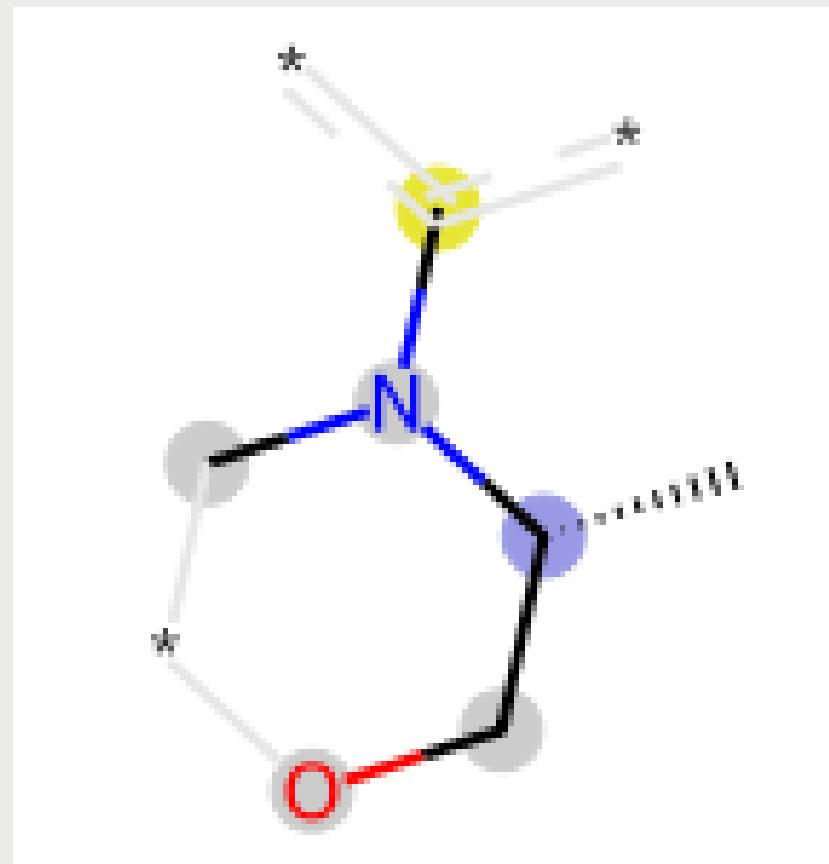
VISUALISASI SEMUA BITS YANG AKTIF



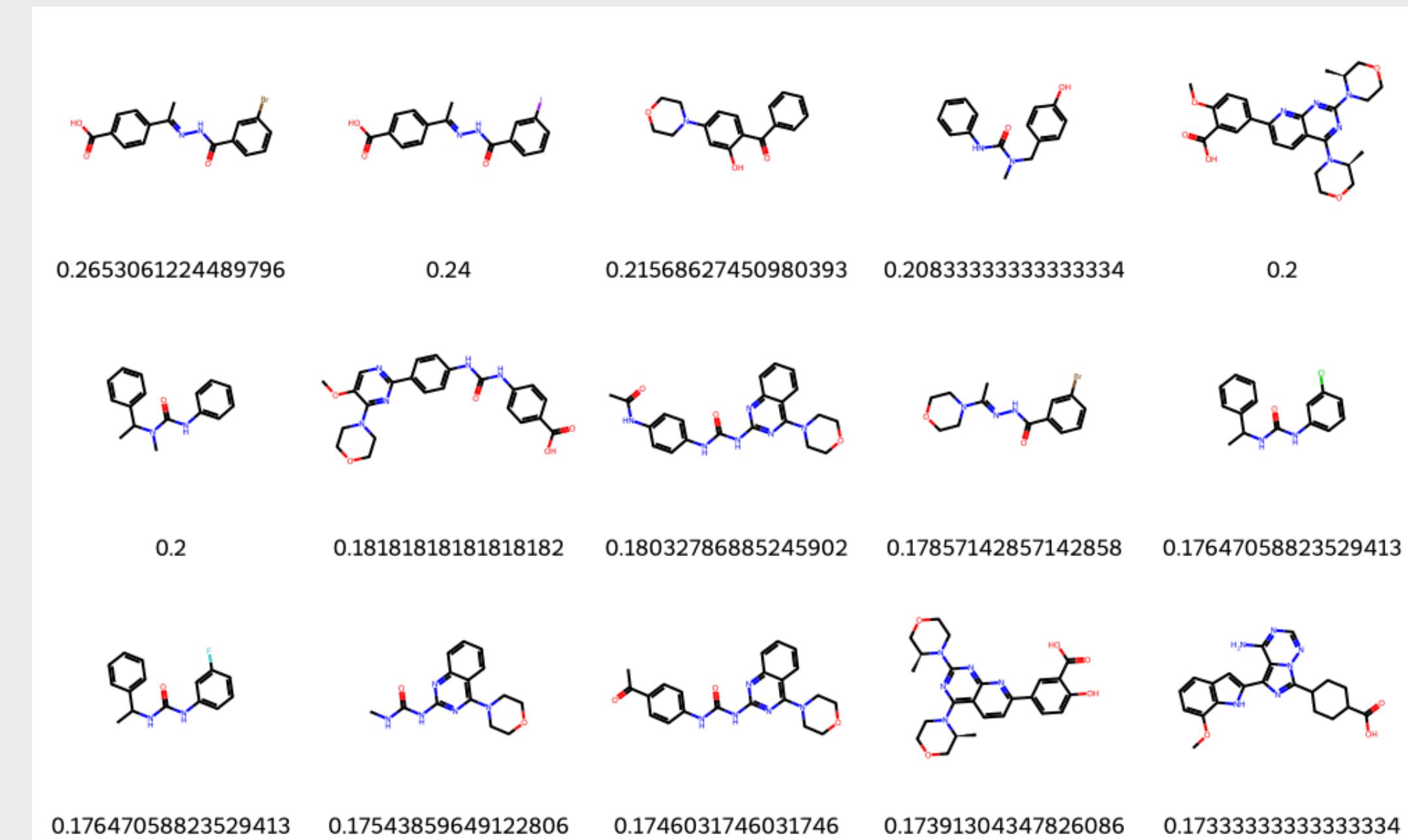
VISUALISASI FRAGMEN DARI PROSES RDKIT



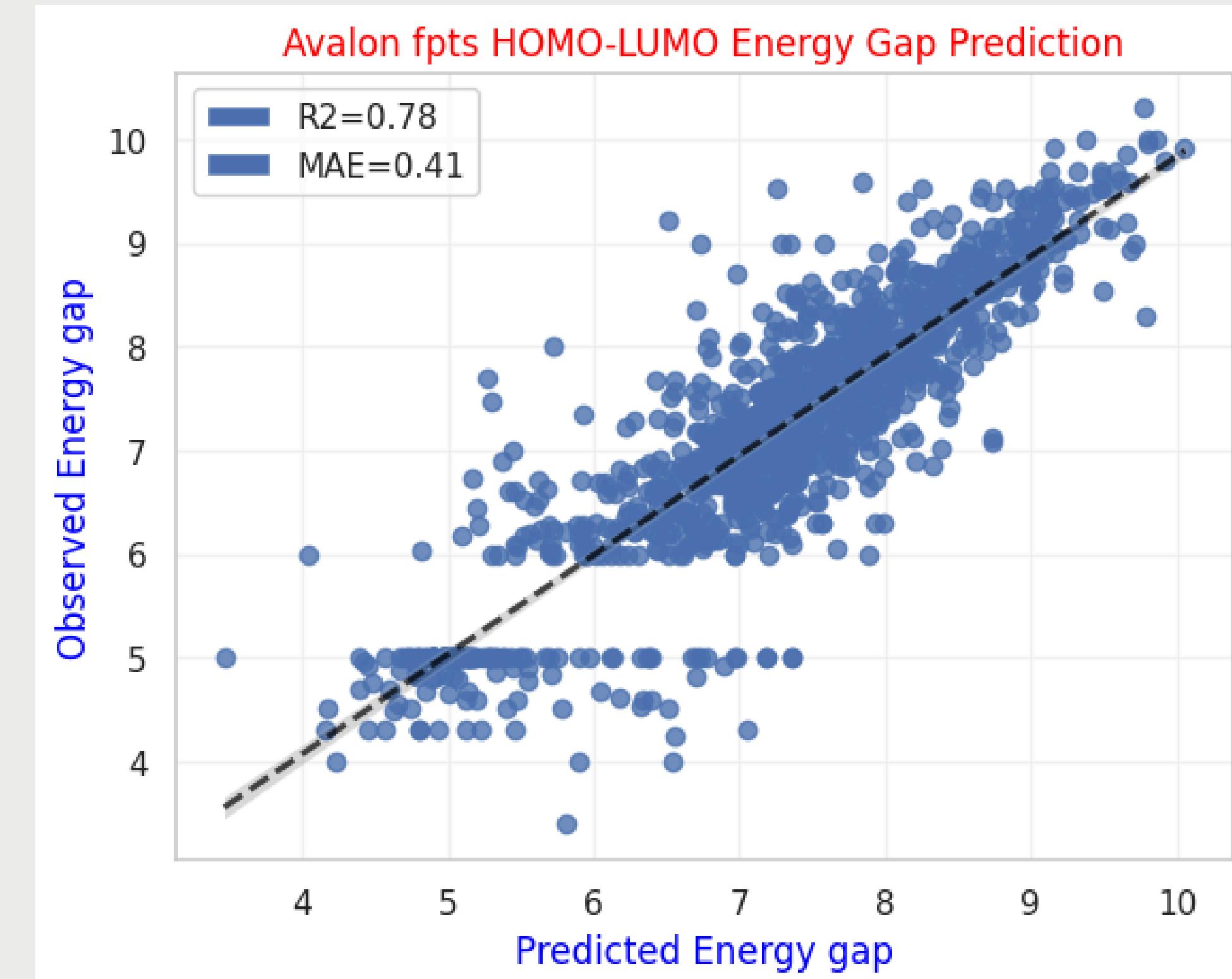
PENCARIAN SENYAWA MIRIP DENGAN KUERI



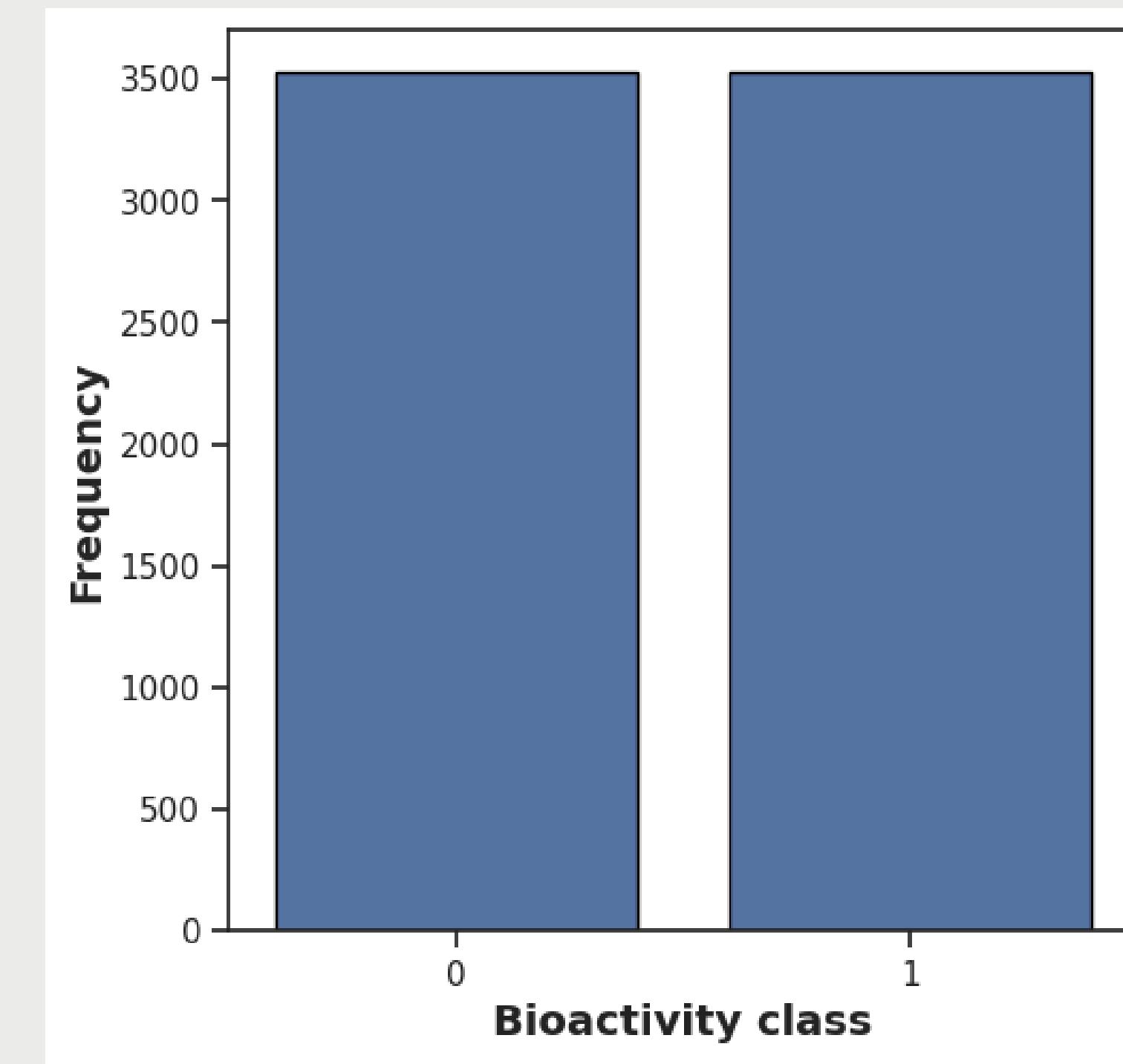
STRUKTUR DAN KOEFISIEN NILAI TANIMOTO



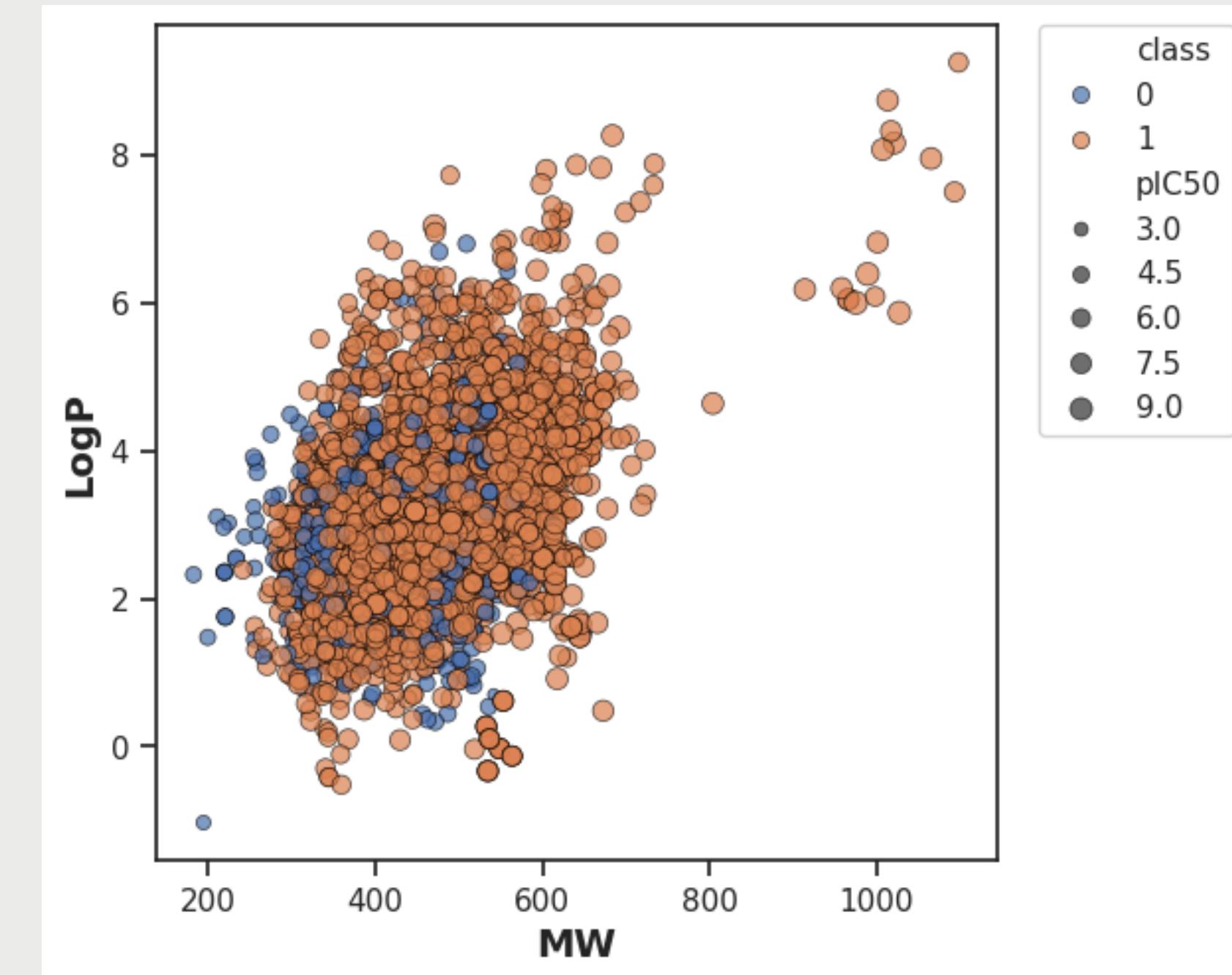
VISUALISASI PREDIKSI CELAH ENERGI HOMO- LUMO



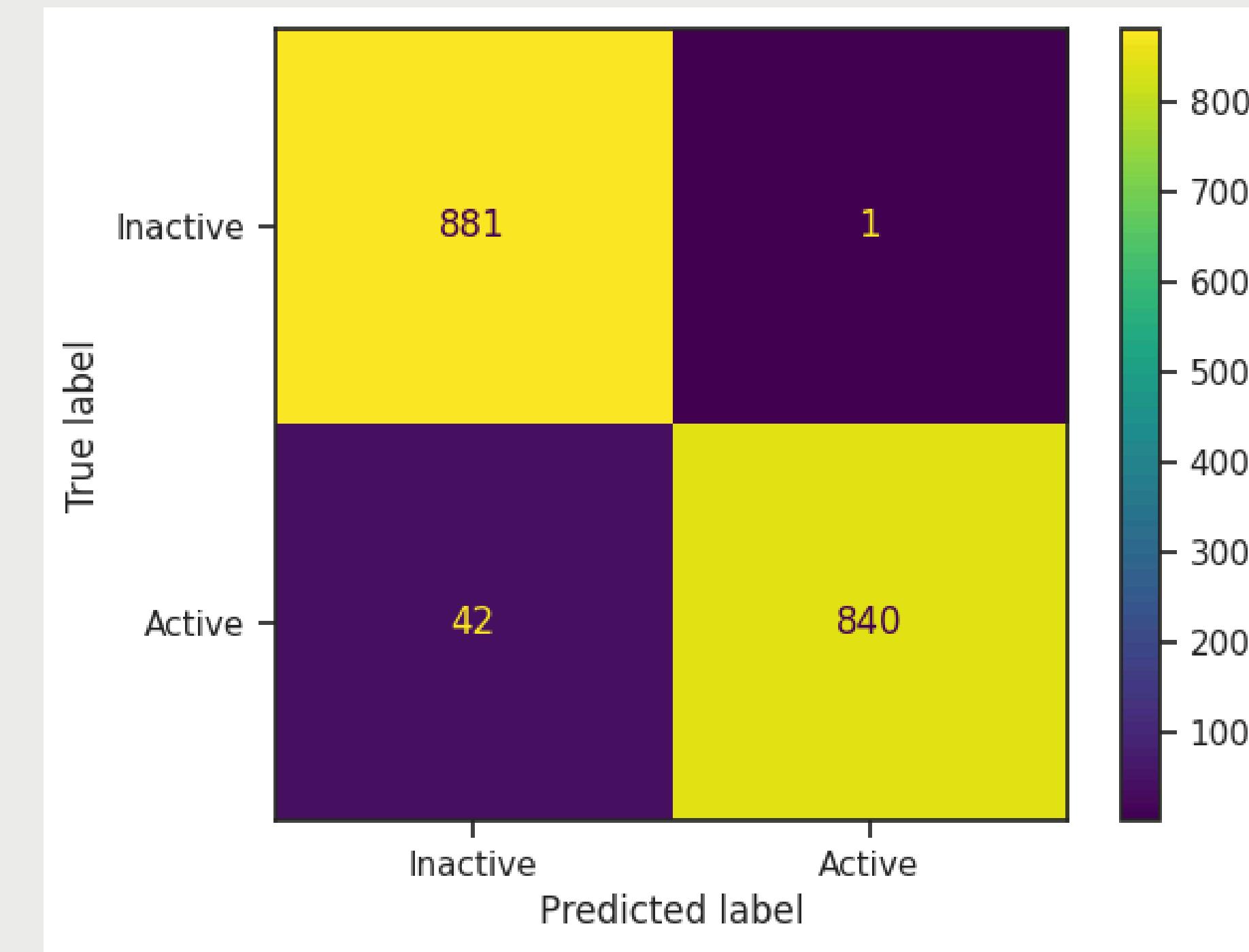
VISUALISASI DATA SMOTE



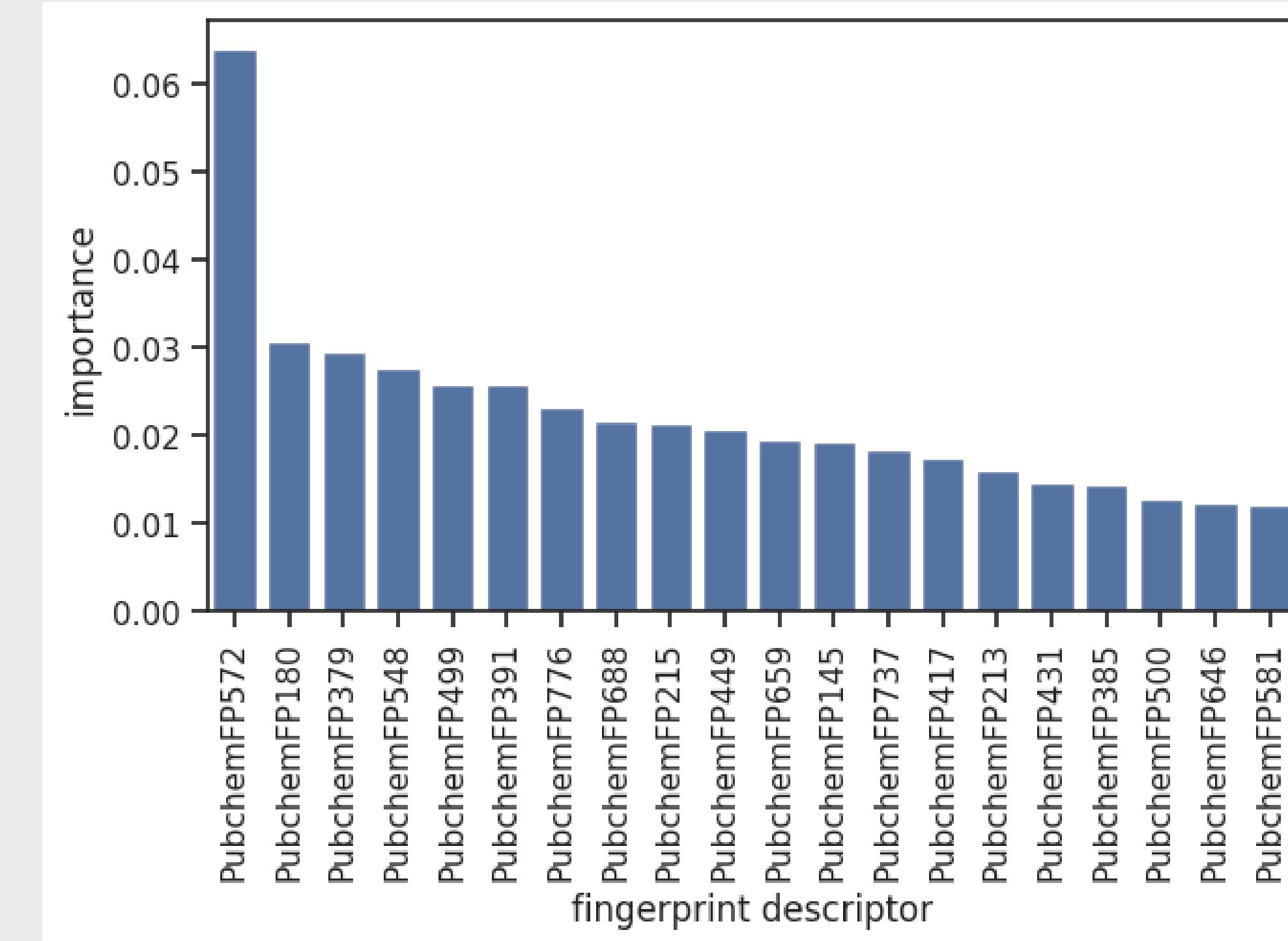
VISUALISASI LOGP DAN MW UNTUK SMOTE



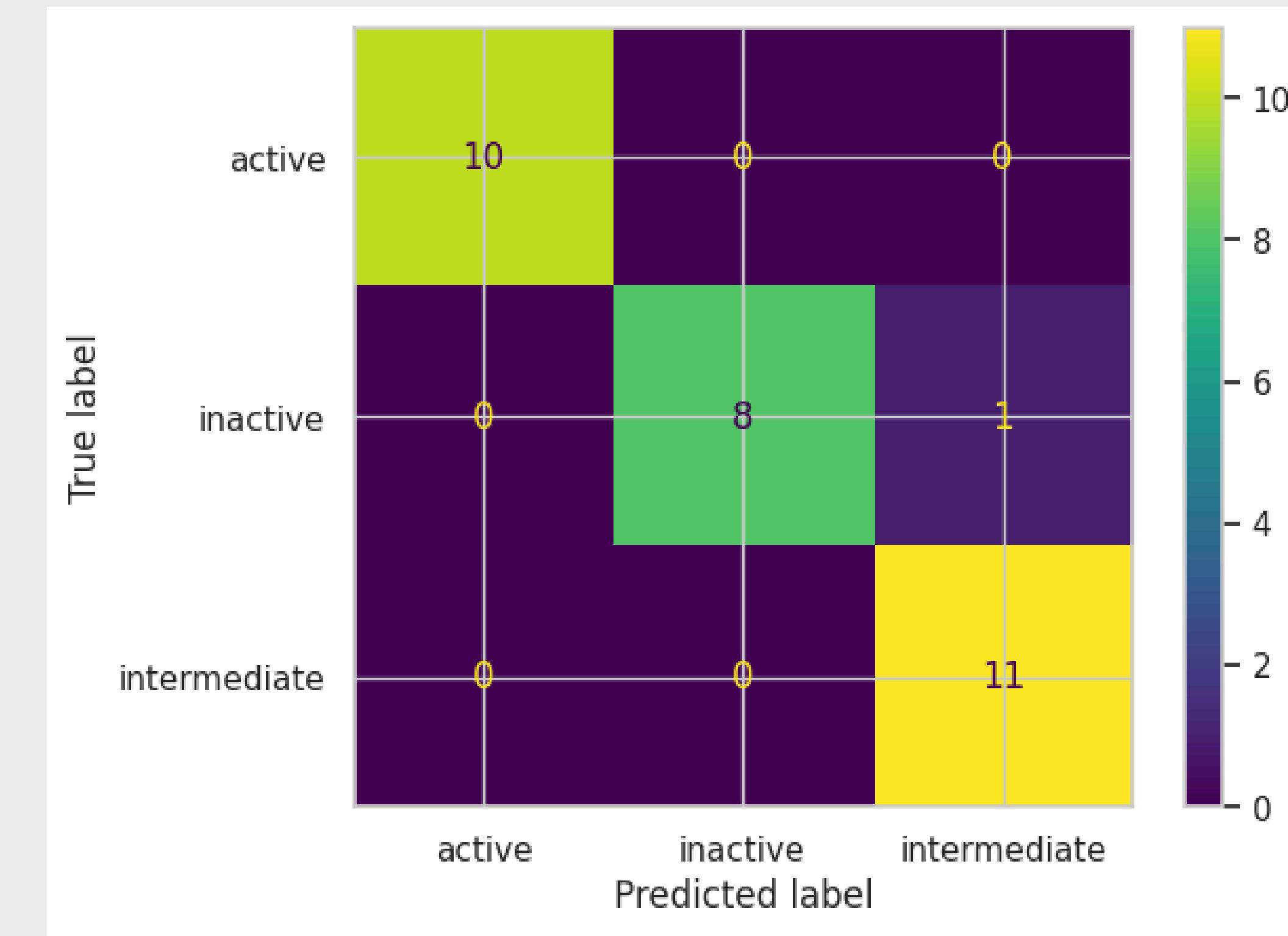
CONFUSION MATRIX XGBOOST



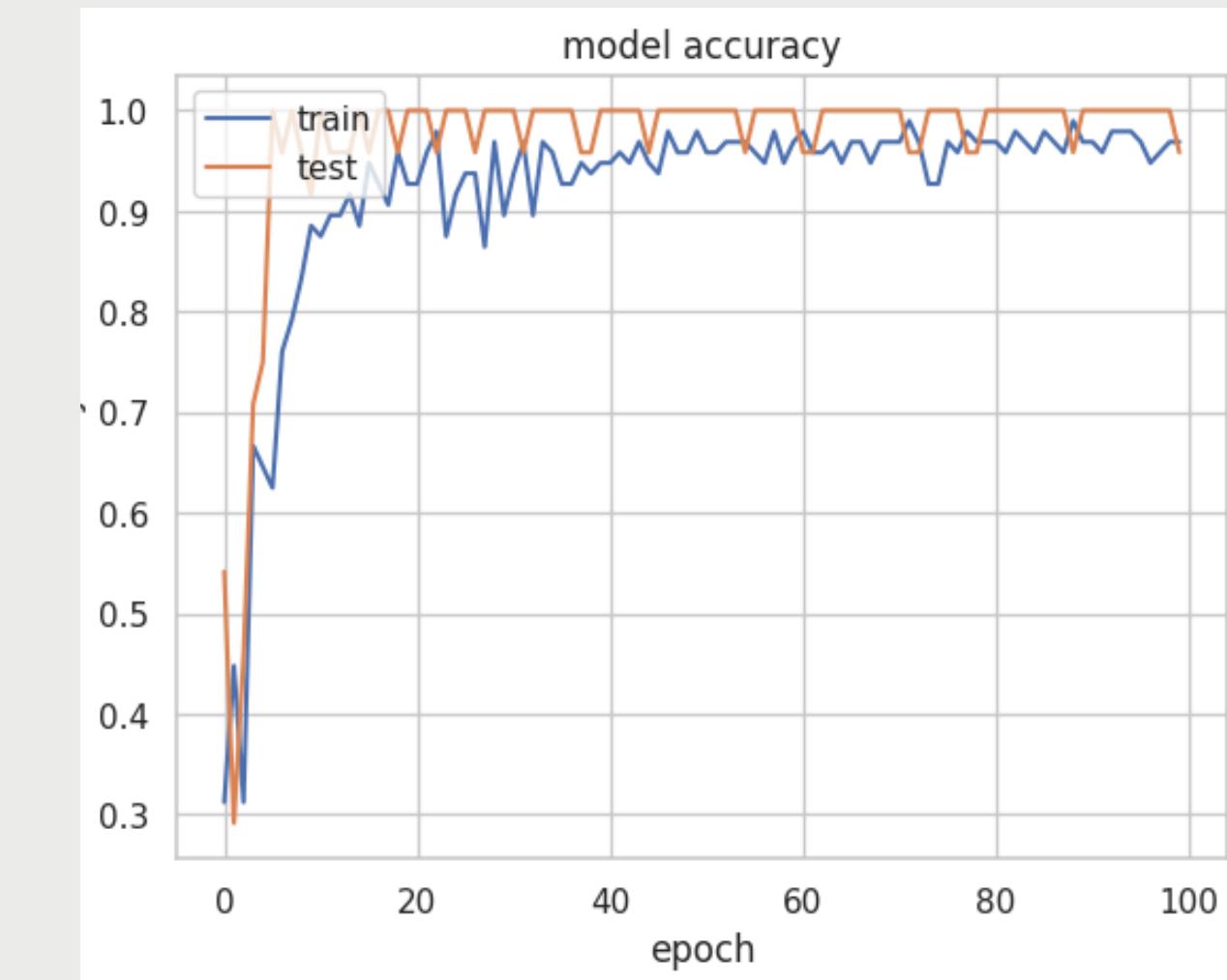
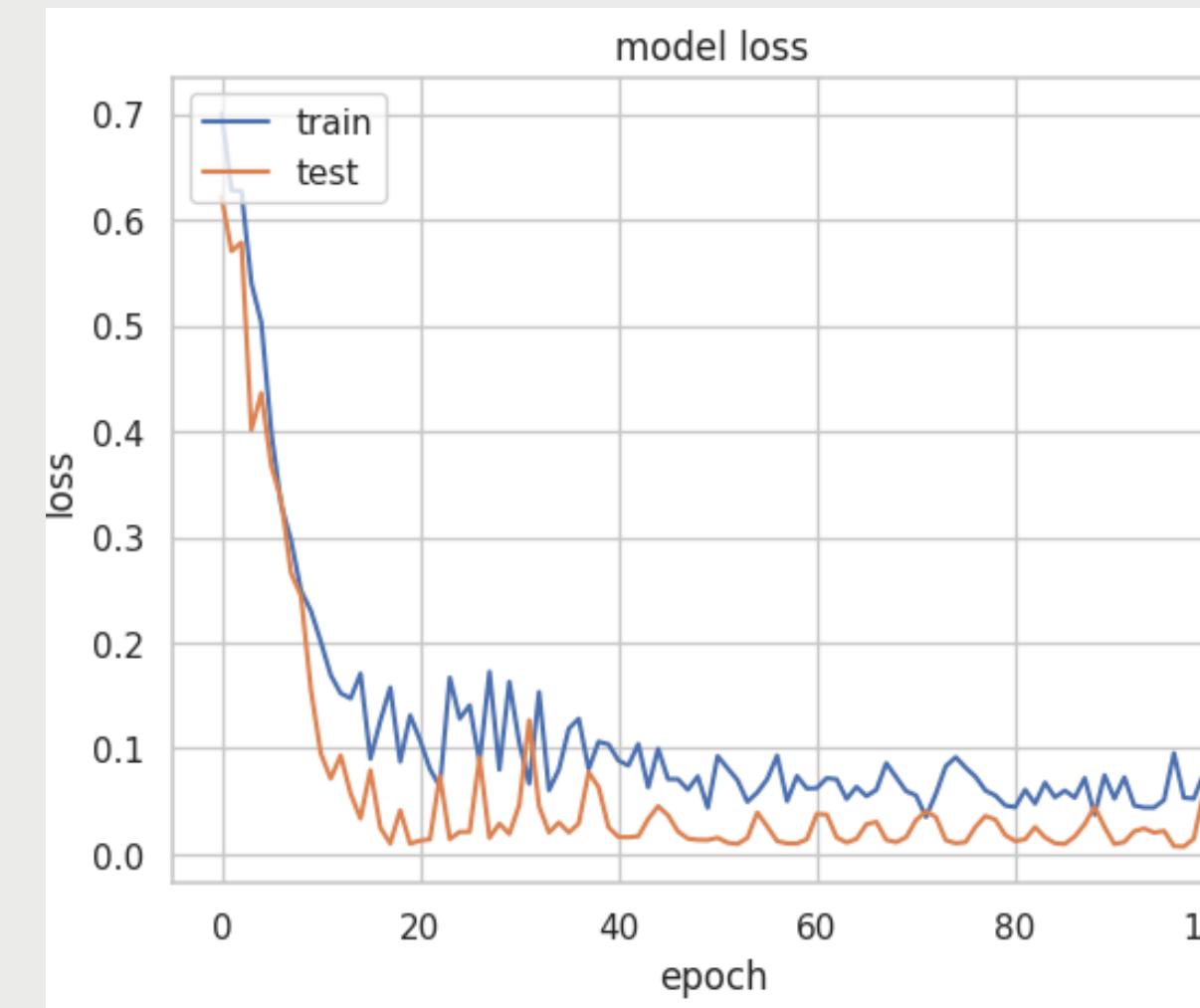
IMPORTANCE FINGERPRINT DESCRIPTOR



CONFUSION MATRIX ANN



EVALUASI MODEL



KESIMPULAN

Penelitian ini menunjukkan bahwa model Artificial Neural Network (ANN) dan XGBoost memiliki kemampuan yang kuat dalam memprediksi bioaktivitas senyawa terhadap protein mTOR. Dengan menggunakan dataset "label2class_MTOR.csv" dan representasi struktur kimia melalui berbagai metode fingerprint molekular seperti Avalon, Atom-Pair, dan Morgan-circular, model ANN berhasil memprediksi kelas bioaktivitas dengan akurasi yang baik. Sementara itu, model XGBoost menunjukkan kinerja unggul dengan akurasi 97,85%, Precision 99,88% untuk label "Active," dan Recall 95,23%. Kedua metode ini menunjukkan efektivitas dalam mengenali pola kompleks pada data molekular, sehingga berpotensi besar untuk dikembangkan lebih lanjut dalam aplikasi desain obat berbasis komputasi yang lebih akurat dan efisien.



**TERIMA
KASIH**

