

PENERAPAN K-MEANS DAN LGBM CLASSIFIER UNTUK IDENTIFIKASI SENYAWA BACE1 SEBAGAI INHIBITOR ALZHEIMER



KELOMPOK 2 BIOINFORMATIKA:

ERICSON CHANDRA (121450026)

KHARISMA GUMILANG (121450142)

DEDE MASITA (121450007)

EUNIKE BUNGA(121450095)

RAMADHITA ATIFA HENDRI(121450131)

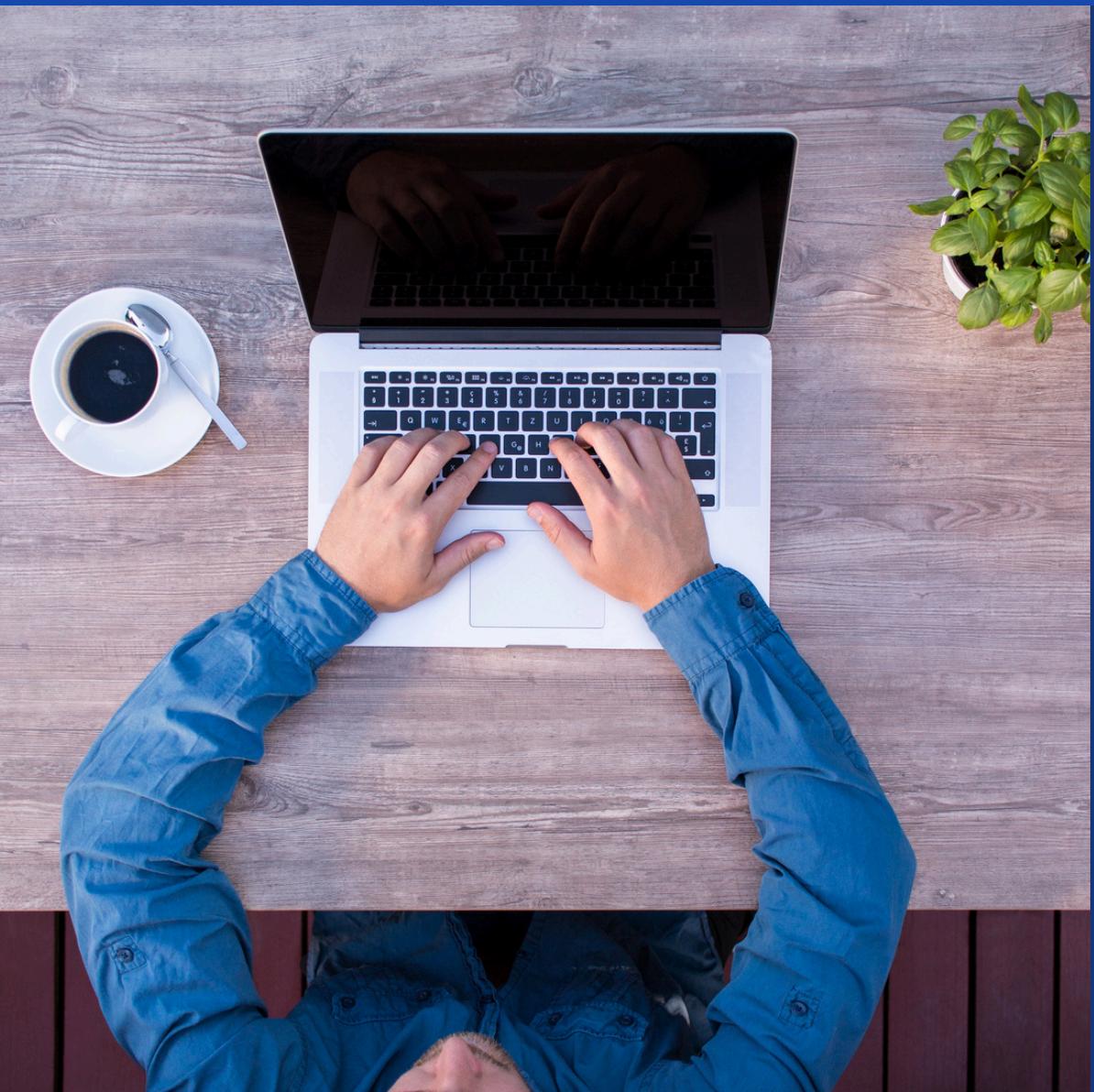
DIVANIA RAHMADANI(121450104)

PENERAPAN K-MEANS DAN LGBM CLASSIFIER UNTUK IDENTIFIKASI SENYAWA BACE1 SEBAGAI INHIBITOR ALZHEIMER

TABLE OF CONTENT

- Pendahuluan
- Tujuan
- Metodologi
- Hasil dan Pembahasan

- Kesimpulan
- Saran



PENERAPAN K-MEANS DAN LGBM CLASSIFIER UNTUK IDENTIFIKASI SENYAWA BACE1 SEBAGAI INHIBITOR ALZHEIMER

PENDAHULUAN

Penyakit Alzheimer: Penyakit neurodegeneratif yang menyebabkan penurunan fungsi kognitif.

Peran Protein BACE1:

- Enzim Beta-Secretase 1 (BACE1) berperan penting dalam pembentukan peptida beta-amiloid.
- Menjadi target utama untuk pengembangan inhibitor.

Pendekatan Studi:

- Menggunakan pembelajaran mesin untuk identifikasi inhibitor potensial.

TUJUAN

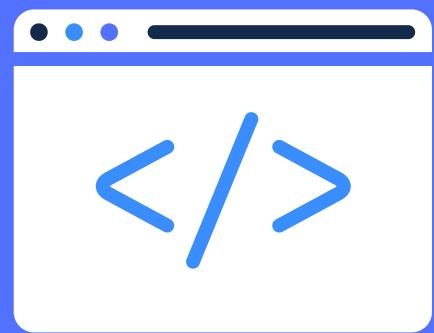
Mengidentifikasi senyawa aktif yang dapat bertindak sebagai inhibitor BACE1.

Menerapkan algoritma K-Means Clustering untuk pengelompokan senyawa.

Menggunakan Light Gradient Boosting Machine (LGBM) Classifier untuk klasifikasi bioaktivitas senyawa.

METODOLOGI

PENERAPAN K-MEANS DAN LGBM CLASSIFIER
UNTUK IDENTIFIKASI SENYAWA BACE1 SEBAGAI
INHIBITOR ALZHEIMER



Dataset

- Data dari database ChEMBL dengan target BACE1
- Data ini memiliki peran penting dalam memahami mekanisme penyakit Alzheimer
- 8.618 baris data dengan atribut utama: SMILES dan pIC50.

Tahapan Preprocessing

- Dataset akan difilter dan hanya menggunakan data dari spesies Homo sapiens.
- Ekstraksi deskriptor Lipinski.
- Normalisasi nilai IC50 menjadi pIC50.

Clustering K-Means

- Mengelompokkan senyawa berdasarkan karakteristik struktural.
- Menggunakan fingerprint Morgan sebagai representasi fitur untuk setiap molekul.

METODOLOGI

PENERAPAN K-MEANS DAN LGBM CLASSIFIER
UNTUK IDENTIFIKASI SENYAWA BACE1 SEBAGAI
INHIBITOR ALZHEIMER

Cross Validation

- menilai kinerja model dengan membagi dataset menjadi beberapa subset (folds).
- menggunakan 10 iterasi dan 30% data sebagai data uji pada setiap iterasi.

Klasifikasi LGBM

- metode gradient boosting yang cepat, terdistribusi, dan memiliki kinerja tinggi yang berbasis pada decision tree.
- Proses pelatihan dilakukan dengan menggunakan fingerprint avalon molekul sebagai fitur (X) dan pIC50 sebagai target (y)

Evaluasi Model

- Akurasi (Accuracy)
- F1-Score
- Area Under the Curve (AUC)
- Receiver Operating Characteristic (ROC)

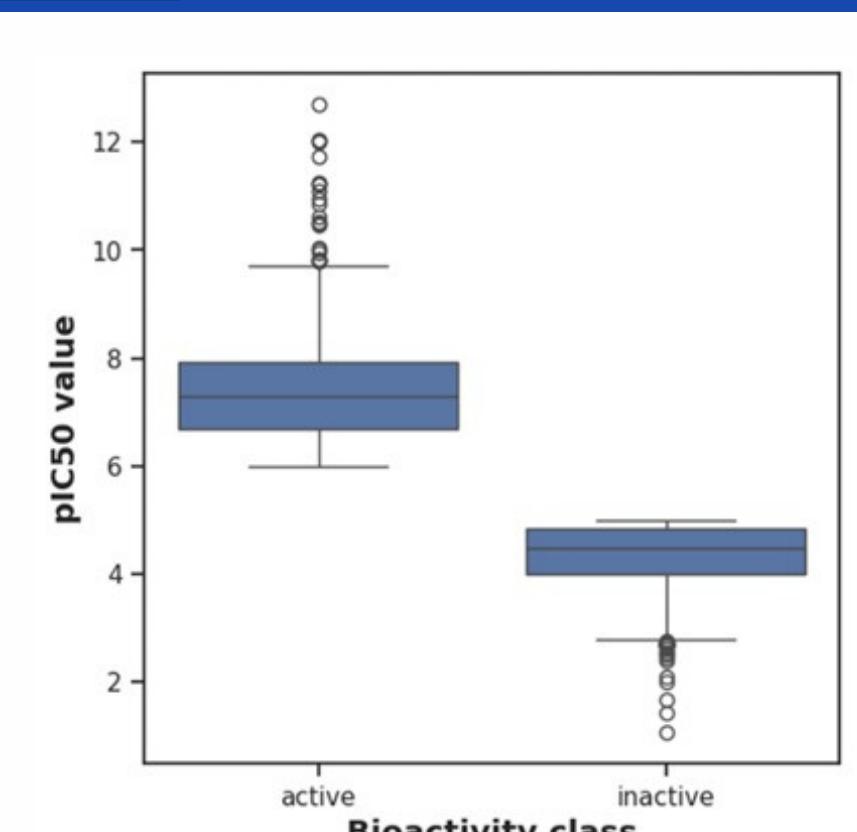
HASIL DAN PEMBAHASAN

Distribusi Data

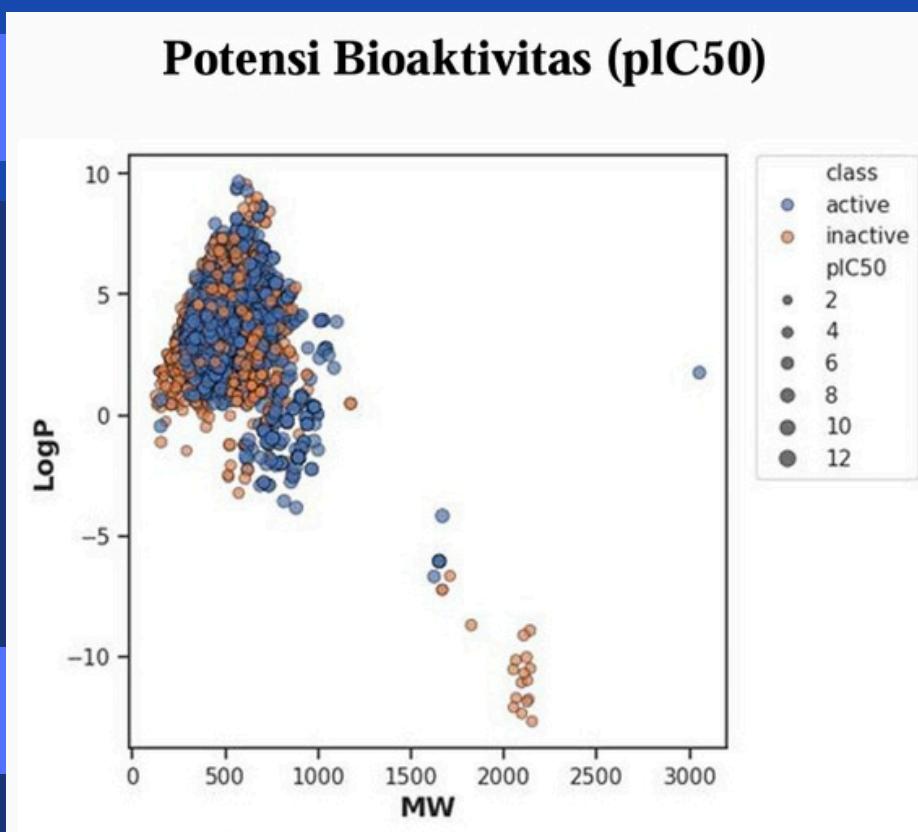
- Sebagian besar senyawa aktif memiliki MW 300-800 dan LogP 0-7.
- Senyawa dengan nilai MW > 1000 cenderung tidak aktif.

Perbedaan Kelas Bioaktivitas

Median pIC50 senyawa aktif lebih tinggi dibandingkan yang tidak aktif, menunjukkan potensi bioaktivitas lebih besar.

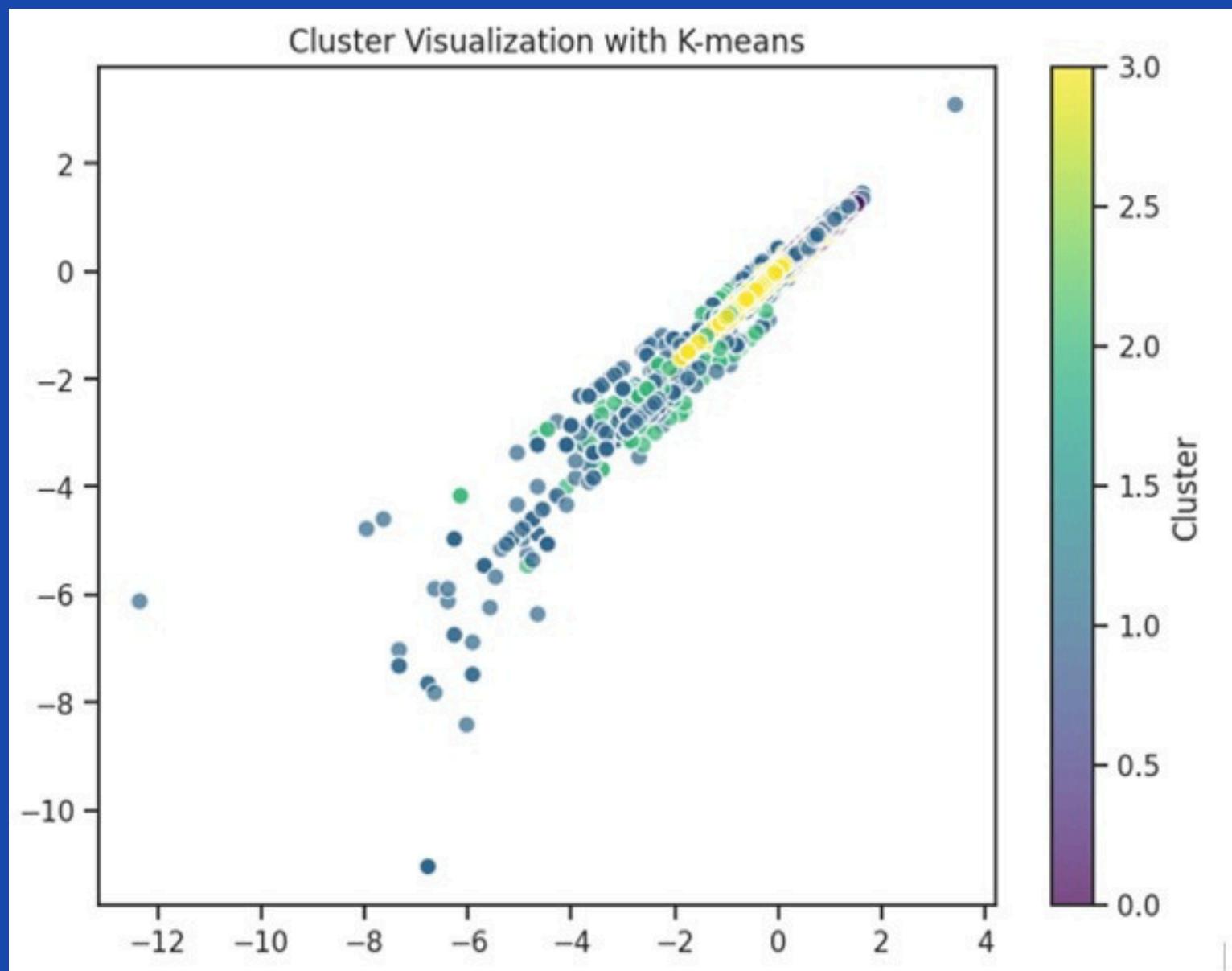


Gambar 3. Kelas Bioaktivitas



Gambar 2. Scatter Plot MW vs LogP

HASIL DAN PEMBAHASAN



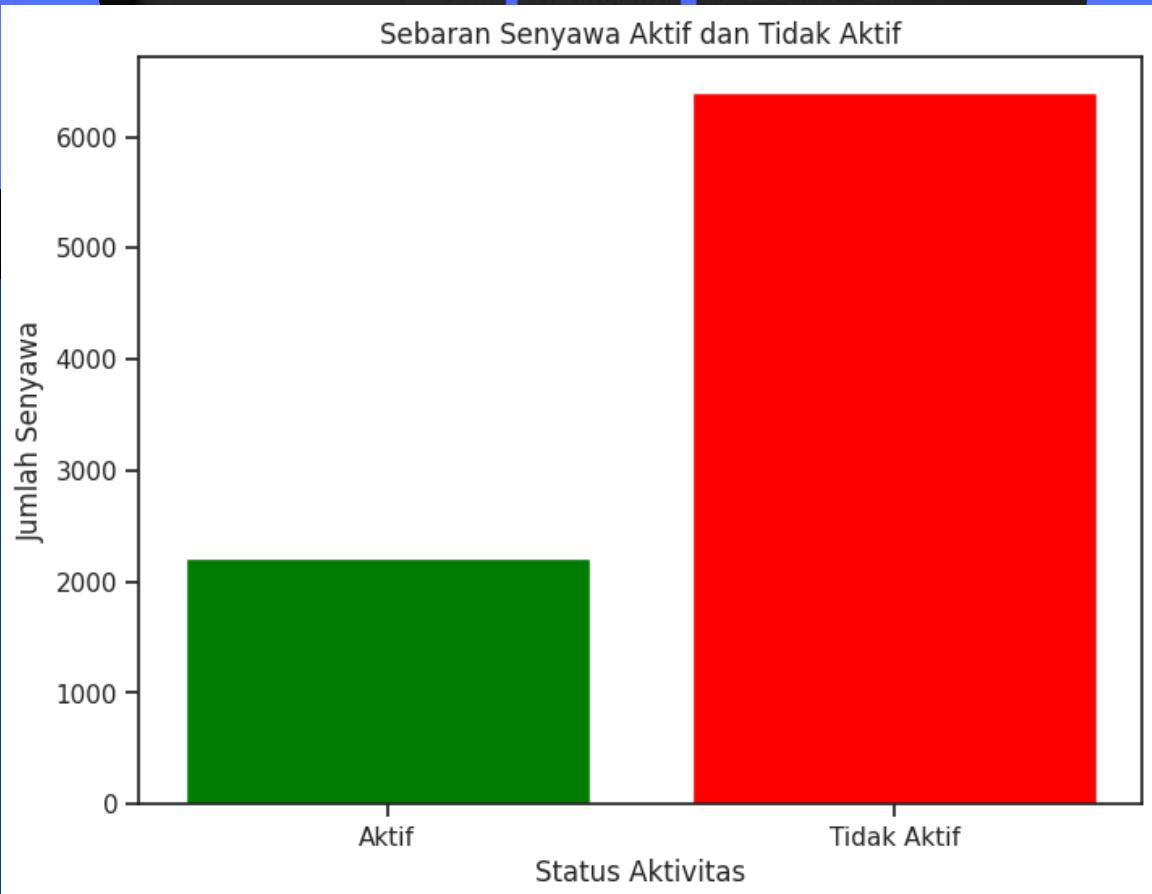
Hasil Klasterisasi

- Cluster 0: Senyawa dengan heteroatom seperti nitrogen dan klorin (3.558 senyawa).
- Cluster 1: Senyawa dengan gugus karboksil dan amida (2.494 senyawa).
- Cluster 2: Senyawa dengan hidroksi dan fenil (1.323 senyawa).
- Cluster 3: Struktur fenol dan alkohol (1.242 senyawa).

Hasil Klasifikasi

Jumlah Senyawa

- Senyawa aktif: 2.219.
- Senyawa tidak aktif: 6.398.



Metrik Evaluasi

- Akurasi (Train): 97%.
- Akurasi (Test): 93%.
- F1-Score: 93%.
- ROC AUC: 95%.

Kinerja Model

Model LGBM menunjukkan akurasi yang baik pada data uji.

KESIMPULAN

- K-Means Clustering berhasil mengelompokkan senyawa berdasarkan karakteristik struktural.
- LGBM Classifier mampu mengklasifikasi senyawa dengan akurasi tinggi (Accuracy (C-V) 93%, F1-Score (C-V), 93% ROC-AUC(C-V) 95%
- Pendekatan pembelajaran mesin mempercepat proses identifikasi inhibitor potensial.

SARAN

- Memperluas cakupan dataset dengan mengumpulkan data yang lebih beragam dan mencakup berbagai jenis senyawa bioaktif
- Penggunaan berbagai jenis fingerprint molekul yang lebih bervariasi, seperti Atom Pair, Topological Torsion, atau Morgan Circular, dapat dieksplorasi lebih lanjut untuk menggali hubungan struktur dan aktivitas yang mungkin tidak teridentifikasi menggunakan fingerprint Avalon saja.
- Pengembangan model dapat ditingkatkan dengan membandingkan performa berbagai algoritma pembelajaran mesin, seperti Random Forest, XGBoost, atau Neural Networks, serta mengoptimalkan hyperparameter untuk meningkatkan akurasi dan reliabilitas model

TERIMA KASIH

