

Prediksi Bioaktivitas Senyawa terhadap Protein MTOR Menggunakan Pendekatan *Artificial Neural Network* (ANN)

Prediction of Compound Bioactivity Against MTOR Protein Using an Artificial Neural Network (ANN) Approach

Anissa Luthfi Alifia¹, Meinisa², Angelica Noviana³, Nadia Fitri Yani⁴,
Sella Dianka Fitri⁵, Ibnu Farhan Al-Ghifari⁶

¹Jurusan Sains Data, Fakultas Sains, Institut Teknologi Sumatera, Lampung Selatan, Indonesia

²Institut Teknologi Sumatera, Lampung Selatan, Indonesia

*E-mail:

anissa.121450093@student.itera.ac.id

meinisa.121450076@student.itera.ac.id

angelica.121450064@student.itera.ac.id

nadia.121450101@student.itera.ac.id

sella.121450043@student.itera.ac.id

ibnu.121450121@student.itera.ac.id

Abstrak

Mammalian target of rapamycin (mTOR) adalah enzim serin/treonin kinase yang berperan penting dalam regulasi sel dan terkait dengan perkembangan kanker, sehingga menjadi target potensial terapi berbasis molekuler. Penelitian ini memprediksi bioaktivitas senyawa terhadap mTOR menggunakan *Artificial Neural Network* (ANN), *Random Forest Regressor* dan XGBoost. Dataset "label2class_MTOR.csv" digunakan dengan berbagai metode *fingerprint* molekular seperti Avalon dan Morgan-circular untuk representasi struktur kimia. Hasil menunjukkan XGBoost memiliki akurasi 97,85%, precision label "active" 99,88%, dan recall 95,23%. ANN dengan 100 epoch dan batch size 128 juga menghasilkan prediksi akurat untuk kelas *active*, *inactive*, dan *intermediate*.

Kata Kunci : *Artificial Neural Network*; Bioaktivitas; mTOR; *Random Forest Regressor*; XGBoost.

Abstract

Mammalian target of rapamycin (mTOR) is a serine/threonine kinase enzyme that plays an important role in cell regulation and is associated with cancer development, making it a potential target for molecular-based therapy. This research predicts the bioactivity of compounds against mTOR using *Artificial Neural Network* (ANN), *Random Forest Regressor* and XGBoost. The dataset "label2class_MTOR.csv" is used with various molecular fingerprinting methods such as Avalon and Morgan-circular for chemical structure representation. The results show that XGBoost has an accuracy of 97.85%, an "active" label precision of 99.88%, and a recall of 95.23%. ANN with 100 epochs and batch size 128 also produces accurate predictions for active, inactive and intermediate classes.

Keywords: *Artificial Neural Network*; Bioactivity; mTOR; *Random Forest Regressor*; XGBoost.

PENDAHULUAN

Mammalian target of rapamycin (mTOR) adalah serin/treonin kinase berukuran besar yang berperan penting dalam berbagai proses selular, seperti regulasi pertumbuhan dan metabolisme sel, perkembangan dan proliferasi fungsi imun adaptif, serta kematian sel melalui apoptosis dan autofagi [1]. mTOR mampu meningkatkan translasi *messenger-RNA* (mRNA) untuk gen yang terlibat dalam pengaturan proliferasi sel, ketahanan hidup, dan progresi siklus sel dari fase G1 menuju S [2-3]. Jalur sinyal pada mTOR berfungsi sebagai integrator yang mendukung interaksi selular dan organisme dengan lingkungan, jalur ini merupakan pusat kontrol pada perkembangan korteks serebri yang berhubungan dengan pertumbuhan sel abnormal [4]. Sinyal mTOR berperan pada seluruh fungsi otak sehingga disfungsi mTOR akan menyebabkan sejumlah kelainan monogenetik[5]. Kelainan ini menjadi dasar perkembangan berbagai jenis kanker, sehingga menjadikannya target utama dalam pengembangan terapi berbasis molekuler. Untuk menemukan inhibitor mTOR yang efektif, maka dibutuhkan metode prediktif yang mampu mengevaluasi bioaktivitas senyawa kimianya.

Pengembangan terapi berbasis molekuler dengan senyawa inhibitor mTOR dilakukan untuk menghambat aktivitas mTOR, sehingga dapat mengurangi pertumbuhan sel-sel kanker. Namun eksperimen biologis dalam evaluasi bioaktivitas senyawa membutuhkan sumber daya waktu yang lama dan sumber daya yang besar, sehingga dibutuhkan pendekatan komputasional untuk mempercepat prosesnya dengan biaya lebih sedikit.

Artificial Neural Network (ANN) menjadi pendekatan yang efektif dalam memprediksi bioaktivitas senyawa karena kemampuannya dalam menangkap pola non-linear antara deskriptor kimia dan aktivitas biologis. Selain ANN, Random Forest Regressor (RFR) juga umum digunakan pada prediksi bioaktivitas

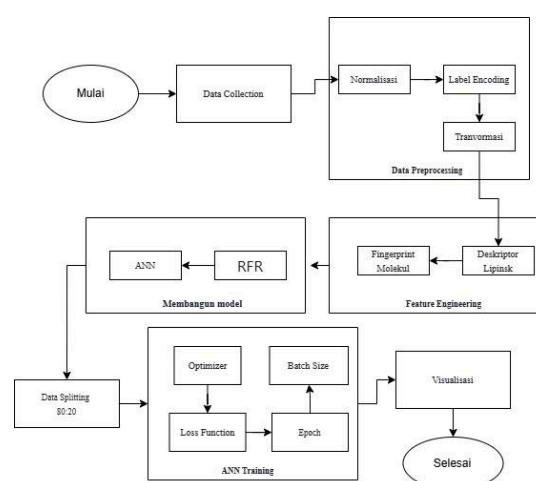
senyawa karena dapat memberikan interpretasi fitur input. Pada penelitian ini digunakan kedua pendekatan tersebut, yaitu ANN dan RFR untuk memprediksi bioaktivitas senyawa terhadap protein mTOR.

Penelitian terdahulu yang dapat digunakan sebagai referensi dalam penelitian ini adalah “*Molecular Docking* Senyawa (8)-Shogaol sebagai obat anti-kanker” oleh Cindy Nur Syaqla dkk yang dilakukan untuk memprediksi hubungan suatu atom dan molekul, lalu penelitian “Peran *Mammalian Target of Rapamycin* dalam Epilepsi” yang ditulis oleh Machlusil Husna dkk yang bertujuan untuk modulasi mTOR sebagai alternatif baru dalam terapi epilepsi.

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

1. Diagram alir

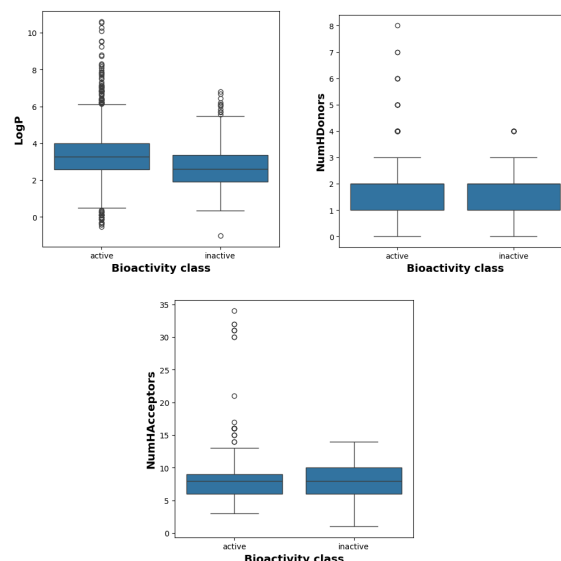
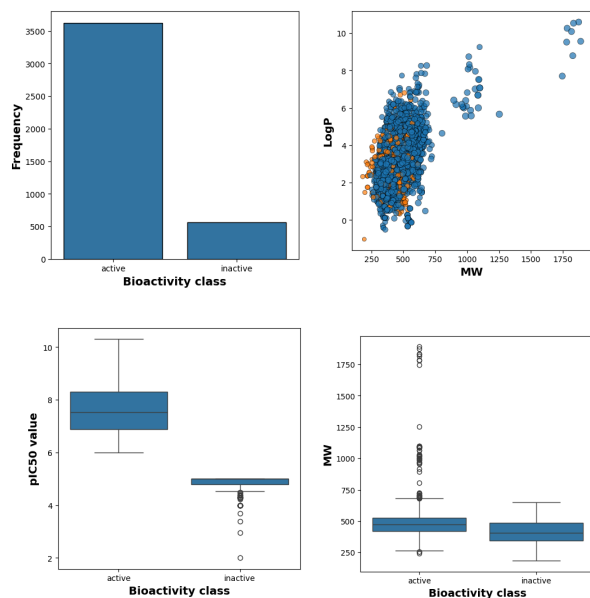


2. Deskripsi dataset

Dataset ini diperoleh dari Database ChEMBL, yang menyimpan data mengenai bioaktivitas senyawa kimia terhadap sasaran biologis. Data yang dimanfaatkan berfokus pada target MTOR (target mamalia dari rapamycin), yang merupakan protein kinase dengan peranan krusial dalam pengaturan pertumbuhan sel, proliferasi, dan metabolisme, dengan bioaktivitas diukur melalui nilai IC50, yaitu konsentrasi senyawa yang dapat menghambat 50% aktivitas dari target. Proses pengumpulan data diawali dengan mencari target menggunakan kode `target.search('mTOR')`, lalu data bioaktivitas disaring menggunakan fungsi `activity.filter` dengan kriteria `target_chembl_id` yang relevant serta nilai aktivitas IC50. Data dari beberapa subset kemudian digabungkan menjadi satu DataFrame, membentuk dataset akhir terdiri dari 4185 senyawa, target, nilai IC50, metode pengujian, dan referensi dari literatur, yang bisa digunakan untuk melakukan analisis hubungan struktur-aktivitas (SAR) atau untuk pengembangan model prediksi.

3. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah proses analisis awal data untuk memahami pola, hubungan, dan anomali melalui visualisasi dan statistik deskriptif [6]. Hasil EDA pada penelitian ini akan ditampilkan pada gambar di bawah ini:



Gambar 1. *Exploratory Data Analysis*

4. Model

4.1 Imbalanced Data

Imbalanced data terjadi ketika jumlah sampel dalam setiap kelas tidak seimbang, sehingga model cenderung memprediksi kelas mayoritas [7]. Kondisi ini dapat menyebabkan performa model tidak optimal dalam mengenali kelas minoritas.

4.2 SMOTE

SMOTE (*Synthetic Minority Oversampling Technique*) adalah metode untuk menangani data tidak seimbang dengan membuat sampel sintetis pada kelas minoritas [8]. Teknik ini bekerja dengan interpolasi antara sampel minoritas yang ada untuk menambah jumlah data. SMOTE membantu meningkatkan kemampuan model dalam mengenali pola pada kelas minoritas.

4.3 Artificial Neural Network

ANN (*Artificial Neural Network*) adalah model pembelajaran mesin yang terinspirasi oleh struktur otak manusia.

Model ini terdiri dari lapisan neuron yang saling terhubung untuk memproses data dan belajar dari pola yang kompleks [9]. ANN sering digunakan untuk tugas klasifikasi, regresi, dan prediksi dengan data berukuran besar.

4.4 Random Forest Regressor

Random Forest Regressor adalah algoritma pembelajaran mesin berbasis ensemble yang berfungsi untuk memprediksi nilai kontinu [10]. Algoritma ini membangun sejumlah *decision tree* dari berbagai subset data dan menggabungkan prediksinya untuk menghasilkan estimasi akhir yang lebih stabil dan akurat.

4.5 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 [11]. Sejak diperkenalkan, XGBoost telah berhasil memenangkan berbagai kompetisi machine learning dan menjadi solusi yang kuat, efisien, serta efektif untuk menyelesaikan berbagai permasalahan klasifikasi.

4.6 Pelatihan Model dan Evaluasi

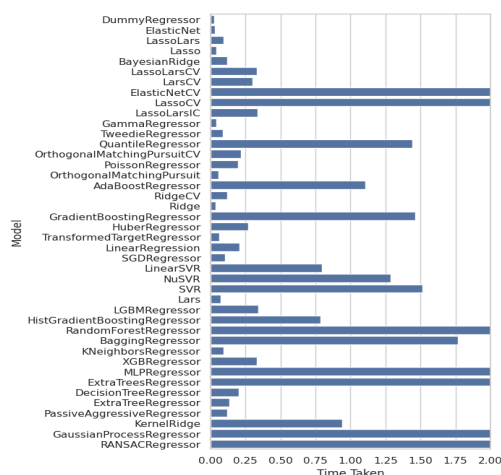
Untuk mengevaluasi kinerja model, metrik yang digunakan adalah *classification report* yang mencakup *precision*, *recall*, *F1-score*, dan *support*.

HASIL DAN PEMBAHASAN

Pada penelitian ini, dataset asli memiliki *missing value* pada beberapa kolom. Baris yang mengandung *missing value* pada kolom-kolom tersebut dihapus, dan duplikasi pada kolom - kolom tersebut dihilangkan untuk memastikan data bersih

dan unik. Setelah itu dilakukan pemrosesan data lebih lanjut seperti memberi label data, *export dataset*, melakukan transform *SMILE* menjadi *LIST*, menghitung lipinski deskriptor, dan mengkonversi data IC50 menjadi pIC50.

Langkah pertama yang dilakukan adalah membagi dataset menjadi data train dan data test dengan persentase 80:20%. Kemudian dilakukan pencarian kombinasi hyperparameter terbaik dengan menggunakan *Random Forest Regressor* dan *GridSearchCV*. Selanjutnya digunakan *LazyRegressor* untuk menguji dan membandingkan berbagai model tanpa menulis kode untuk setiap model secara manual. Hyperparameter yang digunakan ada *Adjusted R-Squared*, *R-Squared*, *RMSE*, dan *Time Taken*. Pada **Gambar 2** ditunjukkan 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*.

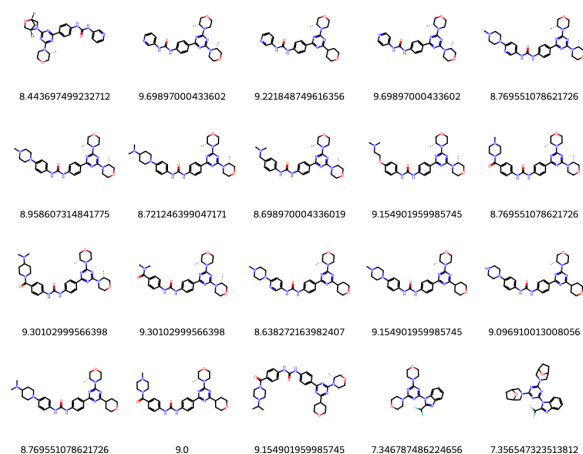


Gambar 2. Plot Calculation Time

1. Preprocessing dan Feature Engineering Machine Learning

Pada tahap ini, dataset yang digunakan adalah "label2class_MTOR.csv" dan kolom "canonical_smiles". Kemudian

ditambahkan kolom “Structure” 2D pada *dataframe*. Langkah selanjutnya, dilakukan visualisasi SD dalam bentuk *grid* seperti pada **Gambar 3**. Gambar ini adalah representasi visual dari struktur molekul berbagai senyawa kimia. Setiap gambarnya menggambarkan susunan atom - atom dalam sebuah molekul, di mana garis - garis pada struktur tersebut mewakili ikatan kimia yang menghubungkan atom - atom tersebut. Struktur dasar pada molekul ini adalah cincin benzena, yang merupakan struktur siklik khas senyawa aromatik. Variasi pada struktur molekul ini disebabkan oleh perbedaan jenis dan posisi atom atau gugus atom yang terikat pada cincin benzena.

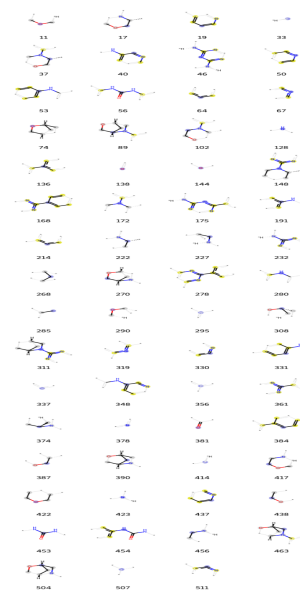


Gambar 3. Visualisasi Struktur Molekul dalam Grid

Proses selanjutnya yaitu melakukan *Generate Fingerprint*, dengan langkah pertamanya adalah *MACC-Keys* yang digunakan untuk menghasilkan *fingerprint* molekular atau representasi digital dari molekul kimia yang menyandikan informasi struktural. Jumlah bits yang digunakan adalah 167 dan ukuran bits MACC yang digunakan yaitu (167,). Proses selanjutnya adalah melakukan *Avalon*, *Atom-Pair*, *Topology-Torsi*, dan *Morgan-circular* untuk menghasilkan *fingerprint* molekular yang mewakili struktur kimia dalam format digital, dengan jumlah dan ukuran bitsnya adalah 512.

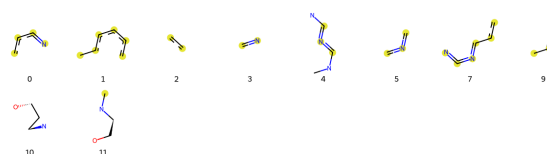
Pada **Gambar 4** ditunjukkan

visualisasi *fingerprint* untuk semua bits yang aktif dari proses tersebut. Struktur - struktur ini terdiri dari berbagai fragmen molekul kecil yang mengandung atom - atom seperti karbon (ditunjukkan dengan garis/vertex), nitrogen (biru), oksigen (merah), dan sulfur (kuning). Keberagaman struktur molekul ini mencakup berbagai gugus fungsi, cincin heterosiklik, dan pola ikatan yang membantu model dalam mempelajari hubungan struktur - aktivitas (SAR) yang penting dalam pengembangan kandidat obat baru.



Gambar 4. Visualisasi Semua Bits yang Aktif

Setelah itu dilakukan proses *RDkit* untuk analisis molekular, pemodelan prediktif, dan pengembangan aplikasi berbasis kimia. **Gambar 5** merupakan contoh visualisasi dari proses *RDkit* yang menunjukkan serangkaian struktur molekul sederhana yang diberi nomor dari 0-11, dengan fokus pada fragmen molekul yang mengandung atom nitrogen (biru), oksigen (merah), dan beberapa gugus yang ditandai dengan highlight kuning.

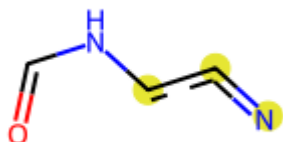


Gambar 5. Visualisasi Fragmen dari Proses RDkit

Dilakukan juga proses *Data Engineering* yang mencakup proses pemanggilan semua fungsi, pemanggilan semua output sebelumnya, juga visualisasi *Morgan* dan *RDkit on bit for single structure and single fingerprint* seperti pada **Gambar 6** dan **Gambar 7**.



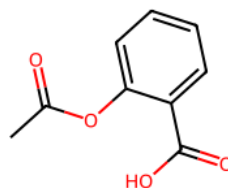
Gambar 6. Visualisasi Morgan on bit for single structure and single fingerprint



Gambar 7. Visualisasi RDkit on bit for single structure and single fingerprint

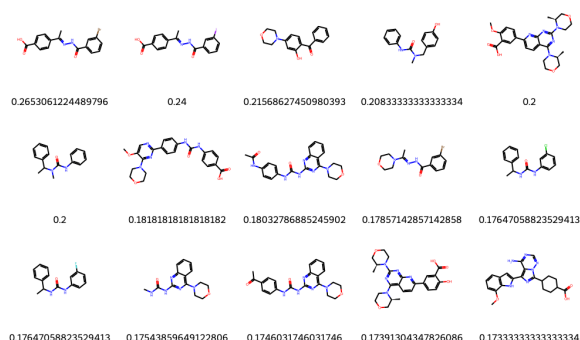
Proses yang selanjutnya dilakukan yaitu pencarian senyawa yang mirip dengan kueri atau senyawa tertentu yang ditunjukkan pada **Gambar 8**. Struktur molekul tersebut adalah Aspirin (Asam asetilsalisilat), yang merupakan obat umum untuk analgesik (penghilang rasa sakit), antipiretik (penurun demam), dan anti-inflamasi. Molekul ini terdiri dari cincin benzena yang tersubstitusi dengan dua gugus fungsional penting : gugus karboksilat (-COOH) dan gugus ester asetat (-OCOCH₃). Aspirin sering digunakan sebagai senyawa referensi atau standar karena mekanisme kerjanya yang telah dipahami dengan baik (menghambat enzim

siklooksigenase/COX) dan profil bioaktivitasnya telah terdokumentasi dengan baik.



Gambar 8. Struktur Molekul Aspirin

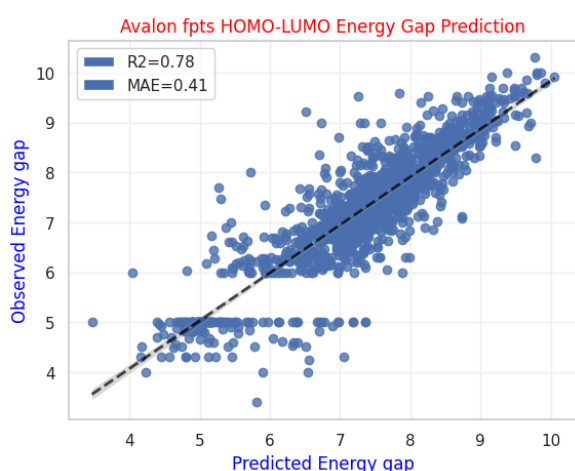
Setelah dilakukan pencarian senyawa atau kueri yang mirip, dilakukan juga visualisasi untuk struktur dan nilai koefisien Tanimoto. Koefisien Tanimoto adalah sebuah ukuran yang digunakan untuk mengukur kemiripan antara dua himpunan yang sering digunakan untuk membandingkan struktur molekul, seperti yang ditunjukkan pada **Gambar 9**. Gambar ini menunjukkan serangkaian struktur molekul organik kompleks beserta nilai numerik (0.17-0.26) yang merepresentasikan skor prediksi bioaktivitas dari model neural network. Struktur - struktur ini memiliki kesamaan motif yaitu mengandung cincin aromatik (benzena) yang terhubung dengan gugus penghubung yang mengandung nitrogen (biru) dan oksigen (merah), serta beberapa memiliki substituen halogen (hijau/cyan). Pola nilai prediksi yang ditampilkan menunjukkan variasi bioaktivitas yang relatif kecil antar struktur, yang mengindikasikan bahwa modifikasi struktur yang ada memberikan pengaruh subtil terhadap aktivitas yang diprediksi.



Gambar 9. Struktur dan Koefisien Nilai Tanimoto

Proses terakhir, dilakukan perhitungan untuk R^2 dan MAE dengan menggunakan *lightGBM* dan *train-test-split Avalon*. Nilai R^2 dan MAE untuk perhitungan dengan *lightGBM* menghasilkan nilai masing-masing 0.79 dan 0.42. Sedangkan untuk perhitungan dengan menggunakan *train-test-split Avalon*, nilai yang dihasilkan sebesar 0.78 dan 0.41. Kedua model ini menghasilkan nilai R^2 dan MAE yang tidak jauh berbeda.

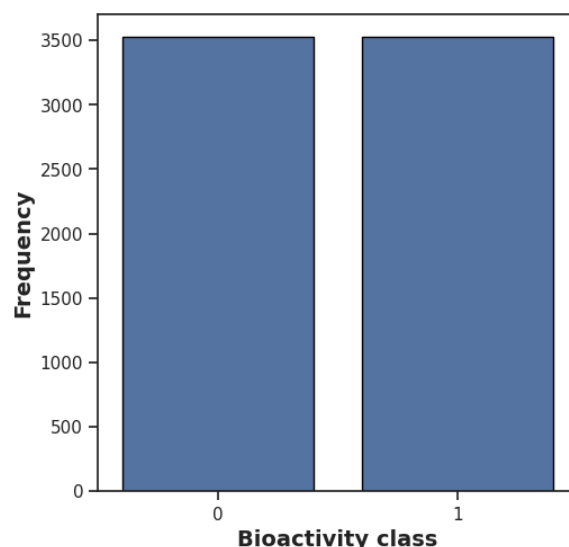
Pada **Gambar 10** ditunjukkan prediksi *energy gap* dan observasi *energy gap*. Grafik ini menunjukkan prediksi celah energi HOMO-LUMO menggunakan *fingerprint Avalon*, dengan sumbu x menunjukkan nilai prediksi dan sumbu y menunjukkan nilai observasi. Garis putus-putus diagonal hitam merepresentasikan prediksi yang ideal dimana nilai prediksi sama dengan nilai observasi. Sebaran titik-titik biru yang relatif mengikuti garis diagonal mengindikasikan performa model yang cukup baik, meskipun terdapat beberapa titik yang menyimpang terutama di area energi gap 4 - 6. Namun, secara keseluruhan model menunjukkan kemampuan prediksi yang baik untuk rentang energi gap antara 4-10 eV.



Gambar 10. Visualisasi Prediksi Celah Energi HOMO-LUMO

2. Pipeline Pengembangan Model Machine Learning

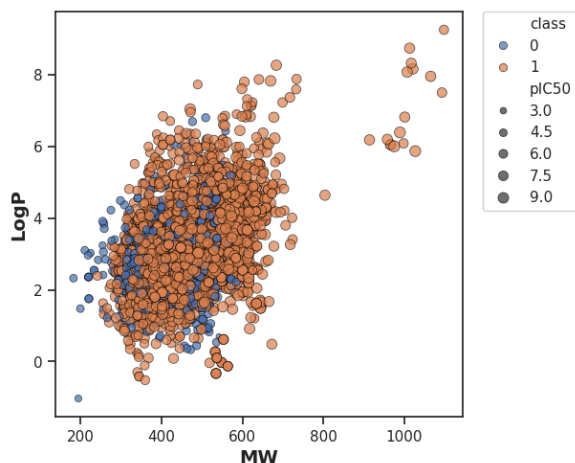
Pada tahap ini, terutama dalam proses *Data Engineering*, digunakan 3 dataset. Dataset 1 (*lipinsky*) dilakukan transformasi dari *class* menjadi *numeric ordinal*, dataset 2 (*finger print*), dan dataset 3 (*combined*). Kemudian dilakukan proses “SMOTE” untuk mengatasi data yang tidak seimbang. **Gambar 11** menunjukkan visualisasi *class* dengan data yang sudah dilakukan proses SMOTE.



Gambar 11. Visualisasi Data SMOTE

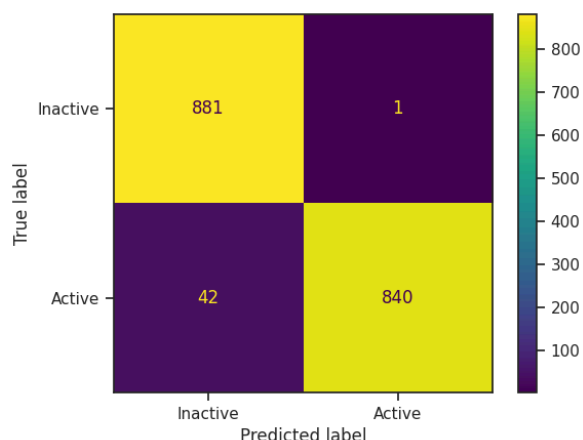
Selanjutnya, dilakukan visualisasi untuk LogP dan MW untuk data hasil proses SMOTE yang ditunjukkan di **Gambar 2**. Grafik ini menunjukkan hubungan antara berat molekul (MW) pada sumbu x dan LogP pada sumbu y, dengan data dikelompokkan menjadi dua kelas (0 dan 1) yang ditunjukkan dengan warna biru dan oranye. Ukuran titik pada grafik merepresentasikan nilai pIC50 yang berkisar antara 3.0 hingga 9.0. Sebaran data menunjukkan bahwa mayoritas senyawa memiliki berat molekul antara 200-800 Da dan nilai LogP antara 0-8, dengan konsentrasi tertinggi pada MW 400-600 dan LogP 2-6. Kelas 1 (oranye) tampak lebih dominan dalam dataset, menunjukkan distribusi yang tidak seimbang antara kedua kelas. Pola sebaran ini sangat relevan untuk pengembangan obat karena berada dalam rentang "aturan Lima Lipinski" yang

umumnya digunakan sebagai panduan dalam desain obat, dimana $MW < 500$ dan $LogP < 5$ dianggap optimal untuk bioavailabilitas oral yang baik.



Gambar 12. Visualisasi LogP dan MW untuk SMOTE

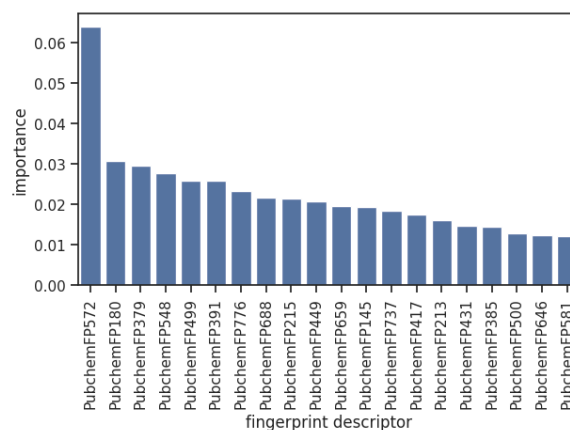
Proses selanjutnya ada pengolahan data dengan *XGBoost* yang ditunjukkan dengan visualisasi di **Gambar 13**. Confusion matrix tersebut menunjukkan bahwa model *XGBoost* memiliki kinerja yang sangat baik dengan akurasi sekitar 97,85%. Model berhasil memprediksi sebagian besar label dengan benar, dengan *Precision* untuk label "Active" mencapai 99,88% dan *Recall* 95,23%. Namun, terdapat 42 *False Negatives*, di mana model melewati beberapa instance "Active," sementara hanya ada 1 *False Positive*.



Gambar 13. Confusion Matrix XGBoost

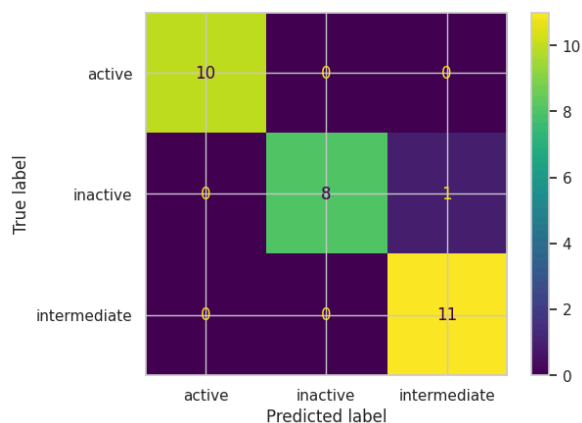
Kemudian, dibuat juga visualisasi

untuk tingkat kepentingan berbagai *fingerprint descriptor* dalam model yang digunakan, ditunjukkan pada **Gambar 14**. Visualisasi ini menunjukkan *fingerprint descriptors* paling penting dalam model adalah PubchemFP572. Descriptor penting seperti PubchemFP180 dan PubchemFP379, menunjukkan fitur kimia yang signifikan terhadap prediksi.



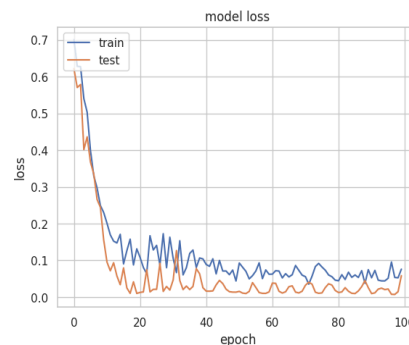
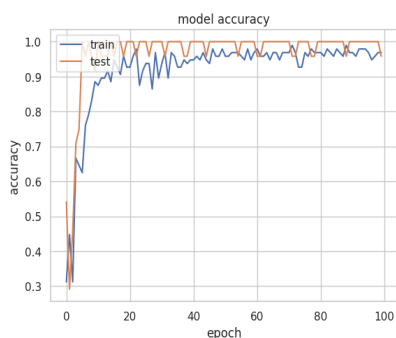
Gambar 14. Visualisasi Tingkat Kepentingan Fingerprint Descriptor

Proses terakhir adalah *Artificial Neural Network* (ANN), dengan epoch yang digunakan adalah 100, *batch size* 128, dan data validasinya sebesar 20%. Dari *hyperparameter* tersebut, didapatkan visualisasi *Confusion Matrix* seperti pada **Gambar 15**. *Confusion matrix* ini menunjukkan kinerja model *Artificial Neural Network* (ANN) untuk tiga kelas : **active**, **inactive**, dan **intermediate**. Model memprediksi sebagian besar kelas dengan benar, seperti terlihat pada diagonal utama (10 untuk active, 8 untuk inactive, dan 11 untuk intermediate). Namun, ada beberapa kesalahan, seperti 1 instance "inactive" yang salah diprediksi sebagai "intermediate."



Gambar 15. Confusion Matrix ANN

Hasil evaluasi model yang didapatkan dari hasil penelitian ini sangat baik seperti pada **Gambar 16**. Grafik ini menunjukkan perubahan loss pada model selama proses pelatihan (*train loss*) dan pengujian (*test loss*) seiring bertambahnya jumlah epoch. Pada awal pelatihan, baik *train loss* maupun *test loss* tinggi, tetapi menurun tajam selama beberapa epoch pertama, menunjukkan bahwa model sedang belajar dengan efektif. Setelah sekitar 20 epoch, *train loss* dan *test loss* stabil pada nilai yang rendah, menunjukkan bahwa model berhasil meminimalkan error dan tidak mengalami overfitting yang signifikan. Perbedaan kecil antara *train loss* dan *test loss* di akhir pelatihan menunjukkan bahwa model memiliki generalisasi yang baik.



Gambar 16. Akurasi dan Loss Model

KESIMPULAN

Penelitian ini menunjukkan bahwa model Artificial Neural Network (ANN) telah berhasil digunakan untuk memprediksi bioaktivitas senyawa terhadap protein mTOR. Dengan memanfaatkan dataset "label2class_MTOR.csv," berbagai metode pembuatan fingerprint molekular seperti Avalon, Atom-Pair, Topology-Torsi, dan Morgan-circular telah diterapkan untuk menghasilkan representasi struktur kimia secara digital. Pendekatan ini memungkinkan ekstraksi fitur penting, termasuk PubchemFP572, yang memberikan kontribusi signifikan terhadap prediksi.

Proses pengolahan data menggunakan model XGBoost menunjukkan kinerja yang sangat baik dengan akurasi 97,85%, di mana Precision untuk label "Active" mencapai 99,88% dan Recall sebesar 95,23%. Selanjutnya, implementasi ANN dengan 100 epoch dan batch size 128 berhasil memprediksi kelas active, inactive, dan intermediate dengan tingkat keakuratan yang baik, meskipun terdapat beberapa kesalahan dalam klasifikasi.

Kesimpulannya, metode ANN dan XGBoost memiliki potensi besar dalam memprediksi bioaktivitas senyawa terhadap protein mTOR, dengan performa yang unggul dalam mengenali pola kompleks pada data molekular. Penelitian ini membuka peluang untuk pengembangan model prediksi yang lebih akurat dan aplikasi lebih lanjut

dalam desain obat berbasis komputasi.

SARAN

Berdasarkan hasil penelitian ini, disarankan untuk melakukan pengembangan lebih lanjut pada model *Artificial Neural Network* (ANN), terutama dengan mengeksplorasi arsitektur yang lebih kompleks atau melakukan penyesuaian hyperparameter agar mengurangi kesalahan klasifikasi pada beberapa kelas. Selain itu, pengujian model dengan dataset yang lebih besar dan beragam juga perlu dilakukan untuk meningkatkan kemampuan generalisasi model dalam memprediksi bioaktivitas senyawa terhadap protein mTOR. Penggunaan teknik validasi silang yang lebih mendalam juga direkomendasikan untuk memastikan performa model yang lebih stabil dan andal. Langkah-langkah ini diharapkan dapat meningkatkan keakuratan dan ketepatan prediksi dalam aplikasi desain obat berbasis komputasi.

UCAPAN TERIMA KASIH

Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada Bapak Tirta Setiawan, S.Pd., M.Si., dosen pengampu Bioinformatika, Arahkan dan bimbingan beliau telah memberikan kontribusi luar biasa dalam penelitian ini, melalui panduan dan pemahaman mendalam di bidang bioinformatika.

DAFTAR RUJUKAN

1. Husna M, Kurniawan SN. Peran mammalian target of rapamycin dalam epilepsi. *Neurona*. 2017;34(4):244–251.
2. Russo E, Citraro R, Constanti A, De Sarro G. The mTOR signaling pathway in the brain: focus on epilepsy and epileptogenesis. *Mol Neurobiol*. 2012;46(3):662–81.
3. Martinova EA. mTOR kinase and its role in the cell stress response. *Biochem Suppl Ser A Membr Cell Biol*. 2012;6(1):9–15.
4. Curatolo P, Moavero R. mTOR inhibitors as a new therapeutic option for epilepsy. *Expert Rev* 2013;13(6):1–13.
5. Lasarge CL, Danzer SC. Mechanisms regulating neuronal excitability and seizure development following mTOR pathway hyperactivation. *Front Mol Neurosci*. 2014;7:18.
6. Radhi M, Amalia, Sitompul DRH, Sinurat SH, Indra E. Analisis big data dengan metode exploratory data analysis (EDA) dan metode visualisasi menggunakan Jupyter Notebook. *JUSIKOM Prima*. 2021;4(2):23–27.
7. Mutmainah S. Penanganan imbalance data pada klasifikasi kemungkinan penyakit stroke. *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*. 2021;1(1):10–16.
8. Kasnah AN, Muldai, Pujianto U. Penerapan teknik SMOTE untuk mengatasi imbalance class dalam klasifikasi objektivitas berita online menggunakan algoritma KNN. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*. 2019;3(2):196–201.
9. Kurniawansyah AS. Implementasi metode artificial neural network dalam memprediksi hasil ujian kompetensi kebidanan (studi kasus: Akademi Kebidanan Dehasen Bengkulu). *Jurnal Pseudocode*. 2018;5(1):37–44.
10. Lestari ES, Astuti I. Penerapan random forest regression untuk memprediksi harga jual rumah dan cosine similarity untuk rekomendasi rumah pada provinsi Jawa Barat. *Jurnal Ilmiah FIFO*. 2022;14(2):131–146.
11. Rizky PS, Hirzi RH, Hidayaturrohman U. Perbandingan metode LightGBM dan XGBoost dalam menangani data dengan kelas tidak seimbang. *J Statistika*. 2022;15(2):228–236