

Klasifikasi Bioaktivitas Pada Protein PARP dan Estimasi Nilai pIC50 Pada protein PARP Pada Kasus Kanker Payudara Menggunakan Support Vector Machine (SVM)

Classification of Bioactivity in PARP Protein and Estimation of pIC50 Values in PARP in Cancer Cases Using Support Vector Machine (SVM)

Putri Intan Kirani¹*, Muhammad Zaki Abdillah², Hartiti Fadlillah³, Muhammad Kaisar Firdaus⁴, Alber Analafean⁵, Yogi Sa'e Tama⁶.

Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera, Lampung Selatan, Indonesia.

**E-mail: putri.121450055@student.itera.ac.id*

Abstrak

Penelitian ini bertujuan untuk menganalisis bioaktivitas protein PARP dalam konteks kasus kanker menggunakan algoritma Support Vector Machine (SVM). Data yang digunakan berasal dari ChEMBL dan diproses melalui virtual screening dengan Python. Studi ini melibatkan dua pendekatan utama: pertama, klasifikasi bioaktivitas PARP menjadi kategori aktif (0) dan inaktif (1) dengan membandingkan performa kernel SVM. Kernel RBF menunjukkan akurasi terbaik sebesar 0,954, diikuti oleh polinomial (0,952), linear (0,888), dan sigmoid (0,499). Model terbaik dipilih menggunakan GridSearchCV. Kedua, estimasi nilai pIC50 pada bioaktivitas PARP, menghasilkan nilai MSE sebesar 0,47, MAE sebesar 0,53, dan RMSE sebesar 0,69. Hasil penelitian menunjukkan potensi SVM sebagai metode andal dalam analisis bioinformatika, khususnya dalam klasifikasi bioaktivitas dan estimasi pIC50.

Kata kunci: Protein PARP; Kanker; Bioaktivitas; Support Vector Machine; pIC50; Kernel SVM

Abstract

This research aims to analyze the bioactivity of PARP protein in cancer cases using the Support Vector Machine (SVM) algorithm. The data utilized is sourced from ChEMBL and processed through virtual screening with Python. The study involves two primary approaches: first, the classification of PARP bioactivity into active (0) and inactive (1) categories by comparing the performance of SVM kernels. The RBF kernel achieved the highest accuracy of 0.954, followed by polynomial (0.952), linear (0.888), and sigmoid (0.499). The best model was selected using GridSearchCV. Second, the estimation of pIC50 values for PARP bioactivity, yielding MSE of 0.47, MAE of 0.53, and RMSE of 0.69. The results demonstrate the potential of SVM as a reliable method in bioinformatics analysis, particularly for bioactivity classification and pIC50 estimation.

Keywords: PARP Protein; Cancer; Bioactivity; Support Vector Machine; pIC50; SVM Kernel

PENDAHULUAN

Poly ADP-Ribose Polymerase (PARP) adalah enzim penting yang terlibat dalam mekanisme perbaikan DNA, khususnya dalam merespons kerusakan DNA yang terjadi akibat berbagai faktor, termasuk radiasi dan stres oksidatif. Enzim ini berfungsi untuk menambah rantai poli(ADP-ribosa) pada protein target, yang kemudian berperan dalam proses perbaikan dan pemulihan sel [1]. Salah satu aspek menarik dari PARP adalah perannya dalam kanker, terutama pada jenis kanker payudara

dan ovarium yang memiliki mutasi BRCA. Mutasi pada gen BRCA1 dan BRCA2 mengurangi kemampuan sel untuk memperbaiki kerusakan DNA, dan penghambatan PARP dapat memperburuk kerusakan DNA pada sel kanker, menyebabkan kematian sel secara selektif [2]. Konsep ini dikenal sebagai letalitas sintesis, yang menjadikan inhibitor PARP sebagai target utama dalam terapi kanker [3].

Pentingnya pengembangan inhibitor PARP dalam pengobatan kanker membuka peluang untuk penerapan teknologi komputasi dalam identifikasi dan pengujian

calon senyawa inhibitor. Salah satu metode yang efektif untuk menemukan kandidat inhibitor adalah melalui pemrograman dan analisis data molekuler menggunakan teknik-teknik machine learning. Dalam era kemajuan teknologi komputasi, pendekatan bioinformatika modern menjadi semakin penting dalam mempercepat proses penemuan dan pengembangan obat. Metode komputasi dapat secara signifikan mengurangi waktu dan biaya yang diperlukan dalam tahap awal pengembangan obat dengan menyaring kandidat potensial sebelum dilakukan pengujian eksperimental [4].

Dalam penelitian ini, dilakukan pendekatan pemrograman untuk menyaring senyawa yang dapat menghambat aktivitas PARP melalui pendekatan bioinformatika komprehensif yang menggabungkan teknik fingerprinting molekuler, pembelajaran mesin, dan analisis hubungan kuantitatif struktur-aktivitas (QSAR) untuk memprediksi aktivitas inhibitor PARP. Proses ini melibatkan pemrosesan data molekuler yang mencakup fitur kimiawi dan biologis dari senyawa, serta klasifikasi status aktivitas senyawa berdasarkan data yang telah diproses. Penelitian ini memanfaatkan data dari ChEMBL sebagai sumber data IC50 untuk inhibitor PARP, yang telah terbukti sebagai sumber data yang handal untuk pengembangan obat berbasis komputasi [5]. Database ChEMBL menyediakan sumber data bioaktivitas yang komprehensif untuk molekul-molekul yang berpotensi menjadi obat [6]. Data yang diekstrak mencakup nilai IC50 sebagai ukuran aktivitas penghambatan dan berbagai deskriptor molekuler termasuk parameter Lipinski.

Penelitian ini mengimplementasikan pendekatan pembelajaran mesin yang komprehensif, meliputi regresi Support Vector Machine (SVM) untuk prediksi pIC50 dan kategorisasi bioaktivitas senyawa. SVM dipilih karena kemampuannya dalam menangani data non-linear dan performa yang baik dalam prediksi aktivitas biologis [7].

Dalam proses analisis, senyawa

diklasifikasikan menjadi kategori aktif dan tidak aktif berdasarkan ambang batas IC50, dengan nilai ≤ 1000 nM dianggap aktif dan ≥ 10000 nM dianggap tidak aktif [9]. Pendekatan ini mengikuti standar yang telah ditetapkan dalam literatur untuk klasifikasi aktivitas senyawa. Deskriptor Lipinski dihitung untuk menilai sifat drug-likeness, yang mencakup parameter fisikokimia kritis seperti berat molekul (MW), LogP, donor ikatan hidrogen, dan akseptor ikatan hidrogen. Dataset mengalami preprocessing yang ketat, termasuk penghapusan nilai yang hilang, struktur duplikat, dan standardisasi representasi molekuler [10].

Melalui pengolahan data yang terstruktur, penelitian ini bertujuan untuk membangun model prediksi yang mampu memisahkan senyawa aktif, tidak aktif, dan intermediari dalam menghambat aktivitas PARP, yang pada gilirannya dapat mempercepat penemuan kandidat senyawa terapeutik yang efektif [11]. Signifikansi penelitian ini terletak pada integrasi berbagai teknik pembelajaran mesin modern dengan analisis struktur molekuler yang komprehensif untuk mengembangkan model prediktif yang handal dalam mengidentifikasi inhibitor PARP potensial. Pendekatan ini diharapkan dapat mempercepat proses penemuan obat dan mengoptimalkan sifat-sifat kandidat obat, serta memberikan wawasan baru tentang hubungan struktur-aktivitas dalam inhibisi PARP.

METODE

Penelitian ini menggunakan pendekatan bioinformatika. Metodologi penelitian meliputi pengumpulan data, preprocessing, ekstraksi fitur molekuler, dan pengembangan model prediktif menggunakan kombinasi Support Vector Machine (SVM) untuk regresi pIC50 dan klasifikasi bioaktivitas senyawa.

1. Pengumpulan Data dan Seleksi Target

Data yang digunakan dalam

penelitian ini adalah data kimia molekuler berupa data aktivitas enzim Poly ADP-Ribose Polymerase (PARP) yang digabungkan dengan dataset fingerprint diambil dari website ChEMBL. Dataset dapat dilihat pada **Tabel 1** bagian lampiran.

2. Lipinski Rule

Lipinski rule merupakan sebuah aturan atau pedoman dalam penilaian kemungkinan suatu senyawa memiliki kemampuan atau sifat yang dapat dijadikan sebagai kandidat obat atas suatu penyakit. Lipinski rule sendiri memiliki empat kriteria yang dijadikan acuan dalam penilaian apakah suatu senyawa dapat dijadikan sebagai kandidat obat atau tidak, empat kriteria tersebut antara lain : Berat Molekul (Molecular Weight), Koefisien partisi (LogP), Jumlah donor ikatan hidrogen, dan Jumlah akseptor ikatan hidrogen.

Pada **Tabel 1** terdapat kumpulan kandidat senyawa kimia yang diprediksi dapat menjadi kandidat obat. Senyawa dikatakan menjadi kandidat obat yang cocok bila memenuhi empat kriteria Lipinski rule yaitu berat molekul harus kurang dari 500 dalton, Koefisien partisi harus kurang dari atau sama dengan 5, jumlah donor ikatan hidrogen harus kurang dari sama dengan 5, dan jumlah akseptor ikatan hidrogen harus kurang dari atau sama dengan 10. Pada Tabel 1 senyawa yang memiliki keempat kriteria ini di masukan sebagai kelas aktif, sementara senyawa yang tidak memenuhi kriteria Lipinski rule dimasukan kedalam kelas inaktif.

3. Fingerprint molekul

Fingerprint Molekul merupakan metode untuk merepresentasikan struktur molekul dalam suatu senyawa kimia ke dalam format vektor biner. Fingerprint molekul digunakan untuk menangkap informasi dalam suatu senyawa seperti jumlah ikatan, jumlah atom, dan sifat topologis dari suatu senyawa kimia yang dapat digunakan dalam pencarian kandidat

obat dan prediksi model senyawa.

	0	1	2	3	4	5	6	7	8	9	...	126	127	128	129	130	131	132	133	134	135
0	1	1	1	0	0	1	1	1	0	0	...	0	0	0	1	0	1	0	1	0	1
1	1	1	1	0	0	1	1	1	0	0	...	0	1	0	1	0	1	0	1	0	1
2	1	1	1	0	0	1	1	1	1	0	...	0	0	0	1	0	1	0	1	0	1
3	1	1	1	0	1	1	1	1	1	0	...	0	0	1	1	0	1	1	1	0	1
4	1	1	1	0	0	1	1	1	0	0	...	0	0	0	1	0	1	0	1	0	1
...
2632	1	1	1	0	0	1	1	1	1	0	...	0	1	1	0	1	0	1	0	1	0
2633	1	0	1	0	0	1	1	1	1	0	...	0	1	1	0	1	0	1	0	1	0
2634	1	0	1	0	1	1	1	1	1	0	...	0	1	1	0	1	0	1	0	1	0
2635	1	0	1	0	1	1	1	1	1	0	...	0	1	1	0	1	0	1	0	1	0
2636	1	0	1	0	0	1	1	1	0	0	...	0	1	1	0	1	0	1	0	1	0

Gambar 2. Fingerprint molekul pada senyawa kimia PARP

Dari hasil fingerprint molekul pada Gambar 2 terdapat baris-baris yang memiliki nilai 0 dan 1, dimana 1 menunjukkan keberadaan dan 0 menunjukkan ketidakhadiran fitur tertentu dalam molekul senyawa kimia.

4. Alat dan bahan

Penelitian dilaksanakan menggunakan Python 3.7 sebagai bahasa pemrograman utama dengan dukungan beberapa library spesifik untuk melakukan analisis data. Beberapa paket yang digunakan antara lain cairosvg, padelpy, gdown, luwiji, jcopml, dan chembl_webresource_client. Untuk visualisasi data, digunakan Matplotlib (versi 3.4.3) dan Seaborn (versi 0.11.2).

ChEMBL Web Resource Client digunakan untuk pengambilan dan pengelolaan data biologis, sementara RDKit 2020.09.1 diimplementasikan untuk kalkulasi deskriptor molekul dan manipulasi struktur kimia. Model SVM diimplementasikan menggunakan Scikit-learn 0.20.3, sedangkan manipulasi dan analisis data dilakukan dengan Pandas dan NumPy. Visualisasi data menggunakan kombinasi Matplotlib dan Seaborn untuk menghasilkan representasi grafis yang informatif. PaDEL-Descriptor digunakan sebagai perangkat lunak tambahan untuk generasi deskriptor molekul, khususnya untuk mendukung analisis prediksi pIC50. Seluruh komputasi dilakukan pada server dengan spesifikasi minimal RAM

8GB dan prosesor Intel Core i5. Dataset utama diperoleh dari database ChEMBL yang menyediakan data aktivitas biologis (IC50) senyawa terhadap target PARP.

Dataset yang digunakan dalam penelitian ini berasal dari Database ChEMBL yang terdiri dari tiga sumber utama, yaitu Dataset 1 (Lipinsky), yang berisi informasi tentang molekul dan kelas aktivitasnya, Dataset 2 (Fingerprint), yang mengandung informasi vektor jari-jari molekul (fingerprint), dan Dataset 3 (Combined), yang merupakan gabungan dari Dataset 1 dan Dataset 2 setelah dilakukan pre-processing.

5. Prosedur kerja

Tahap awal penelitian dimulai dengan pengambilan data aktivitas biologis inhibitor PARP dari database ChEMBL menggunakan Query API. Dataset yang diperoleh kemudian melalui tahap preprocessing yang komprehensif, meliputi penghapusan data duplikat, standarisasi format SMILES, dan penanganan missing values. Pada model regresi SVM, nilai IC50 dikonversi menjadi pIC50 untuk memberikan distribusi yang lebih baik dan interpretasi yang lebih mudah.

Ekstraksi fitur molekuler dilakukan dengan menghitung berbagai deskriptor menggunakan RDKit, termasuk deskriptor Lipinski yang mencakup berat molekul, LogP, serta jumlah donor dan akseptor ikatan hidrogen. Fingerprint molekuler yang digunakan meliputi Avalon (4096 bit) yang telah terbukti efektif dalam representasi struktur molekul untuk prediksi aktivitas biologis. Deskriptor ini dipilih berdasarkan kemampuannya dalam menangkap informasi struktural yang relevan untuk prediksi pIC50 dan klasifikasi bioaktivitas.

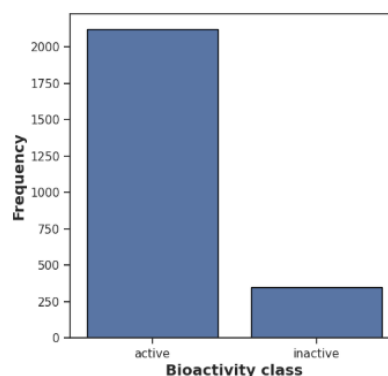
Model prediktif dikembangkan menggunakan dua pendekatan, SVM Regressor untuk prediksi pIC50 dan klasifikasi bioaktivitas. Dataset dibagi menjadi set training (80%) dan testing (20%) menggunakan metode SMOTE untuk menyeimbangkan kedua kelas. Model SVM

menggunakan kernel RBF dengan parameter C dan gamma yang dioptimasi melalui grid search dan 3-fold cross validation untuk memastikan generalisasi yang baik. Evaluasi performa model regresi SVM dilakukan menggunakan metrik R^2 , MAE, dan RMSE, sementara model klasifikasi Analisis statistik menggunakan uji Mann-Whitney U dilakukan untuk mengevaluasi signifikansi perbedaan distribusi deskriptor molekuler antara senyawa aktif dan tidak aktif. Hasil analisis divisualisasikan menggunakan scatter plot untuk hubungan antara nilai prediksi dan eksperimental pIC50, serta confusion matrix untuk

HASIL DAN PEMBAHASAN

1. Visualisasi Data

Dari pengolahan dataset yang ada didapatkan perbandingan antara bioaktivitas yang aktif dengan yang tidak aktif, dimana bioaktivitas dengan kelas aktif memiliki jumlah lebih banyak ketimbang bioaktivitas dengan kelas inaktif. Hal ini menunjukkan perkiraan bahwa jenis-jenis senyawa kimia PARP memiliki tingkat kesesuaian yang cukup tinggi untuk dijadikan sebagai kandidat obat penyakit kanker.

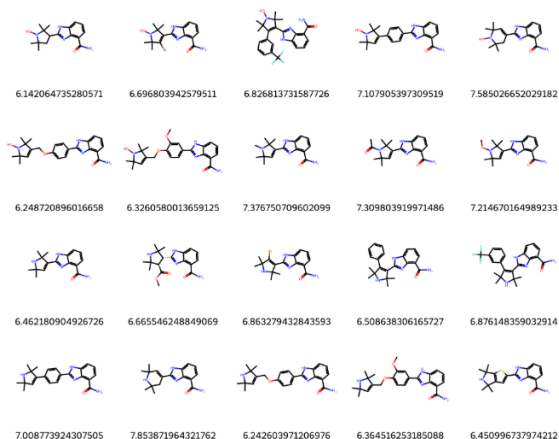


Gambar 3. Perbandingan Bioactive class

Struktur molekul yang didapatkan berdasarkan data senyawa kimia PARP divisualisasikan dalam bentuk 2D yang disimpan dalam format SMILES dimana setiap struktur molekul memiliki nilai pIC50. Perlu diingat bahwa visualisasi molekul-molekul senyawa yang ditampilkan

telah memenuhi setiap kriteria dari lipinski rule.

Visualisasi struktur molekul ini bertujuan agar ini dapat dilihat hubungan antar unsur penyusun struktur molekul, serta mengidentifikasi molekul potensial yang sekiranya memiliki tingkat bioaktivitas yang tinggi sehingga meningkatkan kemungkinan menjadi kandidat obat. Visualisasi molekul adalah sebagai berikut :



Gambar 3. Struktur Molekul

2. Hasil Klasifikasi & Prediksi

Dilakukan proses klasifikasi bioaktivitas protein PARP pada Kanker payudara menggunakan Support Vector Machine. Sebelum model dibangun dilakukan perbandingan dengan beberapa kernel dalam Support vector machine agar diperoleh model terbaik yang bisa terlihat pada tabel 2.

Tabel 2. Perbandingan Akurasi Model

	Linear	RBF	Polinomial	Sigmoid
Akurasi	0.888	0.954	0.952	0.499

Dari tabel 2, terlihat model Support Vector Machine dengan menggunakan kernel Radial Basis Function (RBF) memiliki akurasi sangat tinggi dibandingkan kernel.

Untuk memperkuat hal tersebut dilakukan hyperparameter tuning menggunakan gridsearch. Diperoleh bahwa model terbaik dalam melakukan klasifikasi bioaktivitas protein PARP adalah dengan parameter yang digunakan 'classifier c' bernilai 10, 'classifier_degree' bernilai 2, 'classifier_gamm' menggunakan skala dan 'kernel' yang digunakan adalah RBF dengan akurasi yang dimiliki sebesar 0.95.

Selanjutnya dilakukan pemodelan Support Vector Machine untuk memprediksi nilai pIC50 protein PARP. Prediksi yang dilakukan menggunakan fitur yaitu 'class', 'MW', 'LogP', 'numHDonors', dan 'NumHAcceptors' dengan target yaitu 'pIC50'. Pada fitur 'class' dilakukan proses encoding dan standarisasi terlebih dahulu yang selanjutnya dilakukan pemodelan SVM menggunakan kernel RBF. Terlihat hasil prediksi pIC50 memiliki nilai dominan pada 7, menunjukkan model Support Vector Machine memprediksi pIC50 memiliki nilai sama persis seperti data actual. Nilai pIC50 dominan 7, yang menunjukkan nilai pIC50 senyawa tersebut efektif sebagai obat,

3. Evaluasi Model

Pemodelan telah dilakukan, dilakukan evaluasi model dalam melihat seberapa baik model dalam melakukan Klasifikasi ataupun prediksi. Pada model support vector machine yang dibangun dalam melakukan klasifikasi bioaktivitas protein PARP pada Kanker menunjukkan evaluasi pada Tabel 3.

Tabel 3. Classification report

	Precisi on	recal l	F1-score	Supp ort
0	0.98	0.96	0.97	459
1	0.96	0.98	0.97	452
Accuracy			0.97	911

Macro avg	0.97	0.97	0.97	911
Weight avg	0.97	0.97	0.97	911

gambar Classification Report SVM

Pada Tabel 3 terlihat bahwa model melakukan klasifikasi dengan secara sangat baik, yang bisa terlihat pada nilai akurasi yang dimiliki sebesar 0.97 beserta nilai precision, recall dan f-1 score yang memiliki nilai mendekati 1 yang menunjukkan model sangat baik dalam mengklasifikasikan bioaktivitas. Dari akurasi tersebut, bisa terlihat dari confusion matrix pada gambar x.



Gambar 4. Confusion Matrix

Terlihat dari gambar 4, bahwa model hanya memiliki klasifikasi pada masing-masing kelas yang tidak terlalu banyak, sehingga menyebabkan model memiliki akurasi yang tinggi.

Evaluasi model juga dilakukan pada prediksi nilai pIC50 protein PARP, yang bisa terlihat pada tabel 4. menunjukkan bahwa model yang dibangun memiliki nilai error yang baik dalam melakukan prediksi nilai pIC50.

Tabel 4. Nilai Evaluasi model

Nilai Error	
MSE	0.47
MAE	0.53
RMSE	0.69

KESIMPULAN

Dari proses pemodelan yang dilakukan menggunakan Support vector Machine menunjukkan data bioaktivitas protein PARP memiliki perbedaan signifikan antara kelas aktif dan tidak aktif. Sehingga agar kedua kelas seimbang dilakukan teknik SMOTE.

Dari penelitian ini juga diperoleh visualisasi struktur molekul senyawa PARP yang memenuhi ketentuan Lipinski Rule. Dari model yang dibangun peneliti, menunjukkan akurasi yang sangat tinggi dalam melakukan klasifikasi bioaktivitas protein PARP pada kanker. Dengan menggunakan hyperparameter tuning, model klasifikasi bioaktivitas protein PARP diperoleh akurasi sebesar 0.954. Dan juga telah mampu memprediksi nilai pIC50 dengan akurasi tinggi. Yang mana nilai prediksi memiliki nilai dominan 7, yang menunjukkan nilai pIC50 senyawa tersebut efektif sebagai obat.

PARP merupakan salah satu suatu senyawa berpotensi dalam sebagai obat Kanker bahkan terkhusus Kanker Payudara ataupun Kanker Ovarium. Penggunaan model Support Vector Machine juga sangat efektif dalam membantu identifikasi molekul penyusun obat kanker

SARAN

Peneliti memiliki saran ke peneliti selanjutnya, model yang dibangun peneliti juga bisa berpotensi dalam mengembangkan model yang bisa membangun obat yang optimal dan terbaik dengan perlu adanya

kombinasi unsur senyawa lain yang efektif, kolaborasi penelitian serta penelitian lebih lanjut.

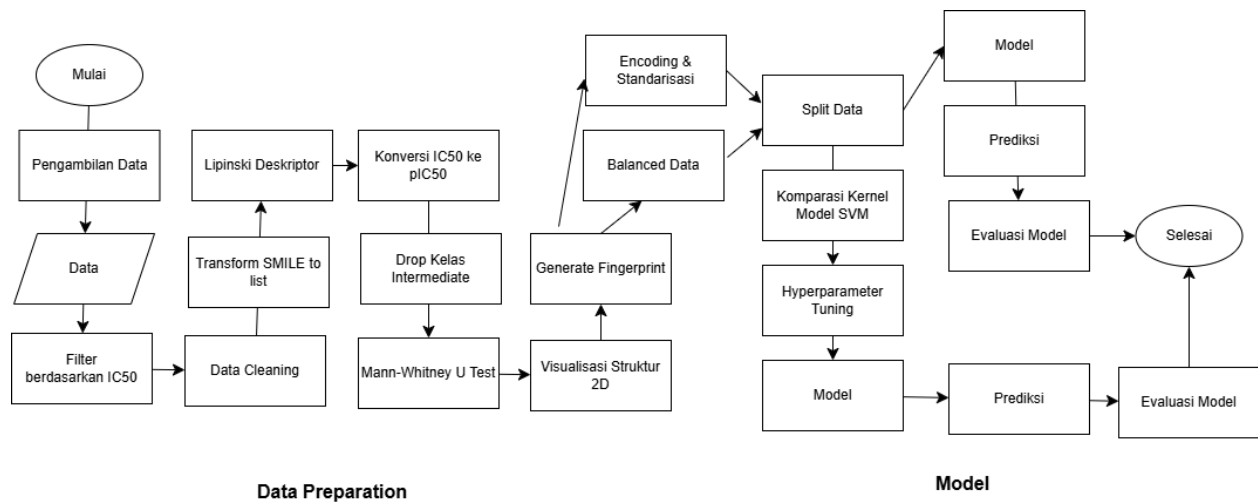
UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada Bapak Tirta Setiawan., S.Pd., M.Si. selaku dosen pengampu mata kuliah Bioinformatika serta kepada seluruh rekan penulis yang telah bekerja sama dalam menyelesaikan Tugas akhir Bioinformatika ini.

DAFTAR RUJUKAN

- [1] M. A. P. Maharani and P. E. Wuyung, "Peran Poly (ADP-ribose) Polymerase-1 pada Perbaikan Kerusakan DNA," *J. Patol. Anat.* vol. 5, no. 1, pp. 1-8, Jan. 2016.
- [2] J. C. Morales, L. Li, F. J. Fattah, Y. Dong, E. A. Bey, M. Patel, J. Gao, and D. A. Boothman, "Review of Poly (ADP-ribose) Polymerase (PARP) Mechanisms of Action and Rationale for Targeting in Cancer and Other Diseases," *Crit. Rev. Eukaryot. Gene Expr.*, vol. 24, no. 1, pp. 15-28, 2014, doi:10.1080/1040841X.2014.993419.
- [3] I. W. G. Suarsana, A. Islam, P. Prihantono, and B. J. Nelwan, "Peran Poly-(ADP ribose) polymerase (PARP) dan Phosphatidylinositol 3-kinase (PI3K) Terhadap Terjadinya Kejadian Metastasis pada Kanker Payudara," **JBN (Jurnal Bedah Nasional)**, vol. 6, no. 1, pp. 30, Jan. 2022, doi: 10.24843/JBN.2022.v06.i01.p05.
- [4] G. L. Wilson and M. A. Lill, "Integrating structure-based and ligand-based approaches for computational drug design," **Future Medicinal Chemistry**, vol. 3, no. 6, pp. 735–750, Apr. 2011, doi: 10.4155/fmc.11.18.
- [5] D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. De Veij, E. Félix, M. P. Magariños, J. F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Marañón, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, F. Atkinson, N. Bosc, C. J. Radoux, A. Segura-Cabrera, A. Hersey, and A. R. Leach, "ChEMBL: towards direct deposition of bioassay data," **Nucleic Acids Research**, vol. 47, no. D1, pp. D930-D940, Jan. 2019, doi: 10.1093/nar/gky107.
- [7] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, and J. P. Overington, "ChEMBL: a large-scale bioactivity database for drug discovery," **Nucleic Acids Research**, vol. 40, no. Database issue, pp. D1100-D1107, Jan. 2012, doi: 10.1093/nar/gkr777.
- [8] R. Rodríguez-Pérez, M. Vogt, and J. Bajorath, "Support Vector Machine Classification and Regression Prioritize Different Structural Features for Binary Compound Activity and Potency Value Prediction," *ACS Omega*, vol. 2, no. 10, pp. 6371–6379, Oct. 2017, doi: 10.1021/acsomega.7b01079.
- [9] K. Ngibad and L. P. Lestari, "Aktivitas Antioksidan dan Kandungan Fenolik Total Daun Zodia (*Evodia suaveolens*)," *ALCHEMY Jurnal Penelitian Kimia*, vol. 16, no. 1, pp. 94–109, Mar. 2020, doi: 10.20961/alchemy.16.1.35580.94-109.
- [10] Tensiska, B. Nurhadi, E. Wulandari, and Y. A. L. Ratri, "Aktivitas Antioksidan Ekstrak Dedak Hanjeli (*Coix lachryma-jobi L.*) dengan Beberapa Jenis Pelarut," *Jurnal Agroindustri*, vol. 10, no. 1, pp. 1–11, May 2020, doi: 10.31186/j.agroind.10.1.1-11.
- [11] J. Garcia and P. Houghton, "Advances in PARP inhibition as cancer therapy," **Cancer Therapy Reviews**, vol. 77, pp. 55-67, 2019, doi: 10.1016/j.ctrv.2019.03.006.

LAMPIRAN GAMBAR



Gambar 1. Diagram Alir Penelitian

LAMPIRAN TABEL

Tabel 1. Dataset

molecule_chembl_id	class	canonical_smiles	MW	LogP	NumH Donors	NumH Acceptors	pIC50
CHEMBL469961	active	CC1(C)CC(c2nc3c(C(N)=O)ccc3[nH]2)C(C)(C)N1O	302.378	239.750	3	4	6.142.065
CHEMBL470166	active	CC1(C)(Br)=C(c2nc3c(C(N)=O)cccc3[nH]2)C(C)(C)N1O	379.258	302.980	3	4	6.696.804
CHEMBL513759	active	CC1(C)C(c2ccc(C(F)(F)F)c2)=C(c2nc3c(C(N)=O)cccc3[nH]2)C(C)(C)N1O	444.457	485.340	3	4	6.826.814
CHEMBL459362	active	CC1(C)C=C(c2ccc(c3nc4c(C(N)=O)cccc4[nH]3)cc2)C(C)(C)N1O	376.460	397.420	3	4	7.107.905
CHEMBL518018	active	CC1(C)C(c2nc3c(C(N)=O)cccc3[nH]2)C(C)(C)N1O	314.389	269.730	3	4	7.585.027

Tugas Besar Bioinformatika RB Kelompok 04

...
CHEMBL36 98974	active	<chem>NC(=O)c1cccc2c1C(=O)N(CCCN1CCN(c3ccc(Cl)c3)CC1)C2</chem>	412.921	260.700	1	4	6.397.940
CHEMBL36 05984	active	<chem>Cc1ccc(CN2CC(C(CN3C(c4ccc(C(N)=O)c4C3=O)CC2)cc1</chem>	377.488	296.192	1	3	7.096.910
CHEMBL36 99003	active	<chem>NC(=O)c1cccc2c1C(=O)N(CC1CCN(Cc3ccc(F)cc3)CC1)C2</chem>	381.451	279.260	1	4	6.795.880
CHEMBL36 99004	active	<chem>NC(=O)c1cc(F)cc2c1C(=O)N(C1CCN(C3CCC(=O)CC3)CC1)C2</chem>	373.428	185.650	1	4	7.096.910
CHEMBL36 05995	active	<chem>COCCN1CCC(N2Cc3cccc(C(N)=O)c3C2=O)C1</chem>	317.389	0.8521	1	4	6.387.216
