

# KLASIFIKASI TIPE ASSAY (PENGUJIAN IKATAN SENYAWA) PADA MOLEKUL NLRP3 MENGGUNAKAN MODEL REGRESI LOGISTIK

fingerprint\_onehot

Kelompok 8



Yosia Letare Banurea | Sasa Rahma Lia | Muhammad Bagas Kurnia | M. Fahrul Aditya | Nadia Silvani | Asrizal

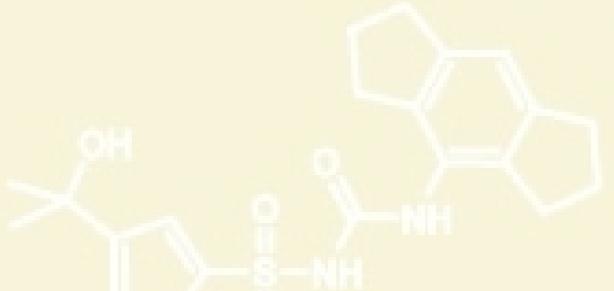
NLRP3

Pyrin

NACHT

LLRR

Walker B: RILFLMDGFDELQGAFDEHI

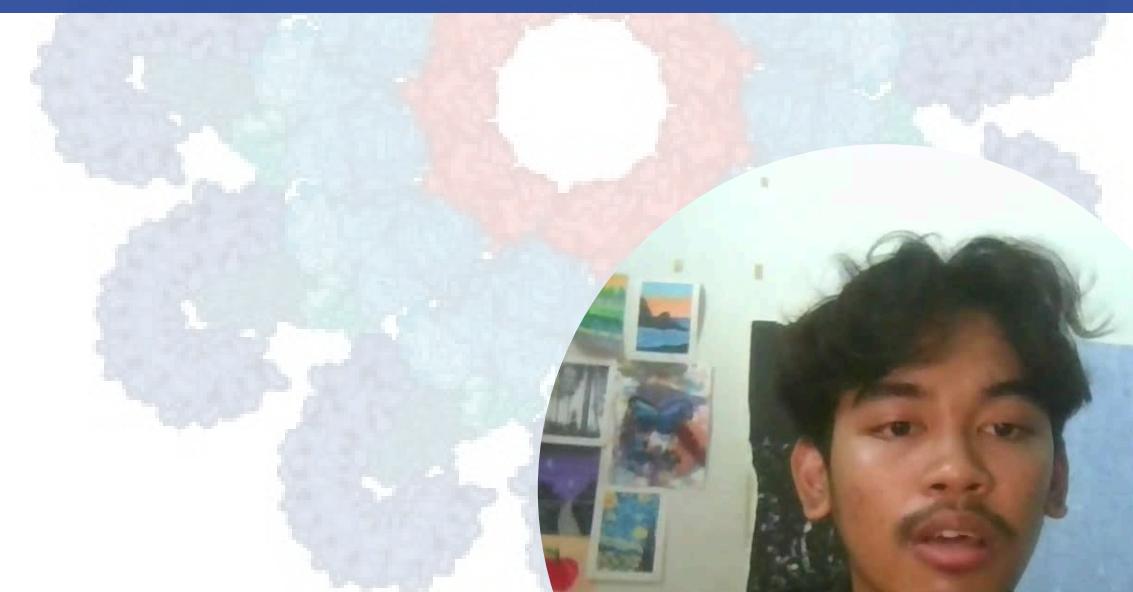


# PENDAHULUAN

Walker B

Inactive NLRP3

Active NLRP3



# PENDAHULUAN

NLRP3 adalah protein inflamasi yang berperan penting dalam respons imun dan sering dikaitkan dengan penyakit kronis, seperti diabetes tipe 2 dan autoimun. Pengujian ikatan senyawa (assay) menjadi langkah penting untuk mengidentifikasi molekul pengatur aktivitas NLRP3. Namun, variasi tipe assay, seperti pengukuran pengikatan senyawa (kelas B) dan efek biologis senyawa (kelas F), membutuhkan klasifikasi yang akurat. Penelitian ini menggunakan regresi logistik untuk mengklasifikasikan tipe assay berdasarkan data bioaktivitas dari database ChEMBL, guna mendukung analisis interaksi molekuler yang lebih efektif.



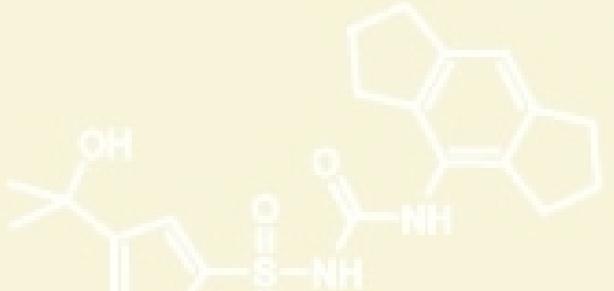
NLRP3

Pyrin

NACHT

OLRR

Walker B: RILFLMDGFDELQGAFDEHI

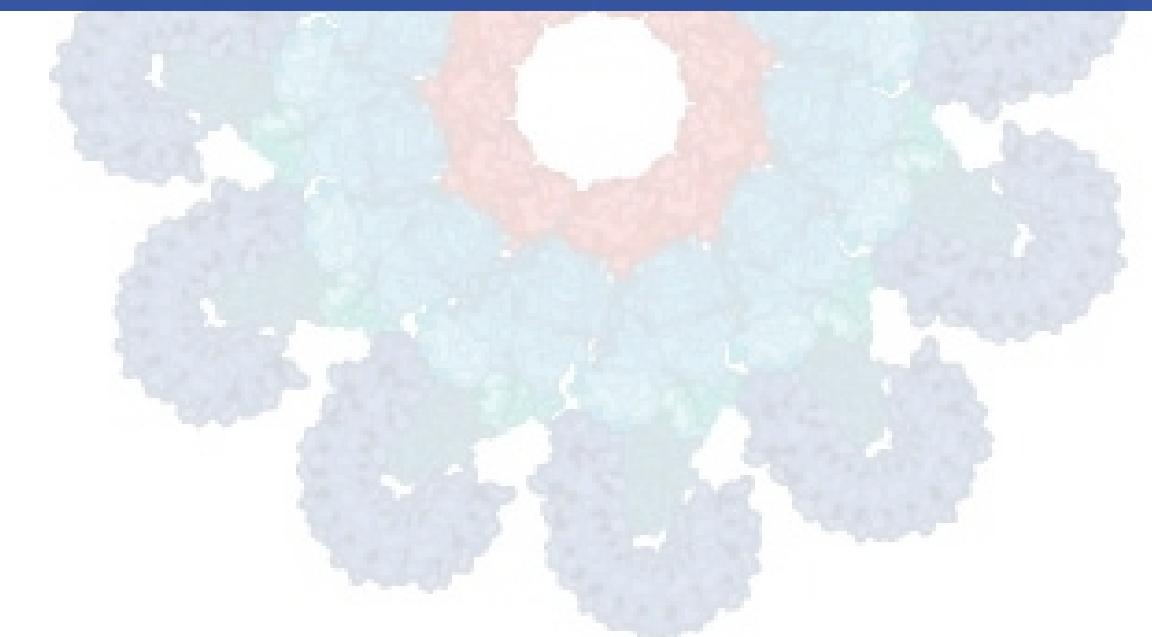


# METODOLOGI PENELITIAN



Inactive NLRP3

Walker B



Active NLRP3

# DESKRIPSI DATASET



Dataset yang digunakan dalam analisis ini berkaitan langsung dengan pengujian ikatan senyawa pada molekul NLRP3, yang merupakan bagian penting dari penelitian imunologi dan peradangan. Dataset ini awalnya berisi 843 sampel. Setelah proses eliminasi untuk menghapus duplikat, jumlahnya berkurang menjadi 833 sampel yang siap untuk dianalisis.

## Fitur Fitur Dataset

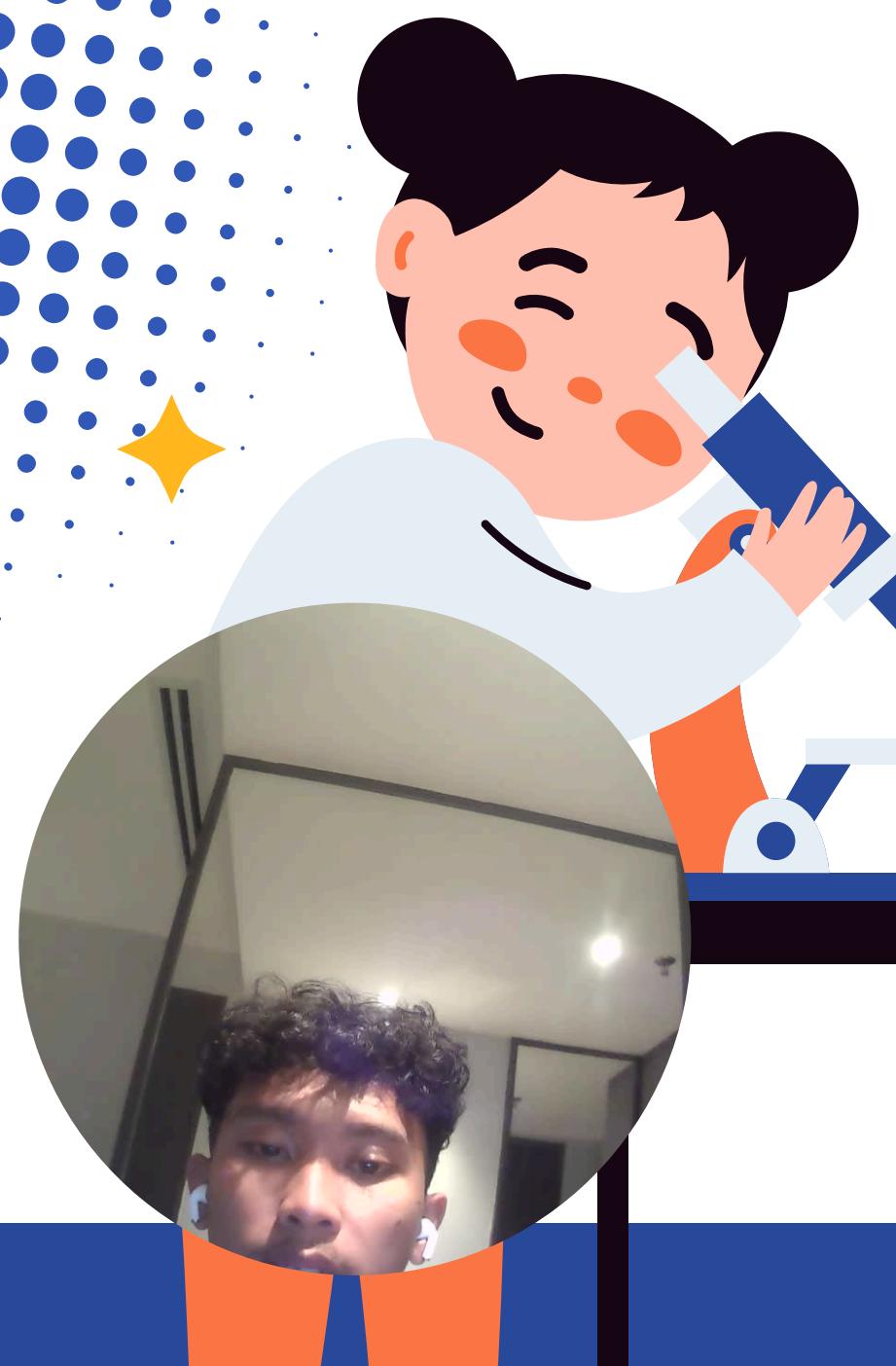
- Molecule Chembl ID: ID unik untuk setiap molekul dari database ChEMBL.
- Canonical SMILES: Notasi tekstual yang menggambarkan struktur kimia molekul.
- Activity Comment: Status aktivitas molekul ('active' atau 'inactive') berdasarkan pengujian.
- pchembl Value: Skor yang menilai efektivitas biologis molekul.
- Potential Duplicate: Tanda apakah entri mungkin duplikat untuk menjaga keunikan data.

## Variabel Target

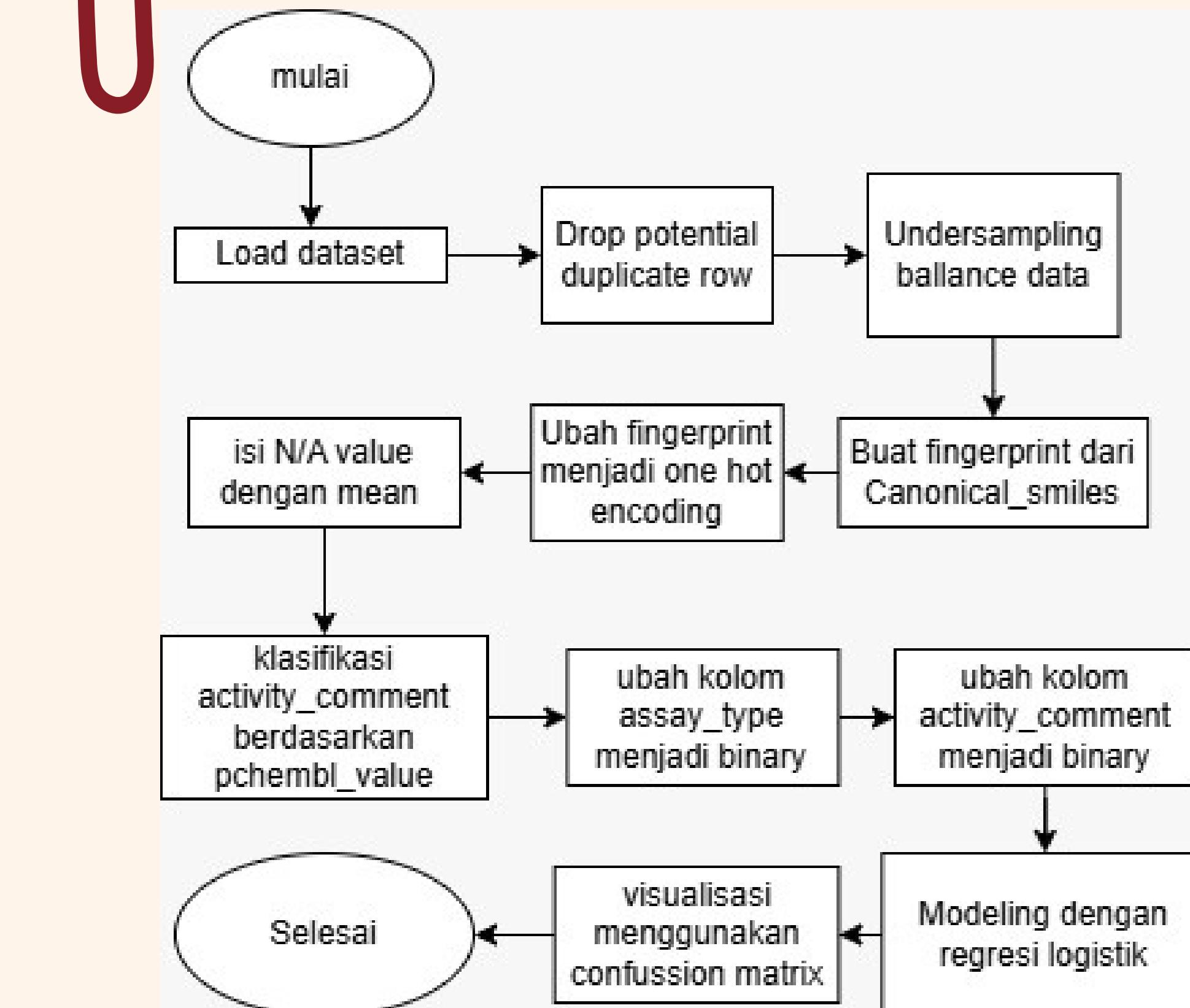
Assay Type, dengan dua klasifikasi:

- B (1): Data yang mengukur pengikatan senyawa ke target molekuler.
- F (0): Data yang mengukur efek biologis dari suatu senyawa





# Flowchart Penelitian



# MODEL REGRESI LOGISTIK

Regresi logistik digunakan untuk memodelkan hubungan antara variabel prediktor (fitur) dan variabel target biner (aktif dan tidak aktif) dalam klasifikasi tipe Assay (pengujian ikatan kimia). Model regresi logistik dipilih karena kemampuannya yang efektif dalam mengatasi masalah klasifikasi biner. Model ini memberikan probabilitas output yang dapat diinterpretasikan langsung sebagai peluang molekul memiliki tipe assay tertentu, yang sangat berguna dalam konteks medis dan farmasi.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

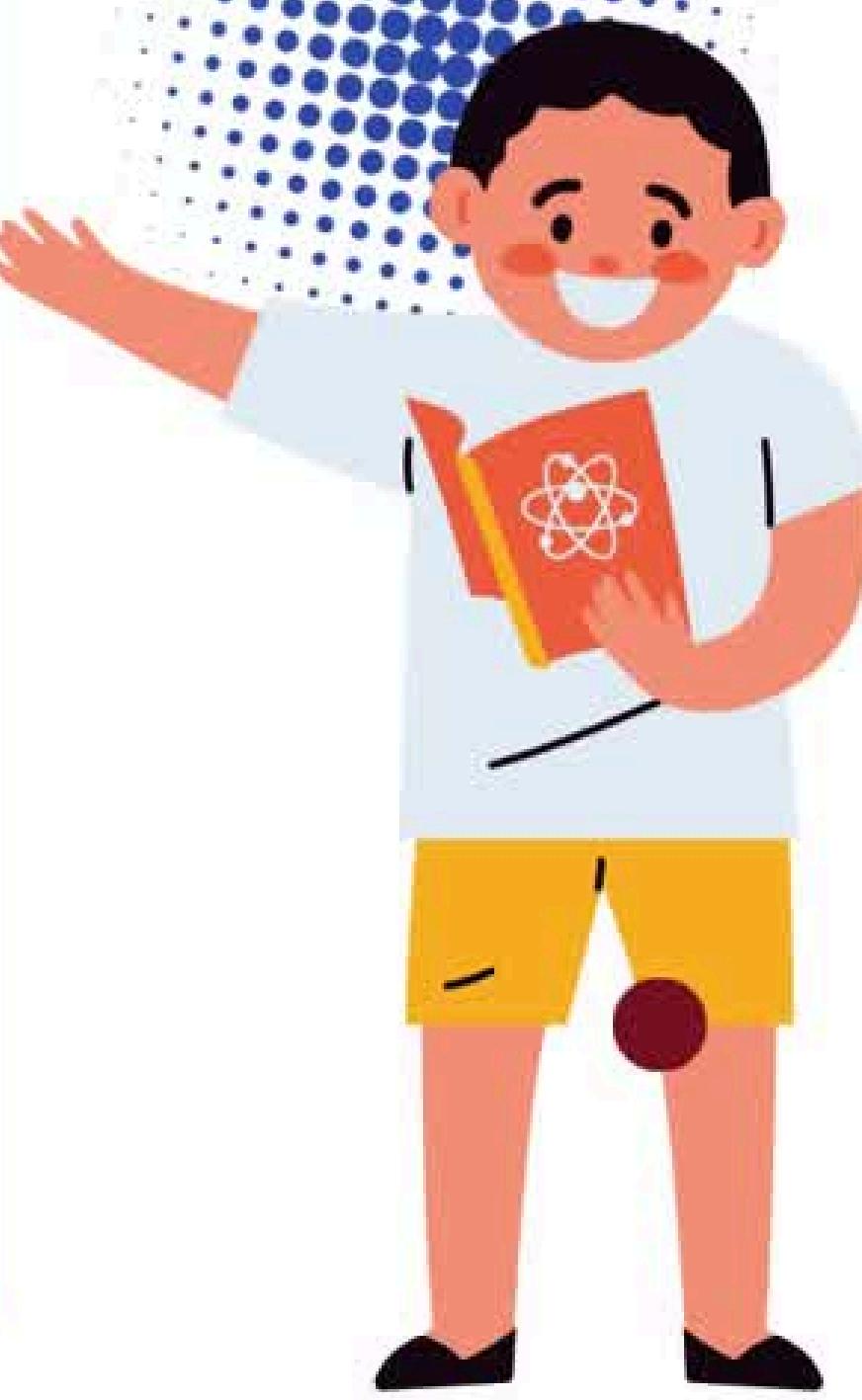
Keterangan:

- p adalah probabilitas bahwa Y=1
- $\beta_0, \beta_1, \beta_2, \dots, \beta_n$  adalah koefisien model yang diestimasi dari data.
- $X_1, X_2, \dots, X_n$  adalah variabel independen.

# METRIK EVALUASI MODEL

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	TP (True Positive)	FP (False Positive) <small>Type I Error</small>
	0 (Negative)	FN (False Negative) <small>Type II Error</small>	TN (True Negative)

Evaluasi yang digunakan untuk mengklasifikasi tipe Assay pada pengujian tipe senyawa NLRP3 menggunakan metrik utama, yakni : akurasi, presisi, recall, dan F1-skor. Selain itu, evaluasi model juga dilengkapi dengan analisis menggunakan confusion matrix untuk mengidentifikasi distribusi kesalahan klasifikasi antara kelas



# METRIK EVALUASI MODEL

## Accuracy:

Ukuran yang menunjukkan seberapa tepat suatu model prediksi dalam memprediksi kejadian yang benar.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

## Recall:

Mengukur kemampuan model dalam mengidentifikasi dengan tepat seluruh sampel yang termasuk dalam kelas positif.

$$Recall = \frac{TP}{TP + FN}$$

## Presisi

Mengukur seberapa akurat model dalam memprediksi kelas tertentu dari semua prediksi positif yang dibuat.

$$Presisi = \frac{TP}{TP + FP}$$

## F1-Score:

Pengukuran dengan menggabungkan presisi dan recall, sehingga memberikan keseimbangan antara keduanya.

$$F1 Score = \frac{2 \times (Presisi \times Recall)}{Presisi + Recall}$$



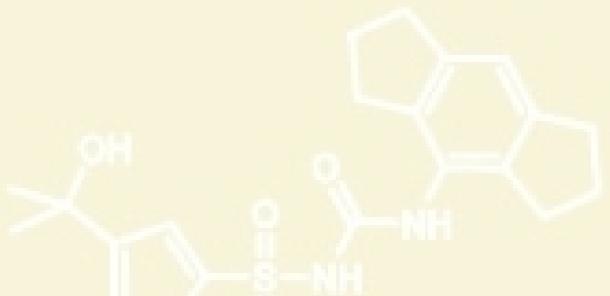
NLRP3

Pyrin

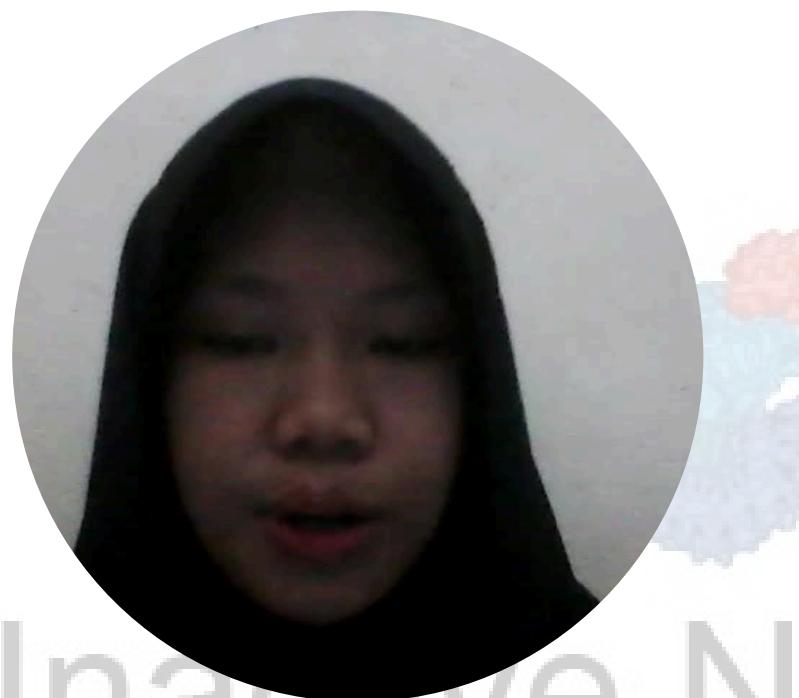
NACHT

LLRR

Walker B: RILFLMDGFDELQGAFDEHI

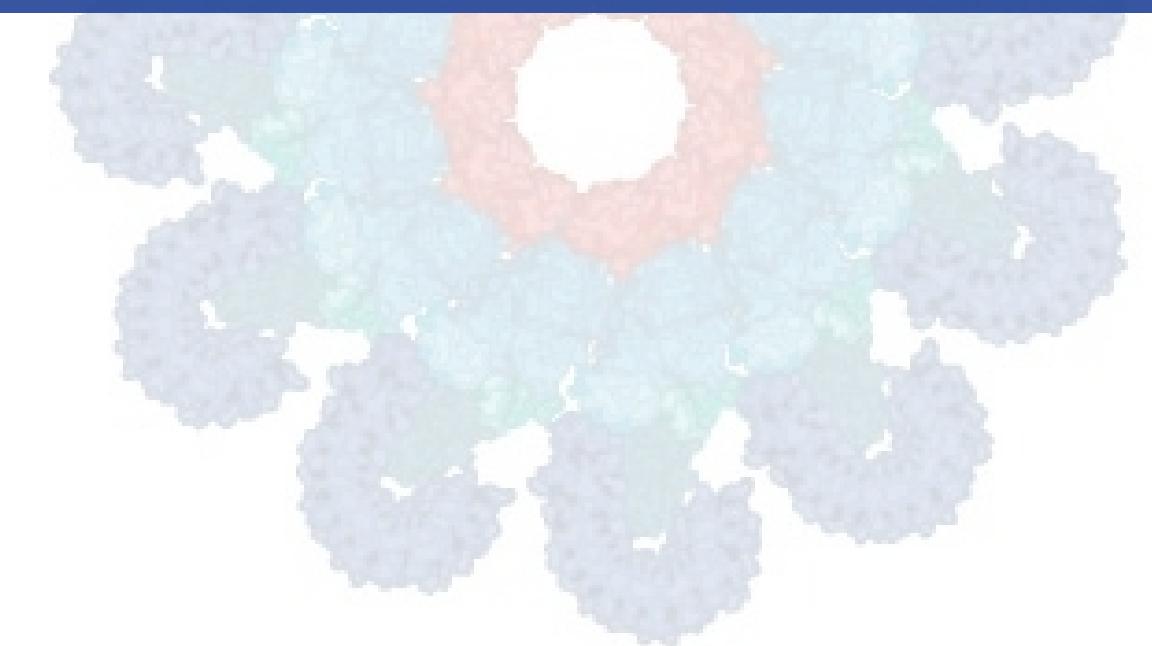
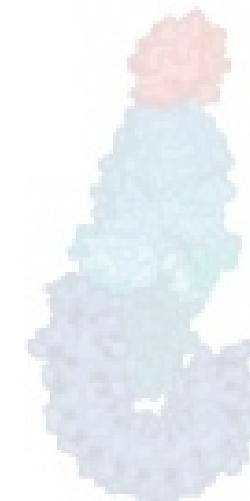


## HASIL DAN PEMBAHASAN



Inactive NLRP3

Walker B



Active NLRP3

# Hasil Pembahasan

Penelitian ini menggunakan dataset yang diambil dari database bioaktivitas ChEMBL, fokus pada molekul NLRP3 yang memiliki peran signifikan dalam respon inflamasi dan penyakit terkait.

Variabel target "assay\_type" diubah menjadi biner, dengan "pchembl\_value" di atas 5 dikategorikan sebagai aktif (1), dan 5 atau lebih rendah sebagai tidak aktif (0). Representasi molekul menggunakan Morgan Fingerprint yang dikonversi menjadi format one-hot encoding memfasilitasi analisis lebih lanjut.

Dataset awal berjumlah 843 entri, yang setelah eliminasi duplikat berkurang menjadi 833 sampel valid. Proses pra-pemrosesan meliputi eliminasi duplikat dan imputasi nilai yang hilang pada "pchembl\_value" dan "activity\_comment" untuk meningkatkan akurasi model.

Teknik undersampling diterapkan untuk menyeimbangkan jumlah sampel antara dua kelas assay, B dan F, masing-masing menjadi 136 sampel. Data yang telah diseimbangkan ini siap untuk analisis menggunakan model regresi logistik, yang bertujuan mengklasifikasikan molekul berdasarkan respons terhadap assay yang diuji.

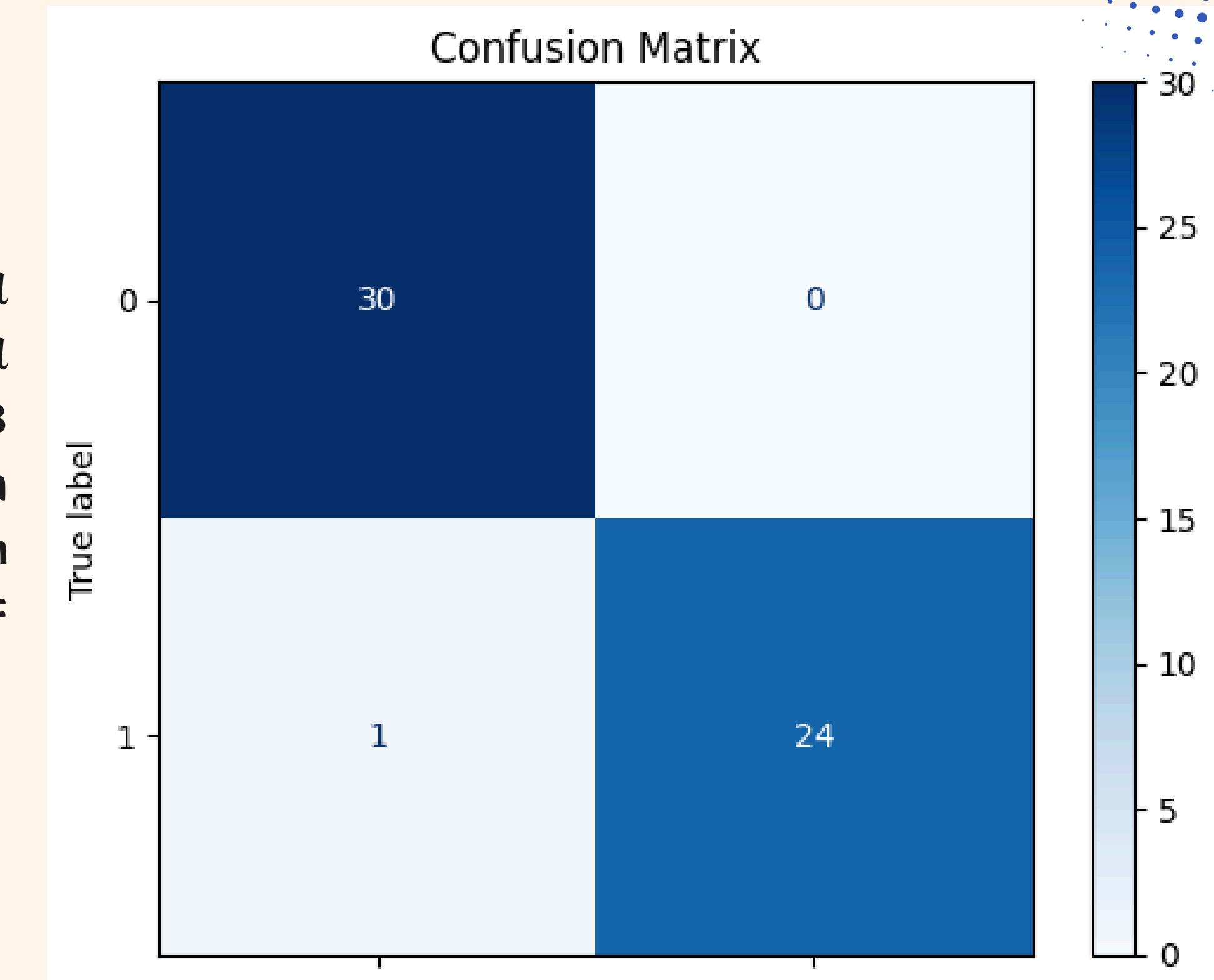
# Tampilan Data



# Hasil Evaluasi



Berdasarkan confusion matrix, dari 55 sampel yang diuji, model berhasil memprediksi 30 sampel kelas F (benar negatif) dan 24 sampel kelas B (benar positif). Tidak ada kesalahan dalam memprediksi kelas F, tetapi ada satu kesalahan di mana kelas B salah diprediksi sebagai kelas F (false negative).



# Hasil Evaluasi

	Precision	Recall	F1-Score	Support
0	0.97	1.00	0.98	30
1	1.00	0.96	0.98	25
Accuracy			0.98	55
Macro Avg	0.98	0.98	0.98	55
Weighted Avg	0.98	0.98	0.98	55

Model regresi logistik mencapai akurasi 98 % , dengan precision 100 % untuk kelas B dan 97 % untuk kelas F. Recall untuk kelas B adalah 96 % , sementara kelas F mencapai 100 %. F1-score untuk kedua kelas adalah 98 % , menunjukkan keseimbangan yang sangat baik antara precision dan recall. Hasil ini menandakan bahwa model sangat efektif dalam mengklasifikasikan tipe assay, dengan kesalahan minimal pada kelas B dan hampir tanpa kesalahan pada kelas F.

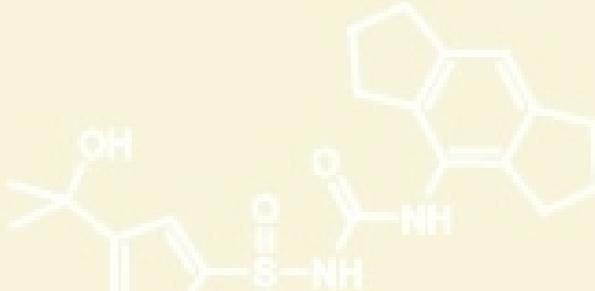
NLRP3

Pyrin

NACHT

LLRR

Walker B: RILFLMDGFDELQGAFDEHI

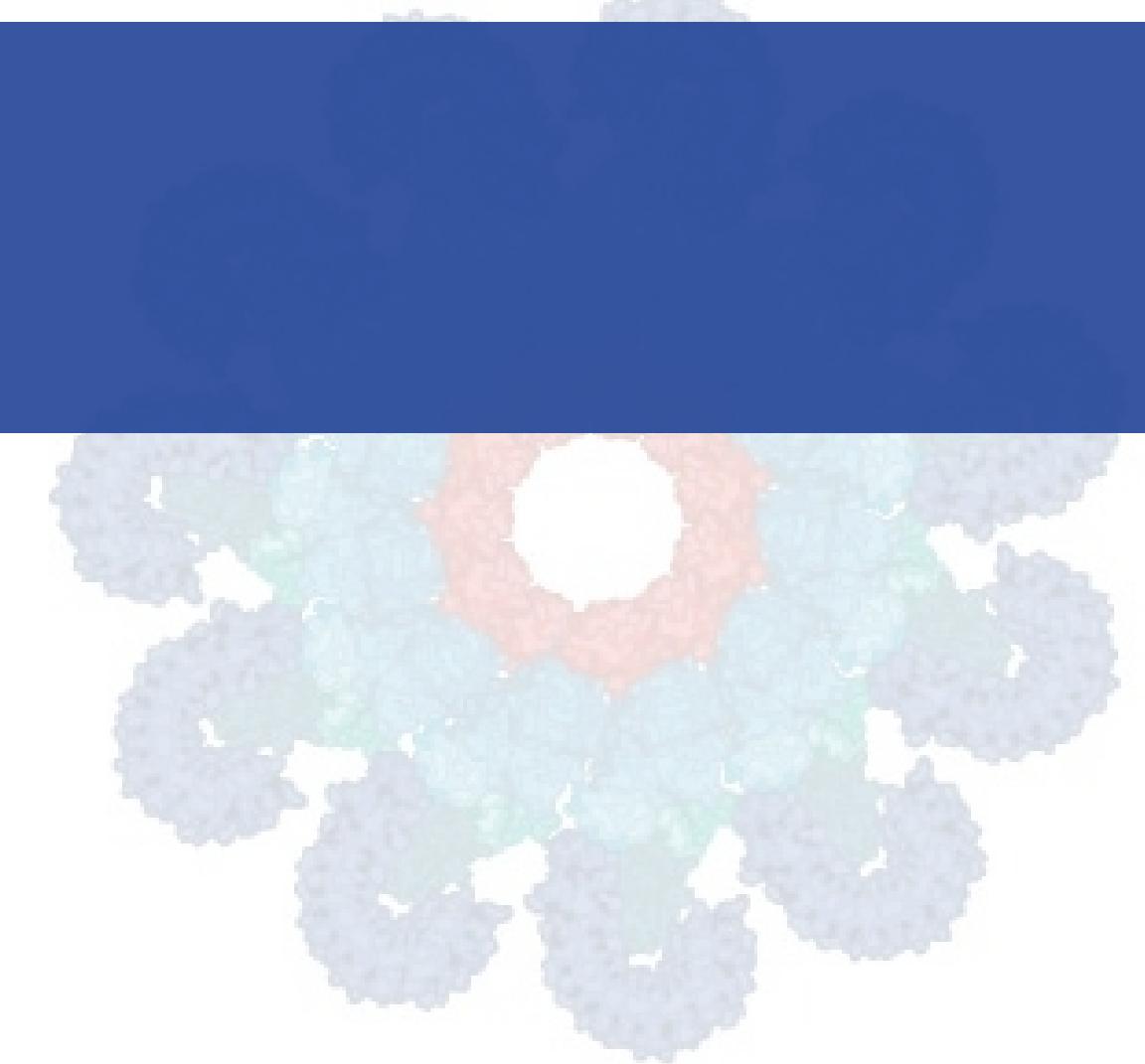
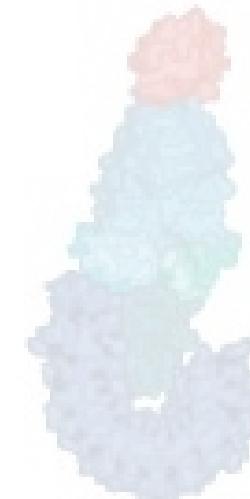


## KESIMPULAN



Inactive NLRP3

Walker B



Active NLRP3



# Kesimpulan

Penelitian ini menunjukkan bahwa model regresi logistik efektif dalam mengklasifikasikan tipe assay pada pengujian ikatan senyawa NLRP3, dengan akurasi mencapai 98 %. Data dari database ChEMBL berhasil diproses dan digunakan untuk membedakan dua kelas assay: pengukuran pengikatan senyawa (B) dan efek biologis senyawa (F). Hasil ini menegaskan pentingnya pemilihan metode analisis yang tepat dalam bioinformatika dan dapat memberikan wawasan baru untuk penelitian lebih lanjut mengenai interaksi molekuler NLRP3. Penelitian ini masih dapat dikembangkan lebih lanjut dengan metode yang lebih mutakhir seperti Deep Learning untuk menjangkau pola-pola tersembunyi dan lebih rumit.

# TERIMA KASIH!

