

BAB 1 PENDAHULUAN

1.1. Latar Belakang

Diabetes melitus merupakan penyakit tidak menular yang sedang menjadi target prioritas dalam tindak lanjut para pemimpin dunia [1]. Terdapat hitungan yang menunjukkan berkisar pada 422 juta orang di dunia menderita diabetes melitus dan sebanyak 1,6 juta kematian lebih tinggi terdapat pada negara yang memiliki penghasilan lebih rendah dan menengah terjadi pada usia sebelum 70 tahun [2]. Data dari International Diabetes Federation (IDF) menunjukkan bahwa penderita diabetes melitus mencapai angka 463 juta orang dan dengan perkiraan akan meningkat menjadi 578 juta orang pada tahun 2030 serta 700 orang pada tahun 2045 [2]. Sebanyak 232 juta dari 463 juta orang belum terdiagnosis sehingga terancam mengalami komplikasi tanpa disadari dan tanpa adanya pencegahan [2]. Pada tahun 2023, databoks mengemukakan bahwa diabetes tipe 2 paling banyak diderita orang Indonesia yaitu sebesar 50,2%. Dampak dari penderita penyakit diabetes ini membuat tubuh kekurangan insulin atau tidak bisa digunakan seperti biasanya [2].

Diabetes digolongkan menjadi tiga jenis: pertama, diabetes tipe 1, yang diduga disebabkan oleh reaksi autoimun dan biasanya di diagnosa pada anak-anak, remaja, dan dewasa muda. Kedua, diabetes tipe 2, yang berkembang secara bertahap dan umumnya didiagnosis pada orang dewasa dan ketiga, diabetes gestasional, yang berkembang pada wanita hamil yang sebelumnya tidak menderita diabetes [3]. Prevalensi diabetes melitus yang terus meningkat memerlukan pengendalian kadar gula darah, dan inhibitor DPP-4 telah terbukti efektif dalam mengelola diabetes tipe 2 [3]. Salah satu pendekatan pengobatan adalah dengan menghambat enzim dipeptidyl peptidase-4 (DPP-4), yang berperan dalam memecah hormon inkretin [4]. Inhibitor DPP-4 menghambat aktivitas enzim ini, sehingga meningkatkan kadar hormon inkretin, yang merangsang sekresi insulin, meningkatkan sensitivitas insulin, dan menghambat produksi glukagon [4].

Berbagai penelitian terdahulu telah menggunakan metode machine learning untuk memprediksi aktivitas inhibitor DPP-4, seperti metode Naïve Bayesian, Deep

Neural Network, dan Support Vector Machine. Namun, metode-metode tersebut masih menghadapi keterbatasan dalam mengelola kompleksitas data molekuler dan optimasi parameter model secara menyeluruh. Penelitian ini mengidentifikasi kesenjangan signifikan dalam prediksi aktivitas inhibitor Dipeptidyl Peptidase-4 (DPP-4) yang masih dihadapi oleh metode-metode konvensional maupun beberapa pendekatan machine learning sebelumnya, terutama terkait dengan pengelolaan kompleksitas data molekuler dan optimasi parameter model. Sebagian besar penelitian terdahulu mengandalkan algoritma standar yang kurang mampu menangkap pola jangka panjang dalam data sekuensial atau belum mengoptimasi parameter model secara menyeluruh. Menggabungkan Long Short-Term Memory (LSTM) yang mampu memodelkan dependensi jangka panjang dengan algoritma Monarch Butterfly Optimization (MBO) yang efektif dalam mencari konfigurasi parameter optimal [12].

Penggabungan kemampuan Long Short-Term Memory (LSTM) dalam memodelkan data sekuensial dengan algoritma Monarch Butterfly Optimization menjadi solusi yang menarik untuk mengoptimasi parameter model, sehingga mampu meningkatkan akurasi prediksi aktivitas inhibitor Dipeptidyl Peptidase-4 (DPP-4) secara signifikan [12].

1.2. Rumusan Masalah

Berdasarkan latar belakang yang disajikan sebelumnya, maka masalah penelitian ini dirumuskan sebagai berikut:

1. Bagaimana performa baseline dibandingkan dengan model LSTM yang telah dioptimasi menggunakan algoritma MBO?
2. Bagaimana Efektivitas Monarch Butterfly Optimization (MBO) untuk optimasi arsitektur LSTM?
3. Bagaimana performa LSTM dalam memprediksi inhibitor DPP-4 sebagai agen anti-diabetes melitus?

1.3. Tujuan dan Manfaat

Tujuan penelitian ini untuk menganalisis dan membandingkan performa model baseline dengan model Long Short-Term Memory (LSTM) yang telah dioptimasi menggunakan algoritma Monarch Butterfly Optimization (MBO) supaya mengetahui peningkatan akurasi prediksi aktivitas inhibitor DPP-4. Selain itu. Penelitian ini juga bertujuan mengevaluasi efektivitas algoritma Monarch Butterfly Optimization dalam mengoptimasi arsitektur model LSTM, spesifik dalam pencarian parameter optimal yang dapat meningkatkan kinerja model. Selanjutnya,

Penelitian ini mengukur kemampuan model LSTM yang telah dioptimasi dengan MBO dalam memprediksi inhibitor DPP-4 sebagai agen anti-diabetes melitus, sebagai langkah awal pengembangan terapi berbasis komputasi yang lebih akurat dan efisien.

1.4. Batasan Masalah

Penelitian ini mengembangkan model Long Short-Term Memory (LSTM) yang dioptimasi dengan Monarch Butterfly Optimization (MBO) untuk memprediksi aktivitas inhibitor Dipeptidyl Peptidase- 4 (DPP-4) sebagai terapi diabetes melitus tipe 2. Data yang digunakan berupa 108 inhibitor dalam format SMILES, dibagi menjadi data train dan test. Model terbatas pada analisis data molekuler ini dan menghadapi tantangan ketidakseimbangan kelas antara inhibitor aktif dan tidak aktif. Optimasi hanya mencakup parameter model menggunakan MBO tanpa test laboratorium.

1.5. Metode Penelitian

Penelitian ini melalui tahapan studi literatur, perancangan sistem, implementasi model, pengujian dan validasi, analisis hasil, serta penulisan laporan, sebagaimana tercermin dalam jadwal pelaksanaan tugas akhir. Eksperimen berfokus pada 108 data inhibitor DPP-4 yang diperoleh dari literatur. Nilai IC_{50} setiap inhibitor dikonversi menjadi pIC_{50} , dengan label aktif (1) untuk $pIC_{50} > 6,526$ dan label tidak aktif (0) untuk $pIC_{50} \leq 6,526$. Dataset sudah memiliki pembagian training dan test yang ditentukan pembuatnya melalui kolom Test setc, di mana nilai 0 menandakan training set awal dan nilai 1 menandakan test set. Berdasarkan

kolom tersebut, diperoleh 81 sampel training set awal (75%) dan 27 sampel test set (25%) dari total dataset.

Sebagai baseline, digunakan arsitektur Long Short-Term Memory (LSTM) 1–3 lapis dengan aktivasi ReLU, optimizer Adam (learning rate 0,001), batch size 32, dan early stopping untuk mencegah overfitting. Performa baseline kemudian dioptimalkan menggunakan Monarch Butterfly Optimization (MBO) dengan penyetelan parameter ukuran populasi 30–40, partition 5/12, dan period 1,2 hingga diperoleh konfigurasi terbaik.

1) Validasi dalam (internal validation).

Training set awal dipecah kembali menjadi training final dan validasi menggunakan metode Stratified Shuffle Split dengan testing size = 0,3, sehingga 30% data digunakan untuk validasi dan 70% untuk training. Pembagian ini menghasilkan 57 sampel untuk training final dan 24 sampel untuk validasi, dengan tujuan memantau learning curve dan mendeteksi overfitting sejak dini.

2) Pengujian akhir (testing).

Menggunakan metode hold-out dengan porsi testing set 25%, model terbaik hasil optimasi diuji pada data yang tidak pernah digunakan selama proses training maupun validasi. Metrik evaluasi berupa akurasi, presisi, recall, F1-score, dan Macro-F1 dihitung dari confusion matrix untuk menilai keseimbangan prediksi pada kedua kelas. Selanjutnya, confusion matrix dianalisis untuk mengidentifikasi dominasi false positive dan false negative serta mengevaluasi dampak ketidakseimbangan kelas.

1.6. Jadwal Pelaksanaan

Jadwal pelaksanaan ini menjadi acuan dalam mengevaluasi tahap-tahap pekerjaan seperti yang tertuang dalam milestone yang sudah ditetapkan.

Tabel 1.1 Jadwal Pelaksanaan Tugas Akhir.

No.	Deskripsi Tahapan	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6
1	Studi Literatur						
2	Perancangan Sistem						
3	Implementasi Sistem						
4	Pengujian dan Analisis						
5	Penulisan Laporan						