

ABSTRAK

KLASIFIKASI GEN ESENSIAL PADA SEKUENS DNA LALAT BUAH (*DROSOPHILA MELANOGASTER*) MENGGUNAKAN METODE *BIDIRECTIONAL LONG SHORT TERM MEMORY (BiLSTM)*

Oleh

PUTRI SANTIKA MAYANGSARI

DNA berfungsi sebagai pembawa informasi genetik yang mengatur berbagai proses kehidupan, seperti pertumbuhan, perkembangan, dan kelangsungan hidup organisme, yang sangat penting untuk memahami fungsi biologis. Untuk mempelajari mekanisme tersebut, klasifikasi gen esensial, yang berperan penting dalam fungsi dasar seluler, sangat penting. Penelitian ini menggunakan metode *Bidirectional Long Short-Term Memory (BiLSTM)* untuk membuat model klasifikasi gen esensial pada sekuens DNA lalat buah. Pola sekuensial gen esensial dan non-esensial diidentifikasi melalui pengolahan dataset sekuens DNA dari *Drosophila melanogaster*. Untuk mencegah *overfitting*, metode regularisasi seperti *early stopping*, *dropout*, dan *L2 regularization* digunakan. Penggunaan *Random Undersampling* bertujuan untuk menyeimbangkan jumlah data tiap kelas dengan mengurangi sampel dari kelas yang lebih besar. Evaluasi model dilakukan menggunakan *Confusion Matrix* dengan metrik evaluasi yang digunakan yaitu *Sensitivity*, *Specificity*, *ROC-AUC*, dan *PR-AUC*. Hasil klasifikasi DNA *Drosophila melanogaster* pada data OEG mendapatkan hasil *Specificity* sebesar 73%, *Sensitivity* sebesar 80%, *ROC-AUC* sebesar 76%, dan *PR-AUC* sebesar 81%. Sedangkan, data CEG mendapatkan hasil *Specificity* sebesar 70%, *Sensitivity* sebesar 64%, *ROC-AUC* sebesar 67%, dan *PR-AUC* sebesar 46%. Pendekatan BiLSTM yang diusulkan membantu dalam pengembangan metode klasifikasi gen esensial.

Kata Kunci: Klasifikasi, *Drosophila melanogaster*, Sekuens DNA, BiLSTM, Gen Esensial.

ABSTRACT

KLASIFIKASI GEN ESENSIAL PADA SEKUENS DNA LALAT BUAH (*DROSOPHILA MELANOGASTER*) MENGGUNAKAN METODE *BIDIRECTIONAL LONG SHORT TERM MEMORY (BiLSTM)*

By

PUTRI SANTIKA MAYANGSARI

DNA functions as the carrier of genetic information that regulates various life processes, such as growth, development, and the survival of organisms, which is crucial for understanding biological functions. To study these mechanisms, the classification of essential genes, which play a vital role in basic cellular functions, is important. This research employs the Bidirectional Long Short-Term Memory (BiLSTM) method to develop an essential genes classification model on the DNA sequences of fruit flies. The sequential patterns of essential and non-essential genes are identified through the processing of DNA sequence datasets from *Drosophila melanogaster*. To prevent overfitting, regularization methods such as early stopping, dropout, and L2 regularization are used. Random undersampling is employed to balance the number of samples in each class by reducing the samples from the larger class. Model evaluation is performed using a Confusion Matrix, with evaluation metrics including Sensitivity, Specificity, ROC-AUC, and PR-AUC. The classification results for *Drosophila melanogaster* DNA on the OEG dataset yielded a Specificity of 73%, Sensitivity of 80%, ROC-AUC of 76%, and PR-AUC of 81%. Meanwhile, the CEG dataset resulted in a Specificity of 70%, Sensitivity of 64%, ROC-AUC of 67%, and PR-AUC of 46%. The proposed BiLSTM approach aids in the development of essential gene classification methods.

Keyword: Classification, *Drosophila melanogaster*, DNA Sequence, BiLSTM, Essential Genes.