

## ABSTRAK

### **KLASIFIKASI GEN ESENSIAL PADA SEKUENS PROTEIN *DROSOPHILA MELANOGASTER* MENGGUNAKAN METODE *BIDIRECTIONAL GATED RECURRENT UNIT (BiGRU)***

Oleh

**YULIA DWI PUTRI**

Gen esensial berperan penting dalam kelangsungan hidup organisme, dan identifikasinya memiliki aplikasi luas dalam biologi sintesis dan medis. Namun, teknik eksperimental sering kali mahal dan rumit, sehingga metode komputasional berbasis pembelajaran mesin menjadi alternatif yang lebih efisien. Metode yang digunakan adalah *Bidirectional Gated Recurrent Unit (BiGRU)* dengan menggunakan *Drosophila melanogaster* sebagai organisme model. Tujuan dari penelitian ini adalah untuk mengevaluasi kinerja metode *Bidirectional Gated Recurrent Unit (BiGRU)* dalam sekuens protein pada *Drosophila melanogaster*. Dataset yang digunakan diperoleh dari penelitian Beder, et al, (2021) dengan 2 jenis dataset yaitu *Cellular Essential Gene (CEG)* dan *Organismal Essential Gene (OEG)*. Ada dua skema arsitektur model dan dua skenario pembagian data, yaitu 80% *training* 20% *validation* dan 90% *training* 10% *validation*. Pada dataset CEG memiliki distribusi kelas yang tidak seimbang sehingga dilakukan proses *Random Undersampling (RUS)*. Hasil kinerja yang paling baik pada dataset OEG didapatkan pada skenario pembagian data 80% *training* 20% *validation* dengan nilai yang didapat adalah 83 % sensitivitas, 79 % spesifisitas, 80% nilai ROC-AUC dan 85% nilai PR-AUC. Hasil yang paling baik pada dataset CEG diperoleh dari pembagian data 90% *training* 10% *validation* dengan nilai yang didapat untuk sensitivitas adalah 73%, spesifisitas 55%, nilai ROC-AUC 64% dan nilai PR-AUC 46%.

Kata Kunci: BiGRU, *Drosophila melanogaster*, Gen Esensial, Sekuens Protein.

## **ABSTRACT**

### **CLASSIFICATION OF ESSENTIAL GENES IN DROSOPHILA MELANOGASTER PROTEIN SEQUENCES USING BIDIRECTIONAL GATED RECURRENT UNIT (BiGRU) METHOD**

**By**

**YULIA DWI PUTRI**

*Essential genes play a crucial role in the survival of organisms, and their identification has wide applications in synthetic biology and medicine. However, experimental techniques are often expensive and complex, making computational methods based on machine learning a more efficient alternative. The method used in this study is the Bidirectional Gated Recurrent Unit (BiGRU), with *Drosophila melanogaster* as the model organism. The aim of this research is to evaluate the performance of the BiGRU method in protein sequences of *Drosophila melanogaster*. The dataset used was obtained from the study by Beder et al. (2021) and consists of two types of datasets: Cellular Essential Gene (CEG) and Organismal Essential Gene (OEG). There are two model architecture schemes and two data-splitting scenarios: 80% training and 20% validation, and 90% training and 10% validation. The CEG dataset has an imbalanced class distribution, so Random Undersampling (RUS) was applied. The best performance for the OEG dataset was achieved with the 80% training and 20% validation split, yielding 83% sensitivity, 79% specificity, 80% ROC-AUC, and 85% PR-AUC. The best performance for the CEG dataset was obtained with the 90% training and 10% validation split, yielding 73% sensitivity, 55% specificity, 64% ROC-AUC, and 46% PR-AUC.*

**Keywords:** *BiGRU, *Drosophila melanogaster*, Essential Genes, Protein Sequence.*