

ABSTRAK

KLASIFIKASI GEN ESENSIAL PADA *DROSOPHILA MELANOGASTER* BERDASARKAN DNA *SEQUENCE* MENGGUNAKAN METODE *GATED RECURRENT UNIT (GRU)*

Oleh

Azahra Alya Hidayah

Gen esensial merupakan gen yang sangat penting dan diperlukan untuk kelangsungan hidup suatu organisme. Informasi terkait esensialitas gen digunakan dalam berbagai penelitian ilmu *science*. Gen esensial dapat diidentifikasi melalui DNA. Gen-gen ini dapat diidentifikasi secara eksperimental dengan berbagai teknik, namun memerlukan sumber daya yang besar. Untuk mengatasi hal tersebut, metode komputasi digunakan untuk mengklasifikasikan gen esensial. Metode yang digunakan adalah *Gated Recurrent Unit (GRU)*. Pada penelitian ini menggunakan *Drosophila melanogaster* yang merupakan salah satu organisme yang sering dijadikan model dalam penelitian *science*. Tujuan dari penelitian ini adalah untuk mengetahui kinerja metode *Gated Recurrent Unit (GRU)* dalam mengklasifikasikan DNA *sequence* 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)*. Terdapat tiga skema arsitektur model dan dua skenario pembagian data, yaitu 80% *training* 20% *validation* dan 90% *training* 10% *data validation*. Pada dataset CEG memiliki distribusi kelas yang tidak seimbang sehingga dilakukan proses *Random Undersampling (RUS)* untuk menyeimbangkan kelas. Hasil kinerja yang paling baik pada dataset OEG didapatkan pada skenario pembagian data 90% *training* 10% *validation* dengan nilai yang didapat adalah 73 % *sensitivity*, 72% *specificity*, 73% nilai ROC-AUC dan 78% nilai PR-AUC. Hasil yang paling baik pada dataset CEG diperoleh dari pembagian data 90% *training* 10% *validation* dengan nilai yang didapat untuk *sensitivity* adalah 72%, *specificity* 48%, nilai ROC-AUC 60% dan nilai PR-AUC 44%.

Kata Kunci : Gen Esensial, *Drosophila melanogaster*, DNA *sequence*, GRU.

ABSTRACT

CLASSIFICATION OF ESSENTIAL GENES IN DROSOPHILA MELANOGASTER BASED ON DNA SEQUENCE USING GATED RECURRENT UNIT (GRU)

By

Azahra Alya Hidayah

Essential genes are genes that are very important and necessary for the survival of an organism. Information about essentiality is used in various scientific research studies. Essential genes can be identified through DNA. Genes can be identified experimentally using various techniques, but it requires substantial resources. To solve this problem, computational methods are used to classify essential genes. The used method is Gated Recurrent Unit (GRU). In this research, *Drosophila melanogaster* was used as the research object. *Drosophila melanogaster* is an organism that is often used as an object in scientific research. The purpose of this research is to determine the performance of the GRU method in classifying *Drosophila melanogaster*'s DNA sequences. The dataset was obtained from research by Beder, et al, (2021) with 2 types of dataset, namely Cellular Essential Gene (CEG) and Organismal Essential Gene (OEG). There are three model architecture schemes and two data separation scenarios, 80% training 20% validation and 90% training 10% validation. The CEG dataset has an imbalanced class distribution so a Random Undersampling (RUS) process is carried out to balance the classes. The best performance results on OEG dataset obtained in data separation 90% training 10% validation scenario, the result are 73% sensitivity, 72% specificity, 73% ROC-AUC and 78% PR-AUC. The best results on CEG dataset obtained in data separation 90% training 10% validation with the values are 72% sensitivity, 48% specificity, 60% ROC-AUC and 44% PR-AUC.

Keywords : Essential genes, *Drosophila melanogaster*, DNA sequence, GRU.