ABSTRAK

KLASIFIKASI DNA-*BINDING* PROTEIN MENGGUNAKAN METODE *BIDIRECTIONAL GATED RECURRENT UNIT* (BIGRU)

Oleh

Jihan Cahya Fatimah

Protein pengikat DNA merupakan protein yang dapat mengikat DNA dan berinteraksi dengan DNA dalam membantu memodulasi fungsi DNA. Mengidentifikasi protein pengikat DNA menjadi salah satu fokus penelitian di bioinformatika karena signifikansinya dalam kegiatan biologis di dalam sel seperti membantu transkripsi sintesis protein, replikasi dan rekombinasi DNA. Protein pengikat DNA biasanya diidentifikasi dengan teknik eksperimental. Namun saat ini peneliti mengidentifikasi protein pengikat DNA menggunakan pembelajaran mesin. Tujuan penelitian ini untuk membuat sebuah model pembelajaran mesin yang digunakan untuk mengklasifikasi protein pengikat DNA. Metode yang digunakan yaitu metode *Bidirectional Gated Recurrent Unit* (BiGRU). Terdapat dua skenario arsitektur percobaan yaitu BiGRU *single layer* dan BiGRU *multi layer*, serta tiga skenario pembagian data yaitu 90% *training* 10% validasi, 80% *training* 20% validasi dan 70% *training* 30% validasi. Hasil penelitian yang didapatkan akan dianalisis nilai akurasi, nilai sensitivitas, nilai spesifisitas, dan nilai *Matthew Correlation Coefficient* untuk mengukur kinerja model yang telah dibuat. Data yang digunakan terdiri dari dua kelas yaitu protein pengikat DNA dan non protein pengikat DNA. Data tersebut diperoleh dari *Protein Data Bank* (PDB) yaitu dataset PDB1075 yang digunakan sebagai data *training* dan dataset PDB186 yang digunakan sebagai data *testing.* Setelah beberapa skenario percobaan dilakukan, didapatkan hasil tertinggi pada arsitektur BiGRU *single layer* yang mendapatkan hasil akurasi 81,72%, sensitivitas 90,32%, spesifisitas 73,11%, MCC 64,40%. Hasil penelitian menunjukkan bahwa metode BiGRU mampu mengklasifikasi protein pengikat DNA.

Kata Kunci : Untaian Protein, Protein Pengikat DNA, Klasifikasi, BiGRU.

ABSTRACT

CLASSIFICATION OF DNA-BINDING PROTEINS USING BIDIRECTIONAL GATED RECURRENT UNIT (BIGRU)

By

Jihan Cahya Fatimah

DNA-binding proteins are proteins that binds to DNA and interact with DNA to support modulate DNA function. Identifying DNA-binding proteins is one of the focuses research in bioinformatics because of the significant role of DNA-binding protein in biological activities cells such as transcription proses, DNA replication and recombination. DNA-binding proteins are usually identified by experimental techniques. However in recent years researchers are identifying DNA-binding proteins using machine learning. Purpose of this research is to create a machine learning model for classifying DNA-binding proteins. The used method is Bidirectional Gated Recurrent Unit (BiGRU). There are two architecture scenarios, BiGRU single layer and BiGRU multi layer, and then three data separation scenarios, 90% training `10% validation, 80% training 20% validation and 70% training 30% validation. The results of the research will be analyzed for accuracy score, sensitivity score, specificity score, and Matthew Correlation Coefficient score to measure the performance of the model. Dataset consists of two classes, DNA-binding proteins and non-DNA-binding proteins. Dataset is retrieved from the Protein Data Bank (PDB), which is a PDB1075 dataset used as training data and a PDB186 dataset used as testing data. After several experimental scenarios, the highest result is on the single layer BiGRU, the result are 81,72% accuracy, 90,32% sensitivity, 73,11% specificity, 64,40% MCC. The results showed that BiGRU was able to classify DNA-binding proteins.

Keywords: Protein Sequences, DNA-Binding Proteins, Classification, BiGRU.