

## ABSTRAK

### KLASIFIKASI *ASETILASI* DAN *METILASI* PADA DNA RAGI MENGUNAKAN METODE *LONG SHORT-TERM MEMORY* (LSTM)

Oleh

MOHAMMAD FAJAR

DNA merupakan molekul di dalam sel yang mengandung informasi genetik yang bertanggung jawab atas perkembangan dan fungsi suatu organisme. DNA merupakan sebuah cetak biru untuk membuat protein. Terdapat protein yang disebut *Histone*. *Histone* terdiri dari empat jenis, yaitu H2A, H2B, H3, dan H4 dan biasanya DNA berada di sekeliling *Histone*. DNA sebagai struktur dasar pembentuk protein, pada prosesnya akan terjadi proses transkripsi yang mengubah DNA menjadi RNA kemudian menjadi protein. Pada protein terjadi modifikasi pada asam amino saat proses, yaitu *Asetilasi* dan *Metilasi*. Pada Tahun 2005 dilakukan penelitian mengenai klasifikasi DNA oleh Pokholok et al., dan didapatkan data DNA dari organisme Ragi. Data tersebut berisikan sekuens DNA ada atau tidaknya *Histone* atau modifikasi DNA. Untuk melakukan klasifikasi DNA diperlukan sumber daya yang besar sehingga untuk mengatasi hal tersebut digunakanlah metode komputasi. Penelitian ini akan dilakukan pengklasifikasian biner terhadap DNA Ragi. DNA Ragi akan diklasifikasikan dengan menentukan apakah terdapat *Histone*, terjadi *Asetilasi* atau *Metilasi*. Metode yang digunakan adalah *Long Short-Term Memory* (LSTM). Penelitian ini bertujuan untuk melakukan *Experiment* seberapa baik kinerja pengklasifikasian secara spasial pada data DNA. *Dataset* yang digunakan diperoleh dari penelitian Pokholok et al., (2005) yang berisikan sepuluh *dataset*. Setiap *dataset* memiliki dua kelas, yaitu positif dan negatif. Pembagian data dilakukan dengan skenario, yaitu 80% *training* dan 20% *testing*. Evaluasi model dilakukan menggunakan *Confusion Matrix* dengan metrik evaluasi yang digunakan yaitu *Accuracy*, *Precision*, dan *Recall*. Hasil klasifikasi DNA Ragi terbaik didapatkan dari *Experiment 4* yang memiliki rerata *Accuracy* Pengujian yang terbaik dengan rerata *Accuracy* sebesar 63,10%, *Precision* sebesar 62,95%, dan *Recall* sebesar 77,80%.

**Kata Kunci:** DNA Ragi, *Histone*, *Asetilasi* dan *Metilasi*, LSTM.

## **ABSTRACT**

### **CLASSIFICATION OF ACETYLATION AND METHYLATION IN YEAST DNA USING LONG SHORT-TERM MEMORY (LSTM) METHOD**

**By**

**MOHAMMAD FAJAR**

DNA is a molecule inside cells that contains genetic information responsible for the development and function of an organism. DNA is a blueprint for making proteins. There are proteins called Histones. Histone consists of four types, namely H2A, H2B, H3, and H4 and usually DNA is around Histone. DNA as the basic structure of protein formation, in the process there will be a transcription process that converts DNA into RNA and then into protein. In proteins, there are modifications to amino acids during the process, namely acetylation and methylation. In 2005, research was conducted on DNA classification by Pokholok et al., and obtained DNA data from yeast organisms. The data contains DNA sequences of the presence or absence of Histone or DNA modifications. To perform DNA classification, large resources are needed so that to overcome this, computational methods are used. This research will do a binary classification of yeast DNA. Yeast DNA will be classified by determining whether there is Histone, Acetylation or Methylation. The method used is Long Short-Term Memory (LSTM). This research aims to experiment how well the spatial classification performance on DNA data. The dataset used is obtained from the research of Pokholok et al., (2005) which contains ten datasets. Each dataset has two classes, namely positive and negative. Data division is done by scenario, which is 80% training and 20% testing. Model evaluation is done using Confusion Matrix with evaluation metrics used, namely Accuracy, Precision, and Recall. The best Yeast DNA classification results were obtained from Experiment 4 which had the best average Testing Accuracy with an average Accuracy of 63.10%, Precision of 62.95%, and Recall of 77.80%.

**Keywords:** Yeast DNA, Histone, Acetylation and Methylation, LSTM.