

IMPLEMENTASI METODE *K-NEAREST NEIGHBOR* (KNN) PADA DATA BIOINFORMATIKA

(Studi Kasus: Klasifikasi *Gene Expression* pada Data *Maize Growth Stages*)

Nurlina Tri Wulandari

Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

INTISARI

Perkembangan teknologi semakin hari semakin pesat dan mempengaruhi perkembangan pada bidang lainnya. Salah satu bidang yang berkembang karena pengaruh perkembangan teknologi yaitu bidang pertanian. Jagung adalah salah satu komoditi tanaman terpenting di pertanian, baik di dunia maupun di Indonesia. Hal ini dikarenakan jagung memiliki kandungan karbohidrat yang tinggi. Oleh karena itu, benih jagung yang digunakan harus bermutu tinggi, baik mutu genetik, fisik maupun fisiologinya dan berasal dari varietas unggul. Penelitian ini bertujuan untuk melakukan dan mengevaluasi klasifikasi gen expression pada tanaman jagung, yaitu pengklasifikasikan tahap pertumbuhan pada organ jagung. Metode yang digunakan adalah K-Nearest Neighbor (KNN). Berdasarkan penelitian diperoleh nilai k optimal yaitu 7 dengan nilai akurasi dari confusion matrix sebesar 85.7 dan ROC dengan nilai AUC sebesar 0.788. Hal ini berarti metode KNN dapat digunakan untuk mengklasifikasi tahap pertumbuhan pada organ jagung.

Kata Kunci : Klasifikasi, *Gene Expression*, Jagung, *K-Nearest Neighbor* (KNN), AUC, *Variable Importance*

IMPLEMENTATION OF *K-NEAREST NEIGHBOR* (KNN) METHOD ON BIOINFORMATIC DATA

(Studi Case: Classification of Gene Expression on Data Maize Growth Stages)

Nurlina Tri Wulandari

*Departement of Statistics, Faculty of Mathematics and Natural Science
Islamic University of Indonesia*

ABSTRACT

*Tecnological development increase rapidly day by day and affect to other fields. One field that develops due to the affect of technological developments is agricultural field. Maize is important agricultural commodities, both in the world and in Indonesia because maize has a high carbohydrate content. Therefore, the maize that used as the seed must be high quality, both genetic, physic and physiological qualities and derived from superior varieties. This study aims to conduct and evaluate the classification of gene expression in maize on classifying stages of growth in the maize organs. The method used is *K-Nearest Neighbor* (KNN). Based on research/study obtained the optimal k is 7 with the accuracy value of confusion matrix 85.7 and ROC with AUC value 0.788. This means that KNN method can be used to classify the growth stages in the maize organs.*

Keyword: *Classification, Gene Expression, Maize, K-Nearest Neighbor (KNN), AUC.*