

**IMPLEMENTASI *K-NEAREST NEIGHBOR* DENGAN *COSINE SIMILARITY* UNTUK KLASIFIKASI ABSTRAK JURNAL INTERNASIONAL ILMU KOMPUTER**

Oleh

Ulva Fatma Fadhila — fadhilaulva@student.upi.edu

1400893

**ABSTRAK**

Penyebaran jurnal internasional atau artikel ilmiah sebagai bahan penunjang penelitian semakin meningkat. Hal tersebut membuat jumlah dokumen jurnal yang tersedia bertambah sehingga mempersulit pencarian dan penyajian jurnal. Diperlukan teknik pengolahan teks yang dapat mengkategorikan dokumen teks dalam jumlah besar sesuai dengan jenisnya, sehingga informasi yang tersedia dapat terorganisasi dengan baik dan mudah di akses sesuai dengan kebutuhan pengguna. Salah satu pemecahan masalah dalam pengkategorian dokumen teks dapat diselesaikan dengan menggunakan metode *text mining* yaitu klasifikasi. Penelitian ini menggunakan algoritma klasifikasi *K-Nearest Neighbor* dengan *cosine similarity* untuk mengkategorikan dokumen jurnal yang maksimal. Secara garis besar penelitian ini melakukan empat tahapan. Tahapan awal merupakan proses *preprocessing* data yang terdiri dari *case folding*, *character removal*, tokenisasi dan *stopword removal*. Tahapan kedua melakukan pembobotan TF-IDF pada setiap *term* dan melakukan *splitting* dataset yang berjumlah 450 jurnal untuk pembagian data *training* dan data *testing*. Metode *splitting* data yang digunakan yaitu *holdout method* dan *10 fold cross validation*. Tahapan ketiga melakukan pemodelan klasifikasi menggunakan *cosine similarity* dan *K-Nearest Neighbor* dengan nilai *k* yang digunakan yaitu 3,6,7,9. Terdapat 3 kategori abstrak diantaranya *Computer and Education*, *Computer and Security* dan *Computer in Human Behavior*. Tahapan terakhir melakukan analisis terhadap hasil dari klasifikasi yang dilakukan. Hasil menunjukkan metode *10 fold cross validation* mampu memberikan hasil yang lebih baik pada *k=9* dengan nilai presisi 80,18% recall 51,11% dan *f1-measure* 62,42%.

Kata kunci: klasifikasi, *k-nearest neighbor*, *cosine similarity*, *holdout method*, *10 fold cross validation*.

Ulva Fatma Fadhila, 2018

IMPLEMENTASI *K-NEAREST NEIGHBOR* DENGAN *COSINE SIMILARITY* UNTUK KLASIFIKASI ABSTRAK JURNAL INTERNASIONAL ILMU KOMPUTER

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

**IMPLEMENTATION OF K-NEAREST NEIGHBOR WITH COSINE  
SIMILARITY FOR CLASSIFICATION ABSTRACT INTERNATIONAL  
JOURNAL OF COMPUTER SCIENCE**

*Arranged by*

Ulva Fatma Fadhila — fadhilaulva@student.upi.edu

1400893

**ABSTRACT**

*The spread of international journals or scientific articles as research supporting material is increasing. This makes the number of available journal documents make it more difficult to find and present journals. Text processing techniques are needed that can categorize large amounts of text documents according to their type, the information available can be accessed properly and easily accessed according to user needs. One of the problem solving in categorizing text documents can be solved by using text mining method that is classification. This research uses the K-Nearest Neighbor classification algorithm with cosine similarity to categorize the maximum journal document. In general, this research conducted four stages. The initial stage is a process of preprocessing data consisting of case folding, character removal, tokenizing and stopword removal. The second stage performs TF-IDF weighting on each term and splits a dataset of 450 journals for data sharing training and data testing. Splitting method used is the data holdout method and 10 fold cross validation. The third stage is modeling the classification using cosine similarity and K-Nearest Neighbor with the value of k used is 3,6,7,9. There are 3 abstract categories including Computer and Education, Computer and Security and Computer in Human Behavior. The last stage of analyzing the results of the classification performed. The result shows 10 fold cross validation method can give better result at  $k = 9$  with precision value 80,18% recall 51,11% and f1-measure 62,42%.*

*Keyword: classification, k-nearest neighbor, cosine similarity, holdout method, 10 fold cross validation.*