

## INTISARI

### PERMODELAN KLASIFIKASI ARTIKEL BERITA *ONLINE* PENYEBAB KASUS NARKOBA MENGGUNAKAN KLASIFIKASI ROCCHIO

Oleh

**Laili Wahyunita**

**14/371985/PPA/04623**

Tindak kasus penyalahgunaan narkoba menunjukkan peningkatan pada beberapa tahun belakangan ini. Salah satu indikasi peningkatan kasus penyalahgunaan narkoba adalah semakin banyak artikel berita *online* yang memuat tentang kasus tersebut. Artikel berita *online* adalah sumber data yang sangat besar dan mudah untuk didapatkan. Semakin banyak situs berita *online* menjadikan data artikel berita semakin bertambah banyak pula. Hal ini menarik perhatian untuk diteliti apakah dapat diambil informasi dan pengetahuan yang ada di sumber data tersebut. Penelitian ini dilakukan untuk menggali informasi penyebab kasus narkoba dari suatu artikel berita *online* yang terkait dengan kasus narkoba.

Penelitian ini melakukan permodelan proses klasifikasi dengan menerapkan metode klasifikasi Rocchio. Rocchio dinilai sangat cocok pada data bertipe teks. Dokumen artikel berita merupakan salah satu bentuk data teks yang tidak terstruktur sehingga memerlukan proses pembersihan data dan pra pemrosesan terlebih dahulu. Klasifikasi Rocchio akan melakukan proses klasifikasi dengan mengubah isi dokumen berita menjadi vektor data. Vektor data menggunakan pembobotan TF-IDF. Dalam menentukan kelas pada data uji, klasifikasi Rocchio menggunakan pendekatan *cosine similarity*. Kelas dengan nilai kedekatan terbesar adalah kelas yang akan terpilih.

Pengujian dilakukan dengan menggunakan data uji sebanyak 50 data. Selain itu, pengujian juga dilakukan dengan metode *blackbox testing*. Pengujian *blackbox testing* bertujuan untuk melihat penilaian dari sudut pandang pengguna umum.

**Kata kunci:** Rocchio, TF-IDF, *cosine similarity*, *blackbox test*

## ABSTRACT

### ***CLASSIFICATION MODELLING OF ONLINE NEWS ARTICLE NARCOTICS CASE CAUSE USING ROCCHIO CLASSIFICATION***

By:

**Laili Wahyunita**  
**14/371985/PPA/04623**

Online news articles are very large data source and easy to obtain. Many a new news sites appear will increase the amount of online news articles also. This attracts the attention to know whether the information and knowledge that exist in the data source. This study conducted to explore the hidden information of root cause in drug case from online news articles related to drug cases.

This research builds model system of classification by applying the Rocchio classification method. Rocchio classification is suitable for the text data. Documents news article is one form of text data that is unstructured so that require cleaning process and pre-processing of data first. Pre-processing using tokenization, *stemming*, and *stopword* that will generate the feature of word document. Rocchio classification will perform the classification process by changing the content of news documents into vector data. Vector data using TF-IDF weighting. In determining the similarity between the test data and train data, the classification Rocchio using *cosine similarity* approach. Class with the greatest proximity value is the class that will be selected.

Tests performed by using 50 test data. In addition, testing was also done using by black box testing method. Testing black box testing aims to look at the ratings from the standpoint of the general *user*.

**keyword:** Rocchio, TF-IDF, *cosine similarity*, blackbox test