



INTISARI

KLASIFIKASI MULTI LABEL PADA ABSTRAK ARTTIKEL PENELITIAN MENGGUNAKAN METODE *MULTINOMIAL NAÏVE* *BAYES*

Oleh

Nurrofiqi Ankisqiantari
22/499561/PPA/06338

Penelitian ini mengklasifikasikan multilabel pada abstrak artikel penelitian. Abstrak artikel penelitian dapat diklasifikasikan menjadi lebih dari satu kategori, maka penelitian ini memfokuskan untuk melakukan klasifikasi multilabel pada abstrak artikel penelitian. Penelitian ini menggunakan dataset berupa abstrak dan kategorinya. Kategori yang digunakan ada 5, yaitu *Artificial Intelligence*, *Electrical & Electronic, Engineering, Information System*, dan *Robotic*. Pengumpulan dan palabelan dataset menggunakan website “*Web of Science*” dan terkumpul sebanyak 1044 data. Website tersebut menyediakan secara otomatis dalam pelabelan multilabel artikel penelitian. Penelitian ini dilakukan menggunakan metode *Multinomial Naive Bayes* (MNB). Pengujian model menggunakan *hamming loss* dan *confusion matrix*.

Proses penelitian diawali dengan *text preprocessing* yaitu *case folding*, *cleaning*, *stopwords removal*, dan *lemmatization*. Kemudian dataset direpresentasikan menggunakan *bag of words*. Dataset dibagi menjadi *data train* dan *data test* dengan perbandingan 80%:20%. Setelah itu data diklasifikasi menggunakan MNB, sehingga mendapatkan nilai probabilitas disetiap label kategorinya. *Threshold* yang ditentukan adalah 0,5. Kategori label yang memiliki nilai lebih dari *threshold* maka label tersebut masuk dalam prediksi.

Penelitian ini membandingkan nilai *threshold* pada MNB (0,03 ; 0,05 ; 0,3 ; 0,5) dan membandingkan tiga metode yaitu MNB, *Logistic Regression*, dan SVM menghasilkan nilai evaluasi yang tidak jauh berbeda. Berdasarkan hasil penelitian klasifikasi MNB dengan *threshold* 0,5 menghasilkan evaluasi yang lebih baik dari pada perbandingan lainnya yaitu *average accuracy* 71%, *average precision* 75%, *average recall* 74%, dan *average f1-score* 75,3%, sedangkan *average hamming loss* mendapatkan hasil sebesar 28,88%. Distribusi data yang tidak seimbang mempengaruhi hasil evaluasi.

Kata Kunci : abstrak, artikel, klasifikasi, multi label, *multinomial naïve bayes*



ABSTRACT

MULTI-LABEL CLASSIFICATION ON RESEARCH ARTICLE ABSTRACTS USING MULTINOMIAL NAÏVE BAYES METHOD

By

Nurrofiqi Ankisqiantari
22/499561/PPA/06338

This research focuses on classifying multi-labels on research article abstracts. Abstracts of research articles can belong to more than one category, so this study aims to perform multilabel classification on abstracts of research articles. The research utilizes datasets consisting of abstracts and their corresponding categories. Five categories are used: Artificial Intelligence, Electrical & Electronic, Engineering, Information System, and Robotic. The datasets were collected and labeled using the “Web of Science” website, resulting in a total of 1044 data points. The website provides automatic multi-label labeling of research articles. The research employed the Multinomial Naïve Bayes (MNB) method for classification, with model testing conducted using hamming loss and confusion matrix.

The research process starts with text preprocessing, including case folding, cleaning, stop-words removal, and lemmatization. The datasets is then represented using a bag of words and split into training and testing data with an 80%:20% ratio. Subsequently, the data is classified using MNB to obtain probability values for each category label. A threshold of 0,5 is set, where categories with values exceeding the threshold are included in the prediction.

The study compares different threshold value in MNB (0,03; 0,05; 0,3; 0,5) and evaluates three methods: MNB, Logistic Regression, and SVM to determine their performance. The results show that MNB classification with a threshold of 0,5 yields better evaluation metrics compared to other threshold, with an average accuracy of 71%, average precision of 75%, average recall of 74%, and average F1-Score of 75,3%. The average hamming loss is calculated to be 28,88%. The evaluation results are influenced by the unbalanced distribution of data.

Keyword: *abstract, article, classification, multi-label, multinomial naïve bayes*