

ABSTRAK

Artikel ilmiah sangat bermanfaat untuk ilmu pengetahuan hingga penyelesaian masalah dalam kehidupan. Untuk menemukan artikel ilmiah peneliti dan pembaca bisa mencari jurnal elektronik yang ada di internet. Internet menyediakan kemudahan akses namun banyak sekali artikel yang tersebar atau tsunami data akan menyulitkan pembaca untuk mencari artikel. Dan diperlukan ketelitian untuk menghindari publikasi jurnal predator. Untuk itu, dibutuhkan rekomendasi untuk mengelola informasi agar pembaca mendapat rekomendasi artikel yang sesuai dengan kebutuhan.

Penelitian ini membantu pembaca mengetahui artikel yang dibutuhkan dengan mengimplementasikan *content based filtering* dalam memberi rekomendasi. Dengan data artikel ilmiah yang berasal dari data sekunder pada website Kaggle milik Ammarabbasi. *Content based filtering* adalah metode yang memperhatikan kemiripan konten atau informasi. Pada penelitian ini dataset akan melalui proses preprocessing sebelum diolah. Kemudian data akan melalui pembobotan kata dengan TF-IDF, lalu dilanjutkan perhitungan similaritas menggunakan *cosine similarity*.

Rekomendasi yang dibangun akan diuji dengan menggunakan *confusion matrix*. Dari tabel *confusion matrix* akan didapat nilai *precision* dan *recall*. Dengan *threshold* 0.10 hasil yang diperoleh dari perhitungan *precision@k* yaitu k=5 88%, k=8 82%, dan k=10 83%. Sedangkan hasil *recall@k* pada yaitu k=5 51%, k=8 73%, dan k=10 87%.

Kata kunci: artikel ilmiah, rekomendasi, *content based filtering*, TF-IDF, *cosine similarity*

ABSTRACT

Scientific articles are very useful for science and solving problems in life. To find scientific articles, researchers and readers can search for electronic journals on the internet. The internet provides easy access, but there are so many articles scattered around or a tsunami of data that makes it difficult for readers to find articles. And care is needed to avoid publication in predatory journals. For this reason, recommendations are needed for managing information so that readers get recommended articles that suit their needs.

This research helps readers find out what articles they need by implementing content-based filtering in providing recommendations. With scientific article data originating from secondary data on Ammarabbasi's Kaggle website. Content based filtering is a method that pays attention to similarities in content or information. In this research, the dataset will go through a preprocessing process before being processed. Then the data will go through word weighting with TF-IDF, then continue with similarity calculations using cosine similarity.

The recommendations built will be tested using a confusion matrix. From the confusion matrix table, precision and recall values will be obtained. With a threshold of 0.10, the results obtained from the precision@k calculation are k=5 88%, k=8 82%, and k=10 83%. Meanwhile, the results for recall@k are k=5 51%, k=8 73%, and k=10 87%.

Keywords: scientific articles, recommendations, content based filtering, TF-IDF, cosine similarity