



## K-NN METHOD FOR REVIEW ANALYS PRODUCT ACCOUNTING SOFTWARE

Sucitra Sahara <sup>1</sup>, Rizqi Agung Permana <sup>2</sup>

<sup>1</sup> Program Studi Sistem Informasi, Universitas Bina Sarana Informatika

<sup>2</sup> Program Studi Teknologi Komputer, STMIK Antar Bangsa

email: <sup>1</sup> [sucitrasahara@gmail.com](mailto:sucitrasahara@gmail.com), <sup>2</sup> [rizqiagung@gmail.com](mailto:rizqiagung@gmail.com)

### ARTICLE INFO

#### Article History:

Received : 21 October 2020

Revised : 07 December 2020

Accepted : 10 December 2020

Published : 11 December 2020

#### Keywords:

K-Nearest Neighbor

Review Analysis

Software Product

#### IEEE style in citing this article:

S. Sahara and R. A. Permana, "K-NN Method for Review Analys Product Accounting Software", *Jurnal.ilmiah.informatika*, vol. 5, no. 2, pp. 85-92, Dec. 2020.

### ABSTRACT

Many companies have not implemented accounting software in financial management. Even though the current era of technology is increasingly updated and developing, more and more superior products are being issued by software development companies, especially in accounting software. There are not a few software products whose quality is still below standard or incomplete with features and facilities. So that researchers concentrate on companies or individual businesses that still use manual methods in processing their finances by helping and making it easier to choose the software product they will choose. Researchers first carry out the accounting software product selection stage based on an opinion or opinion of the public who have bought and used the software they choose and they pour this opinion into online media such as comments on a product selling site. Thousands of comments will be processed and grouped into data sets and this time the researcher processes the data classification using the k-Nearest Neighbor (K-NN) algorithm. By using the K-NN method, it is expected to be able to produce the expected accuracy value so that the data set processing is stronger and more valid. It turns out that after applying the data accuracy value obtained by 80.50%, it can be concluded that the K-NN method is very suitable for the concept of text mining this time and for selecting the data set in the form of text.

© 2020 Jurnal Ilmiah Informatika (Scientific Informatics Journal) with CC BY NC licence

## 1. PENDAHULUAN

Di negara Indonesia jika diambil data secara random sebanyak 60% perusahaan belum menggunakan *accounting* software dalam pengelolaan keuangan. Era teknologi sekarang sudah semakin update dan berkembang, dan sudah semakin

banyak produk produk unggulan yang dikeluarkan oleh perusahaan pengembang software khususnya pada *software accounting*. Tidak sedikit juga produk software yang kualitasnya masih dibawah standar atau belum lengkap secara fitur dan fasilitasnya. Oleh karenanya calon

pengguna software harus mengetahui terlebih dahulu produk mana yang dibidang layak untuk digunakan dalam pengelolaan *accounting* nya. Sehingga peneliti sedang konsen terhadap perusahaan atau usaha perorangan yang masih menggunakan cara manual dalam mengolah keuangan mereka dengan cara membantu dan mempermudah dalam pemilihan produk software yang akan mereka pilih. Peneliti terlebih dahulu melakukan tahap penyeleksian produk *software accounting* berdasarkan suatu pendapat atau opini masyarakat yang telah membeli dan menggunakan software yang mereka pilih dan opini tersebut mereka tuangkan ke media online seperti komentar pada suatu situs penjualan produk online. Dari ribuan komentar akan diolah dan dikelompokkan menjadi data set dan peneliti kali ini mengolah klasifikasi data dengan menggunakan metode algoritma K-Nearest Neighbor (K-NN), dimana algoritma K-NN adalah salah satu algoritma yang paling sesuai. Peneliti menemukan bahwa algoritma K-NN mampu mengolah data set yang sudah dikumpulkan dan dikelompokkan menjadi teks positif dan teks negatif khususnya dalam penyeleksian suatu teks. Dengan menggunakan metode K-NN diharapkan mampu menghasilkan nilai akurasi yang diharapkan sehingga pengolahan data set nya lebih kuat dan valid.

Beberapa penelitian sebelumnya sudah dilakukan dalam klasifikasi sentimen terhadap review produk yang tersedia secara online seperti Analisa sentiment apikasi *smartphone* dengan membandingkan metode *Support Vector Machine* (SVM) dan *Naïve Bayes*, Karena SVM adalah classifier, kemudian diberikan satu set contoh pelatihan, masing-masing ditandai sebagai milik salah satu dari dua kategori, algoritma pelatihan SVM membangun model yang memberikan contoh baru ke dalam satu kategori atau

yang lain[1]. Kategorisasi teks merupakan solusi yang tepat untuk mengelola informasi yang saat ini berkembang dengan sangat cepat dan melimpah. Kategorisasi teks membuat pengelolaan informasi tersebut menjadi efektif dan efisien. Dengan menggunakan kategorisasi teks, dapat dilakukan penyusunan dokumen menurut kategorinya, penyaringan terhadap email *spam*, melakukan penggalian opini (*opinion mining*) dan analisis sentimen. Algoritma kategorisasi teks saat ini telah banyak berkembang, antara lain: *Support Vector Machines* (SVM), *Naive Bayessian* (NB), pohon keputusan, *K-Nearest*. Metode K-NN adalah Metode yang paling umum digunakan untuk menyelesaikan masalah ini, berdasarkan teknik *cross-validasi*, secara otomatis akan menentukan jumlah cluster dalam dataset[2]. Pada penelitian lain metode Pendekatan berbasis jarak: K-NN diperkenalkan dalam kerangka MIL.

Banyak dari konsumen yang akan membeli produk khususnya *Accounting*, bingung dalam menentukan produk mana yang paling baik dari sekian produk yang ada, didukung betapa banyaknya produk software yang kurang akurat dalam menghentikan virus pada perangkat komputer mereka, dan banyak pula produk palsu. Kegiatan pemalsuan produk ini semakin marak pada masyarakat karena permintaan yang terus meningkat sehingga pemilik merek merasa kewalahan akan kegiatan pemalsuan barang ini[3].

Pada penelitian kali ini, penggunaan metode pengklasifikasian K-Nearest Neighbors diterapkan untuk klasifikasi *text* pada opini *public* mengenai *review* produk *software accounting* yang terdapat pada situs penjualan produk dalam bentuk komentar, dan data komentar akan di proses dan diklasifikasikan berdasarkan object yang sudah ditentukan, pengolahan data tersebut akan menghasilkan 2

kategori yaitu komentar positif dan negatif, hasil akhir akan dituangkan pada aplikasi review produk *software accounting* yang menunjukkan software tersebut baik atau tidak. Dikatakan bahwa memiliki terlalu banyak komputer dalam kehidupan kita sehari-hari menyebabkan masalah keamanan, karena orang tidak dapat melindungi komputer mereka dari peretas dan virus internet[4]. Trend analisis ini menunjukkan peningkatan penargetan platform pada komputer sebagai sasaran malware sehingga aplikasi review pada produk *software accounting* yang penulis buat sangat perlu untuk diprediksi dengan akurat sebelum konsumen membeli produk tersebut, dan sebagai bahan pertimbangan perusahaan dalam memilih produk *software accounting*.

## 2. METODE PENELITIAN

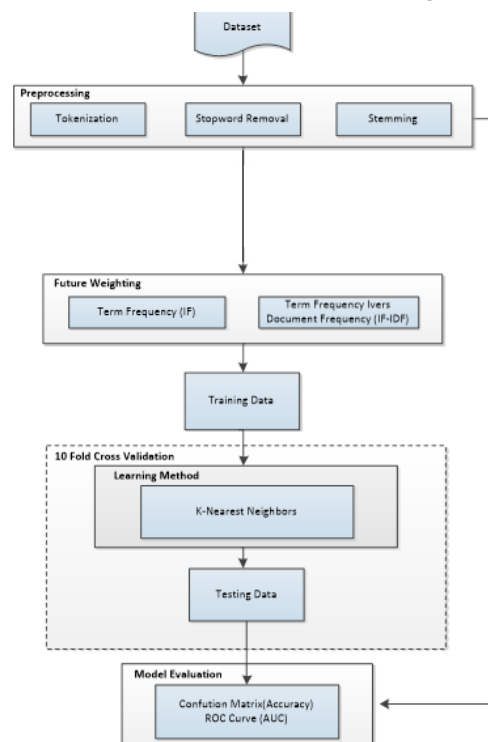
Kata atau Teks yang akan dilakukan proses text mining, pada umumnya memiliki beberapa karakteristik diantaranya adalah memiliki dimensi

tinggi, terdapat noise pada data, dan terdapat struktur teks yang tidak baik. Cara yang digunakan dalam mempelajari suatu data teks, adalah dengan terlebih dahulu menentukan fitur-fitur yang mewakili setiap kata untuk setiap fitur yang ada pada dokumen[5].

Setelah data dasar (komentar review) menjadi data terstruktur dan berupa nilai numerik maka data dapat disajikan sebagai sumber data yang dapat diolah lebih lanjut. Proses pengubahan bentuk menjadi data yang terstruktur sesuai kebutuhannya untuk proses dalam data mining, yang biasanya akan menjadi nilai-nilai numerik, proses ini sering disebut dengan *text processing* [6].

Pendekatan analisis mencakup tiga fase: pra-pemrosesan, di mana data mentah ditransformasikan menjadi format yang dapat digunakan, terutama dengan membersihkan, menetapkan atribut, dan mengintegrasikan data [6].

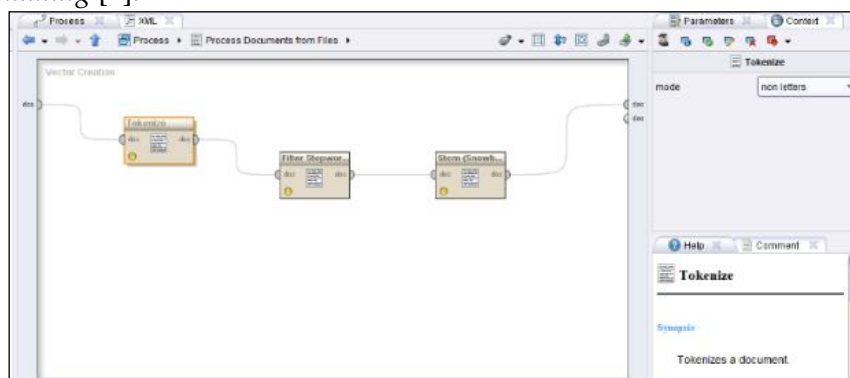
Berikut gambar metode usulan peneliti yang dirancang dalam *review product Accounting*:



Gambar 1. Metode Usulan

Beberapa proses yang dilakukan adalah sebagai berikut:

- Tokenization* adalah Peneliti menggunakan *Tokenize* untuk memisahkan kata atau huruf dari tanda baca dan simbol.
- Stopwords Removal* adalah kata yang dianggap tidak perlu dalam pengolahan data sentimen review, sebagai contoh *if, the, of, or, etc*
- Stemming* adalah Proses perubahan bentuk kata menjadi kata dasar. Metode perubahan bentuk kata menjadi kata dasar ini menyesuaikan struktur bahasa yang digunakan dalam proses *stemming* [7].



Gambar. 2. Tahap *Preprocessing* Data

Beberapa proses yang dilakukan pada Gambar 2 adalah *Tokenization* Peneliti menggunakan *Tokenize* untuk memisahkan kata atau huruf dari tanda baca dan simbol. *Stopwords Removal* kata yang dianggap tidak perlu dalam pengolahan data sentimen review, sebagai contoh *if, the, of, or, etc.* *Stemming* Proses perubahan bentuk kata menjadi kata dasar. Metode perubahan bentuk kata

### 3. HASIL DAN PEMBAHASAN

Data training yang digunakan dalam pengklasifikasian *text* terdiri atas 100 review positif pada produk *Accounting software* dan 100 review negatif pada produk *Accounting software*. Data review yang akan diolah masih berupa kumpulan *text* atau kalimat yang dipisah dalam bentuk dokumen *txt*. Sebelum diklasifikasikan, data tersebut harus melewati proses tahapan agar data dapat diolah dengan baik.

- Tahap *Preprocessing* *Tokenization*, *Stopwords Removal*, *Stemming* data sebelum di proses pengujian data.

menjadi kata dasar ini menyesuaikan struktur bahasa yang digunakan dalam proses *stemming*.

- Hasil Eksperimen Pengujian Data menggunakan Metode *k-NN*, Hasil Nilai query instance dalam penelitian disini ditentukan dengan cara melakukan uji coba memasukkan nilai *k* (jumlahangga terdekat) yaitu 1 sampai 10.

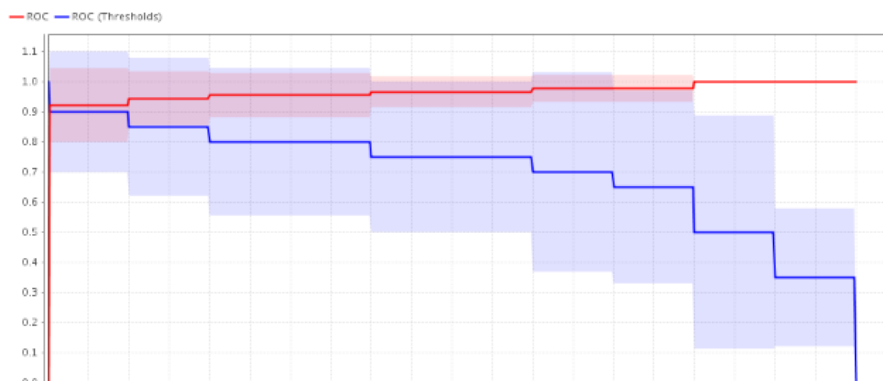
Tabel 2. Hasil Eksperimen *k-NN*

k	k-NN		Populations Size
	Accuracy	AUC	
1	67.50%	0.520	5
2	66.50%	0.652	5
3	76.00%	0.402	5
4	70.30%	0.645	5
5	80.50%	0.784	5
6	70.00%	0.790	5
7	75.50%	0.700	5
8	70.00%	0.667	5
9	64.50%	0.708	5
10	61.50%	0.795	5

Hasil eksperimen menunjukkan pada penerapan metode k-Nearest Neighbors pada tabel 2, bahwa k-Nearest Neighbors menentukan nilai k=5 dan population Size=5 menghasilkan  $accuracy=80.50\%$ ,

nilai  $AUC=0.784$  adalah nilai yang paling tertinggi.

### c. Model Avaluasi Kurva ROC (AUC Optimistic)



Gambar. 3. Kurva ROC (AUC Optimistic)

### Analisis Evaluasi Hasil dan Validasi Model

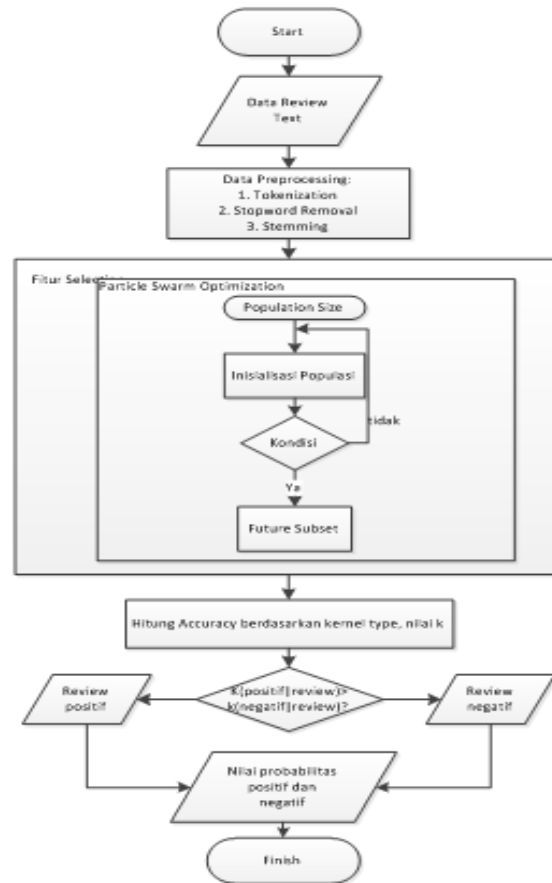
Dari hasil pengujian yang peneliti lakukan dari awal pembahasan, pengukuran akurasi menggunakan confusion matrix dan kurva ROC membuktikan bahwa hasil pengujian algoritma k-Nearest Neighbors (k-NN) menghasilkan akurasi yang cukup tinggi. Nilai akurasi untuk model algoritma k-NN nilai k=5 dan population Size=5 menghasilkan  $accuracy=80.50\%$ , nilai  $AUC=0.784$  adalah nilai yang paling tertinggi. Kesimpulan pengujian ini adalah penerapan algoritma k-NN dapat meningkatkan nilai akurasi yang merupakan solusi yang baik dalam permasalahan pada klasifikasi sentimen analisis review produk khususnya kali ini pada produk *Software accounting*.

### Pembahasan

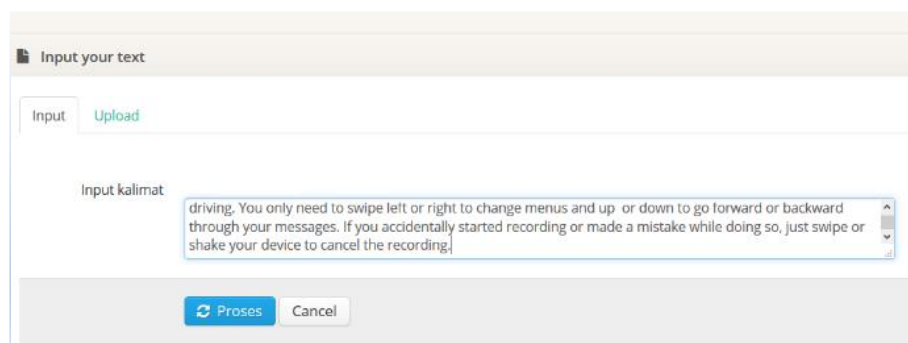
Penerapan algoritma k-NN dapat menghasilkan nilai akurasi pada klasifikasi review produk *Accounting software* dalam mengidentifikasi antara review positif dan review negatif, dengan menggunakan model klasifikasi teks pada review kali ini, pembaca dapat dengan mudah mengidentifikasi mana review yang positif

maupun yang negatif. Dari data review yang sudah ada, dipisahkan menjadi kata-kata, lalu diberikan bobot pada masing-masing kata tersebut. Dapat dilihat kata mana saja yang berhubungan dengan sentimen yang sering muncul dan mempunyai bobot paling tinggi. Dengan demikian dapat diketahui *review* tersebut termasuk *review* positif atau *review* negatif. Dalam penelitian ini, hasil pengujian model akan dibahas melalui *confusion matrix* untuk menunjukkan model yang terbaik. Tanpa menggunakan metode pemilihan fitur, k-Nearest Neighbors sendiri sudah menghasilkan akurasi yang cukup tinggi sebesar algoritma k-NN nilai k=5 dan population Size=5 menghasilkan  $accuracy=80.50\%$ , nilai  $AUC=0.784$ .

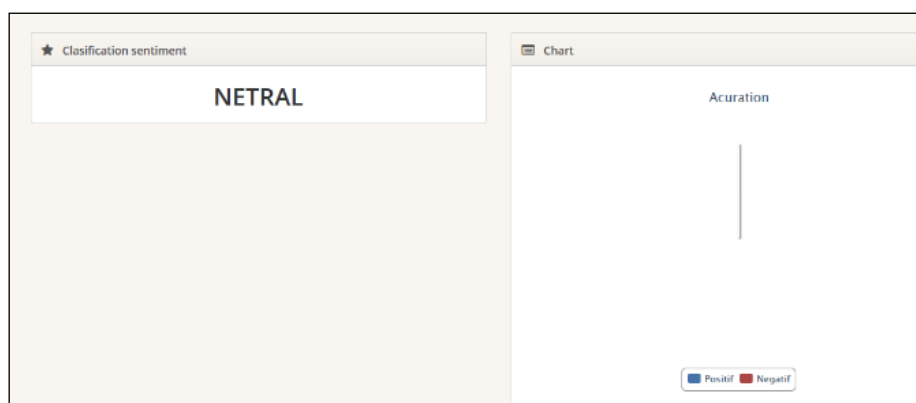
Peneliti menyediakan aplikasi untuk menguji model menggunakan dataset yang berbeda dan belum diklasifikasikan sesuai dengan kelasnya. Diaplikasikan dengan menggunakan bahasa pemrograman PHP berbasis *Web*. Dapat dilihat Gambar 4 diagram alir proses klasifikasi pada aplikasi yang dibentuk, dan Gambar 5 dan 6 hasil aplikasi yang di buat.



Gambar 4. Diagram Alir Tahapan Proses Klasifikasi Algoritma k-NN



Gambar 5. Proses Input Teks



Gambar 6. Hasil Penilaian Produk berdasarkan Komentar

Pada Gambar 6 Hasil Penilaian Produk berdasarkan Komentar menggambarkan hasil dari komentar yang dimasukkan termasuk nilai produknya netral artinya tidak jelek dan tidak bagus. Ada juga hasil yang menunjukkan hasil positif yang menunjukkan komentar tersebut baik dan hasil negatif yang menunjukkan komentar pada produk tersebut tidak bagus.

Implikasi penelitian ini mencakup beberapa aspek, di antaranya:

- a. Implikasi terhadap aspek sistem Hasil evaluasi menunjukkan K-Nearest Neighbors merupakan metode yang cukup baik dalam mengklasifikasi teks review produk pada produk *Accounting*. Metode ini dapat membantu para calon pengguna software (perusahaan) dalam menentukan aplikasi *Accounting* apa yang layak mereka gunakan pada pengolahan keuangan mereka, supaya tidak salah membeli aplikasi *software accounting* yang banyak tersedia diberberapa media penjualan online, yang dampak buruknya bisa mengakibatkan tidak lengkap fitur yang di buat oleh para pengembang *software accounting*, bahkan ada beberapa produk *Accounting* yang palsu bahkan didetect sebagai virus, dan mengurangi waktu para calon pembeli software dalam membaca review dan komentar mengenai produk *Accounting* software.
- b. Implikasi Aspek Manajerial Membantu para pengembang dan vendor sistem yang berkaitan dengan dunia komputer khususnya *Accounting*, baik dari sumber sosial media atau dari situs resmi para pengusaha dibidang *Accounting* software.
- c. Implikasi terhadap aspek penelitian lanjutan Penelitian selanjutnya bisa menggunakan metode pemilihan fitur ataupun dataset dari domain yang

berbeda, seperti review hotel, review restoran, dan banyak lainnya yang bisa dicari dalam bidang pengembangannya.

#### 4. UCAPAN TERIMA KASIH

Alhamdulillah puji syukur kepada Allah SWT, karena atas kehendak dan ridha Allah peneliti dapat menyelesaikan penelitian ini. Kami sadari penelitian ini tidak akan selesai tanpa doa, dukungan dan dorongan dari berbagai pihak. Adapun dalam kesempatan ini peneliti ingin mengucapkan banyak terima kasih kepada penyelenggara jurnal ini.

#### 5. KESIMPULAN

Klasifikasi text dengan data berupa review produk *Accounting* software yang peneliti buat, ternyata akurasi data lebih optimal dan nilai nya cukup tinggi dengan metode pengklasifikasian K-Nearest Neighbors. Hal ini dikarenakan K-NN metode yang dapat sesuai dengan klasifikasi data dan mudah dipahami. K-NN juga sering digunakan pada beberapa peneliti dalam klasifikasi teks dan memiliki performa yang baik. Dari pengolahan data yang sudah dilakukan. Data review yang peneliti olah dapat diklasifikasi dengan baik ke dalam bentuk positif dan negatif. Akurasi k-Nearest Neighbors sendiri sudah menghasilkan akurasi yang cukup tinggi sebesar algoritma K-NN nilai  $k=5$  dan population Size=5 menghasilkan accuracy=80.50%, nilai AUC=0.784. Untuk memudahkan penelitian, dibuatlah aplikasi review produk *Accounting* software untuk mengklasifikasikan review positif dan negatif yang ditampilkan dalam bentuk chart menggunakan bahasa pemrograman PHP. Model yang terbentuk dapat diterapkan pada seluruh data review produk *Accounting* software dari berbagai sumber, sehingga dapat dilihat secara langsung hasilnya dalam bentuk positif

dan negatif (*chart*). Hal ini dapat membantu seseorang untuk menghemat waktu saat mencari produk *Accounting* yang akan digunakan baik atau tidak. Kedepannya peneliti akan mencari model fitur selection lain yang dapat digunakan untuk klasifikasi data berupa text supaya perbandingan hasilnya dapat lebih optimal, mengimplementasikan model K-NN pada data review produk atau jasa lainnya, dalam pengembangan pengujian selanjutnya adalah memperbaiki proses data preparation dimana kualitas data yang akan diolah menjadi lebih baik sehingga pengolahan pada proses text mining menjadi lebih optimal.

## 6. REFERENSI

- [1] Y. Yin, D. Han, and Z. Cai, "Explore Data Classification Algorithm Based on SVM and PSO for Education Decision," *J. Conver. Inf. Technol.*, 2011, doi: 10.4156/jcit.vol6.issue10.16.
- [2] F. Gorunescu, *Data Mining: Concepts and Techniques*. 2011.
- [3] P. Kasih, "Perlindungan Hukum Bagi Masyarakat Terhadap Peredaran Barang Palsu," *J. Univ. Atma Jaya Yogyakarta*, vol. 5, no. 6, pp. 12–21, 2016.
- [4] M. Berndtsson, J. Hansson, B. Olsson, and B. Lundell, *Thesis Guide - A Guide for Students in Computer Science and Information Systems*. 2010.
- [5] D. P. Langgeni, Z. K. A. Baizal, and Y. F. A. W, "Clustering Artikel Berita Berbahasa Indonesia," *Semin. Nas. Inform.*, 2010.
- [6] M. Abdous, W. He, and C. J. Yen, "Using Data Mining for Predicting Relationships Between Online Question Theme and Final Grade," *Educ. Technol. Soc.*, vol. 15, no. 3, pp. 77–88, 2012.
- [7] M. S. Hudin, M. A. Fauzi, and S. Adinugroho, "Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi (Studi Kasus: Universitas Brawijaya)," *J. Pengemb. Teknologi Inf. dan Ilmu Komput.*, vol. 2, no. 11, pp. 5518–5524, 2018.