



# Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play menggunakan Metode Algoritma Support Vector Machines (SVM) Dan Naive Bayes

Ulfa Kusnia<sup>1</sup>, Fachrul Kurniawan<sup>2</sup>

<sup>1,2</sup>Magister Informatika, UIN Maulana Malik Ibrahim Malang, Malang, Indonesia

Email : <sup>1</sup>[19841003@student.uin-malang.ac.id](mailto:19841003@student.uin-malang.ac.id), <sup>2</sup>[Fachrlk@ti.uin-malang.ac.id](mailto:Fachrlk@ti.uin-malang.ac.id)

## INFO ARTIKEL

### Sejarah Artikel:

Diterima 8 Juni 2022  
Direvisi 10 Juni 2022 2022  
Disetujui 19 Juni 2022  
Dipublikasi 19 Juni 2022

### Katakunci:

Analisis Sentimen  
Aplikasi Media Berita Online  
Google Play  
Support Vector Machine (SVM)  
Naive Bayes

## ABSTRAK

Media berita online sebagai media massa yang paling banyak dikonsumsi publik yang bisa mengungguli media era sebelumnya misalnya media elektronik dan media cetak. Keunggulan media online dibanding media cetak biasanya yaitu praktis, real time dan up to date. Penilaian umum atas layanan dan berita yang diberikan sangat penting untuk menjaga dan meningkatkan kinerja media berita online. Adapun evaluasi publik dapat dilihat melalui page Google Play dalam kolom opini user. Analisis sentimen bisa menganalisis opini tersebut, dengan proses menganalisa dan mengekstraksi data teks yang tidak struktur untuk menghasilkan informasi sentimen yang terdapat dalam kalimat opini pada aplikasi. Dalam penelitian analisis sentimen ini, mengimplementasikan algoritma Naive Bayes dan SVM (Support Vector Machines), serta mengetahui kecenderungan opini publik di google play tentang aplikasi media berita online. Subjek diambil menggunakan scraping sebanyak 5615 opini. Hasil eksperimen menunjukkan bahwa keakuratan SVM (Support Vector Machines) 88 % lebih mengungguli dari pada Naive Bayes 87%, dan didapatkan pula kecenderungan opini publik di google play tentang aplikasi media berita online condong positif, hal tersebut dapat dilihat dari jumlah opini positif sebesar 5160 sedangkan negatif 455.

## ABSTRACT

Online news media as the most widely consumed mass media by the public can outperform the media of the previous era, such as electronic media and print media. The advantages of online media compared to print media are usually practical, real time and up to date. A general assessment of the services and news provided is very important to maintain and improve the performance of online news media. The public evaluation can be seen through the Google Play page in the user opinion column. Sentiment analysis can analyze these opinions, by analyzing and extracting unstructured text data to generate sentiment information contained in the opinion sentences in the application. In this sentiment analysis research, implementing the Naive Bayes algorithm and SVM (Support Vector Machines), and knowing the tendency of public opinion on google play about online news media applications. Subjects were taken using scraping as many as 5615 opinions. The experimental results show that the accuracy of SVM (Support Vector Machines) is 88% more superior than Naive Bayes 87%, and there is also a tendency for public opinion on google play about online news media applications to tend to be positive, this can be seen from the number of positive opinions of 5160 while negative 455.

### Keyword:

Sentiment Analysis  
Online News Media  
Google Play  
Support Vector Machine (SVM)  
Naive Bayes

### DOI Artikel:

10.35891/exploit.v14i1.3116

©2022 diterbitkan oleh Prodi Teknik Informatika Fakultas Teknik Universitas Yudharta Pasuruan

## 1. Pendahuluan

Seiring pesatnya perkembangan digitalisasi di aspek kehidupan dunia. Hal ini menyebabkan banyak perusahaan menciptakan digital baru terutama aplikasi media berita online untuk memenuhi kebutuhan informasi masyarakat. Media berita online sebagai media massa paling banyak dikonsumsi publik, karena bisa diakses dimana & kapan saja sejauh didukung teknologi internet. Sehingga bisa mengungguli media generasi sebelumnya yaitu media elektronik & media cetak [1]. Aplikasi media berita online bisa diakses melalui Google Play. Google Play bisa diakses melalui web dan android [2]. Adanya fitur kolom opini dalam Google Play, memungkinkan user menilai aplikasi pada bentuk penilaian & opini. Hal ini sebagai tujuan penelitian, untuk melihat evaluasi publik dari opini menurut user aplikasi media berita online. Opini user biasanya bisa berupa opini positif misalnya apresiasi & saran sedangkan opini negatif misalnya keluhan. Dengan demikian bisa menggambarkan tentang kinerja aplikasi media berita online pada para user. Untuk mensortir opini tersebut, tentunya bukan hal yang sederhana karena umumnya jumlah opini sangat banyak. Maka dari itu dilakukan teknik scraping atau penambangan teks. Teknik scraping atau penambangan yaitu mekanisme mengambil sebuah data semi-terstruktur dari page di internet.[4] Pada proses analisis sentimen untuk mengasosiasikan dan memproses tekstual data secara otomatis untuk menghasilkan kelompok sentimen(kelas positif atau negatif) yang terkandung dalam sebuah kalimat.[3]

Penelitian ini menggunakan, proses pembagian terstruktur atau klasifikasi mengenai machine learning dengan metode algoritma Support Vector Machine dan Naive Bayes. Beberapa karya ilmiah telah dilakukan dalam mengklasifikasikan sentimen pada opini yang tersedia secara online. Diantaranya, Sentiment Analysis On Gojek And Grab User Reviews Using Svm Algorithm Based On Particle Swarm Optimization [4], A comparative approach to Naive Bayes classifier and support vector machine for email spam classification [5], dan Employing Machine Learning techniques on Sentiment Analysis of Google Play Store Bangla Reviews [6].

Teknik Machine Learning metode algoritma Support Vector Machine dan Naive Bayes acap kali dilakukan peneliti pada bidang penambangan teks, lantaran bisa mengkomputasi data menggunakan dimensi tinggi sehingga tingkat akurasi yg didapatkan lebih baik. Kedua

prosedur algoritma ini juga dipakai pada penelitian ini pada penjabaran *opini user* Aplikasi Media Berita *Online*, bertujuan supaya yg terpilih algoritma yg paling akurat sehingga bisa menganalisis pengguna review Aplikasi Media Berita *Online*.

**2. Kajian Teori**

**2.1 Media Berita online dan Google Play**

Media Berita *online* yaitu Media yang disajikan menggunakan teknologi internet, yang tersaji secara cepat dan *ontime*. [7]. *Google Play* merupakan layanan konten digital milik *Google* terdiri dari toko produk *online* misalnya lagu, buku, aplikasi, game, atau pemutar media. *Google Play* bisa diakses melalui web dan android. [3].

**2.2 Analisis Sentimen**

Analisis sentimen adalah teknik penggalian informasi berupa sudut pandang seseorang terhadap suatu masalah dan juga merupakan proses memahami dalam memproses data tekstual secara otomatis untuk menghasilkan informasi kategori kelas negatif dan positif yang terkandung dalam kalimat. Analisis sentimen memiliki efek dan manfaat yang luar biasa, menyebabkan penelitian analisis sentimen tumbuh pesat.[3].

**2.3 Text Mining**

*Text Mining* adalah proses inovasi akan informasi atau isu terkini yang sebelumnya tidak terungkap menggunakan mekanisme dan menganalisa data pada jumlah besar. Dalam Menganalisa keseluruhan atau sebagian *unstructured text*, *text mining* membuat asosiasi satu bagian text dengan lainnya berdasarkan aturan tertentu. Hasil yang diharapkan adalah kata baru yang tidak terungkap jelas sebelumnya.[8].

**2.4 Metode Algoritma Naive Bayes**

Algoritma *Naive Bayes* adalah pengklasifikasi statistik yg bisa memprediksi probabilitas keanggotaan kelas data, classifier ini diklasifikasikan ke kelas tertentu sesuai teori probabilitas [9]–[11].

**2.5 Metode Algoritma Support Vector Machine**

Algoritma *Support Vector Machines* adalah teknik asparse. Seperti metode nonparametrik, SVM mengharuskan semua data pelatihan tersedia, yaitu, disimpan dalam memori selama fase pelatihan, ketika parameter model SVM dipelajari. Namun, setelah parameter model diidentifikasi, SVM hanya bergantung pada subset dari instance pelatihan ini, yang disebut vektor dukungan, untuk prediksi di masa mendatang.[12].

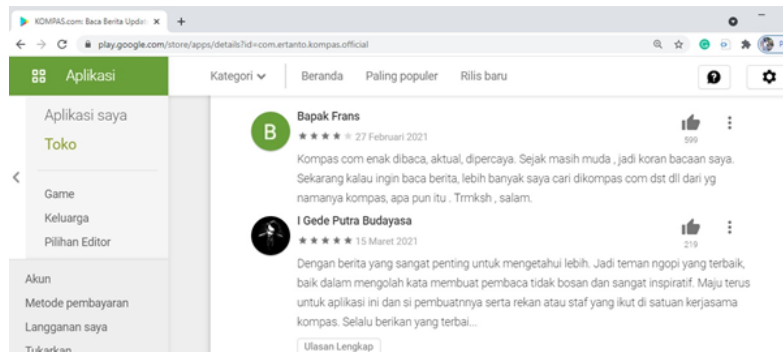
**3. Metodologi Penelitian**

**3.1 Studi Pustaka**

Studi pustaka akan dipaparkan penelitian – penelitian terdahulu yang berhubungan dengan permasalahan yang diteliti dan menjadi acuan konseptual. Dalam penelitian ini, metode pengumpulan data yang dilakukan merupakan studi pustaka, & studi literature.

**3.2 Pengumpulan Data**

Data yang digunakan adalah *opini* pengguna Aplikasi Media Berita *Online* pada kolom komentar di *Google Play* pada halaman situs <https://play.google.com/store/apps> . Komponen - komponen yang digunakan nama *user*, tanggal, rating, dan *user opini*, komponen tersebut dapat dilihat pada Gambar 1. Adapun data dikumpulkan menggunakan proses *scraping* dengan bantuan *tools scrape similar* pada *Google Crome*. Hasil *scraping* yang diperoleh setelah proses pengolahan dapat dilihat pada tabel berikut.



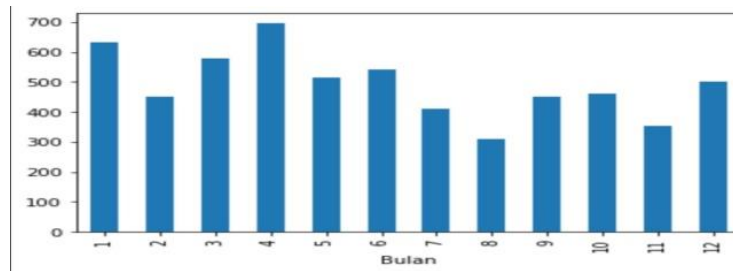
Gambar 1 Laman Kolom *Opini*

No.	Ulasan
0 1	jelek
1 2	Bagus sih beritanya lengkap, cuma saat membuka...
2 3	pemai yg lincah
3 4	Terkemuka dan Akurat
4 5	Baik
...	...
5889 5890	i Will try
5890 5891	update sekali
5891 5892	acaranya akurat .mantap
5892 5893	beritanya sangat terupdate
5893 5894	baca kompas sdh dari kecil

Gambar 2 Hasil *Scraping*

Berdasarkan gambar 2 dapat dilihat rating pada *Google Play* mempunyai range 1 sampai 5 dengan nilai 1 untuk kriteria “Sangat Tidak Suka”, nilai 2 untuk kriteria “Tidak Suka”, nilai 3 untuk kriteria “Cukup”, nilai 4 untuk kriteria “Suka”, dan nilai 5 untuk kriteria “Sangat Suka”. User yaitu pengguna aplikasi media berita online’. Ulasan atau review, yakni pendapat pengguna terhadap aplikasi.

Gambaran umum mengenai informasi dari aplikasi berita *online* berdasarkan jumlah *opini* dan *rating* yang diberikan pengguna.



Gambar 3 Grafik jumlah *opini*

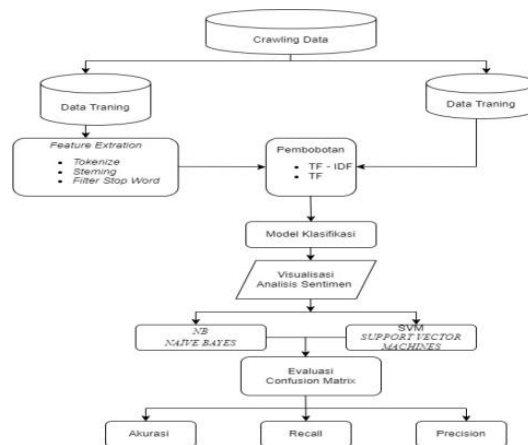
Berdasarkan grafik pada Gambar 3 dilihat jumlah ulasan dari para pengguna pada setiap bulannya tahun 2020. Jumlah ulasan terbanyak terdapat pada bulan April dengan 700 ulasan. Jumlah ulasan terkecil terdapat pada bulan Agustus dengan 300 ulasan. Pada bulan April terjadi kenaikan jumlah ulasan yang tinggi dibandingkan bulan-bulan lainnya. Hal ini disebabkan tingginya kebutuhan masyarakat terhadap berita terutama informasi tentang wabah corona atau covid 19.

3.3 Analisis Data

Pada tahap ini data akan diklasifikasikan kedalam kategori positif atau negatif dan data akan melalui proses *preprocessing* atau prapemrosesan untuk data *opini* yang masih belum terstruktur dengan baik dan belum seragam. Bertujuan untuk menyeragamkan dan menstrukturkan data yakni dengan mengurangi volume kosakata terutama pada karakter-karakter selain huruf, karena akan mempengaruhi nilai akurasi dari pengolahan data.

3.4 Metode Yang digunakan

Pada penelitian ini, metode yang digunakan yaitu metode algoritma *Support Vector Machine* dan *Naive Bayes*. Adapun design sistem pada gambar 2 berikut :



Gambar. 4 Design Sistem

Berdasarkan Gambar 4. Design sistem diatas. Proses analisis sentimen *opini* dimulai dari *scraping* dengan bantuan web *scraping extension for chrome* pada halaman situs aplikasi media berita *online* di *google play*, yang disebut dataset sebanyak 5615. Setelah data *opini* didapatkan kemudian data akan melewati tahap *text preprocessing* yaitu proses pembersihan data dilakukan dengan menggunakan *text mining* dengan tahap akan dilakukan yakni *tokenizing*, *stemming* dan *stopword*. Dilanjut dengan pembobotan *term*, menggunakan TF, IF dan IDF merupakan algoritma ilmu komputer yang digunakan untuk menghitung bobot *term* dari sebuah dokumen dilakukan dengan menghitung masing-masing nilai, yang terdiri model klasifikasi *Naive Bayes* dan *Support Vector Machine*. Setelah itu dilakukan evaluasi *confusion matrik* untuk mengetahui nilai *Accuracy*, *Recall* dan *Precision*.

3.5 Penerapan Sistem

Dilakukan penerapan sistem yaitu peng-codean menggunakan bahasa pemrograman python dengan memanfaatkan google colab. Dataset terdiri dari data Latih dan Uji. Proses Latih digunakan untuk menghasilkan model klasifikasi data *opini* yang selanjutnya akan digunakan sebagai dasar untuk mengklasifikasikan data *opini* dengan data uji atau data mentah yang baru.

4. Hasil dan Pembahasan

Dataset yang diujikan diambil dari *opini* pengguna Aplikasi Media Berita *Online* pada kolom comentari di *Google Play* pada halaman situs <https://play.google.com/store/apps> diambil tahun 2020, didapat sebanyak 5615 *opini*. Dengan perhitungan *opini* positif sebanyak 5160 sedangkan *opini* negatif sebanyak 455. Selanjutnya akan dilakukan *training* dan *testing* data set dan didapatkan nilai *Accuracy*, *Recall* dan *Precision*.

1. Proses analisis *opini* dan permofma metode yang diterapkan terdiri dari :

a. *Preprocessing* Data

Pada tahapan *preprocessing*, proses pembersihan data dilakukan dengan menggunakan *text mining*. Adapun tahap-tahap yang akan dilakukan yakni *tokenizing*, *stemming* dan *stopword*.

- **Tokenizing**  
Proses pemisahan teks dokumen menjadi kata per kata yang tidak saling berpengaruh (*independent*). *Tokenizing* berguna untuk mempermudah perhitungan frekuensi kemunculan kata dalam dokumen.

Tabel 1 *Tokenizing*

<i>Input</i>	<i>Output</i>		
Berita tidak seimbang bukan dari sumber asli. Muat berita HOAAX.	berita tidak seimbang	bukan dari sumber asli	muat berita hoax

- **Stemming**  
Tahap stemming merupakan suatu proses untuk mengubah kata – kata yang terdapat dalam suatu dokumen ke dalam kata – kata akarnya dengan menggunakan aturan – aturan tertentu. Proses stemming bahasa Indonesia dilakukan dengan menghilangkan sufiks, prefix, dan konfiks pada dokumen.

Tabel 2 *Stemming*

<i>Input</i>	<i>Output</i>
Terimakasih atas <b>kebaikan</b> nya, ke pada Kompas com, untuk <b>mem</b> bantu masyarakat Indonesia ini sangat cocok untuk <b>mem</b> berikan berita/kabar baik,	Terimakasih atas baik nya, ke pada Kompas com, untuk bantu masyarakat Indonesia ini sangat cocok untuk beribaru/kabar baik,

- **Stopword**  
Tahap menghilangkan kata yang tidak sesuai dengan topik dokumen.

Tabel 3 *Stopword*

<i>Input</i>	<i>Output</i>
Berita <b>gak</b> seimbang bukan <b>dari</b> sumber asli. Muat berita HOAAX.	Berita seimbang bukan sumber asli. Muat berita HOAAX.

2. Pelabelan Kelas Sentimen  
Pelabelan klasifikasi yang dilakukan pada analisis sentimen dapat dikelompokkan menjadi kelas sentimen positif dan negatif
3. Klasifikasi dengan Metode *Machine Learning*  
Proses klasifikasi dilakukan dengan menggunakan data training dan data testing dalam proses pembuatan machine learning, rasio yang digunakan dalam perbandingan data latih dan data uji adalah 80% : 20%.
4. Tahap evaluasi  
Dilakukan untuk menentukan nilai uji dari model yang berhasil diuji, dengan menghitung satu ukuran tertentu terhadap himpunan data uji, yaitu data yang tidak digunakan dalam proses pembuatan model classification tersebut. Dari hasil pengujian model menghasilkan *Confusion Matrik*.
5. Hasil Experiment algoritma *Support Vector Machine* dan *Naive Bayes*  
Uji coba dilakukan dengan melakukan input data set dengan rasio perbandingan training 80% : testing 20%. Dari hasil pengklasifikasian data opini kelas sentimen positif 5160 dan negatif 455. Adapun hasil masing – masing uji coba machine learning metode algoritma *Support Vector Machine* dan *Naive Bayes* sebagai berikut :

Tabel 4 Kinerja pada Pengklasifikasi *Support Vector Machine*

Traning Size	Testing Size	Akurasi %	
		Support Vector Machine	Naive Bayes
80%	20%	88.4648	87.2774

Berdasarkan tabel 4 diatas menunjukkan bahwa penerapan metode algoritma Support Vector Machine(SVM) menghasilkan tingkat akurasi yang lebih unggul dibandingkan metode algoritma Naive Bayes. Dengan pengujian dataset Training 80% :Testing 20%.

Tabel 5 Hasil *Confusion Matrik*

Aktual	Prediksi		<i>Precision</i>	<i>Recall</i>
	Positif	Negatif		
Positif	5160	0	1 / 100%	1 / 100%
negatif	0	456	1 / 100%	1 / 100%

Berdasarkan Tabel 4 dapat diketahui hasil evaluasi dari *machine learning* terhadap data uji. *Precision* adalah proporsi masing-masing kelas sentimen prediksi yang tepat dari semua kelas sentimen yang diprediksi. Pada kelas positif kemampuan sistem dalam memprediksi kelas positif dengan tepat dari semua kelas positif yang diprediksi sebesar 100%, sedangkan kelas negatif sebesar 100%. *Recall* adalah proporsi masing-masing kelas sentimen aktual yang tepat terprediksi dari semua kelas sentimen yang diprediksi. Pada kelas positif kemampuan sistem untuk menemukan kelas positif aktual yang terprediksi dengan tepat sebesar 100%, sedangkan kelas negatif sebesar 100%.

6. Menentukan Hasil Analisis Sentimen Opini  
Untuk mengetahui kecenderungan opini yang diuji termasuk kelas positif atau negatif, dapat dilihat dari banyaknya opini yang ada, yaitu positif 5160 opini dan negatif 455 opini. Untuk memperkuat data kecenderungan opini yang diuji dapat dilihat pada hasil pengujian model algoritma *Support Vector Machine* diketahui nilai akurasi 88% dan *Naive Bayes* 87%.

## 5. Kesimpulan

Berdasarkan hasil prosedur pengujian yang dilakukan terhadap Review Aplikasi Media Berita *Online* Pada *Google Play* Menggunakan Algoritma *Support Vector Machine (SVM)* dan Algoritma *Naive Bayes*. Hasil pengklasifikasian data *opini* kelas sentimen positif menjadi 5160 dan negatif menjadi 455. Hasil *Experimen machine learning* dengan perbandingan data *training* 80% dan data *testing* 20%, menghasilkan nilai akurasi untuk algoritma *Support Vector Machine (SVM)* 88% dan algoritma *Naive Bayes* 87%. Untuk data sentimen Riwew Aplikasi Media Berita *Online* membutuhkan upaya yang tidak mudah saat tahap *prossesing* diawal yaitu bagian pelabelan. Faktor penyebab sulitnya pelabelan yaitu penanganan sentimen negasi kata atau kalimat yang membalik makna pesan dalam bahasa indonesia, sehingga belum dapat ditentukan polaritasnya dengan optimal. Sebagai perbaikan untuk peneliti berikutnya perlu menggunakan metode machine learning lain. Sebagai pembanding hasil analisis sentimen algoritma *Support Vector Machine (SVM)* dan algoritma *Naive Bayes*.

## 6. Daftar Pustaka

- [1] B. Bayquni, "Partisipasi khalayak media online terhadap liptan6. com dalam memenangkan persaingan industri media massa di Indonesia," *Jurnal Pustaka Komunikasi*, vol. 1, no. 2, pp. 228–237, 2018.
- [2] S. Yunus, *Jurnalistik terapan*. Bogor: Ghalia Indonesia, 2010.
- [3] A. Mustopa, E. B. Pratama, A. Hendini, and D. Risdiansyah, "Analysis of user reviews for the pedulilindungi application on google play using the support vector machine and naive bayes algorithm based on particle swarm optimization," in *2020 Fifth International Conference on Informatics and Computing (ICIC)*, 2020, pp. 1–7.
- [4] H. Hermanto, A. Y. Kuntoro, T. Asra, N. Nurajijah, L. Effendi, and R. Ocanitra, "Sentiment Analysis On Gojek And Grab User Reviews Using Svm Algorithm Based On Particle Swarm Optimization," *Jurnal Pilar Nusa Mandiri*, vol. 16, no. 1, Art. no. 1, Mar. 2020, doi: 10.33480/pilar.v16i1.1304.
- [5] T. M. Ma, K. Yamamori, and A. Thida, "A comparative approach to Naïve Bayes classifier and support vector machine for email spam classification," in *2020 IEEE 9th Global Conference on Consumer Electronics (GCCE)*, 2020, pp. 324–326.
- [6] M. M. J. Soumik, S. S. M. Farhavi, F. Eva, T. Sinha, and M. S. Alam, "Employing machine learning techniques on sentiment analysis of google play store bangla reviews," in *2019 22nd International Conference on Computer and Information Technology (ICIT)*, 2019, pp. 1–5.
- [7] S. Santoso, "Analisis Resepsi Audiens Terhadap Berita Kasus Meiliana di Media Online," *Komuniti: Jurnal Komunikasi dan Teknologi Informasi*, vol. 12, no. 2, pp. 140–154, 2021.
- [8] Adiwijaya, "Text Mining dan Knowledge Discovery," 2006.
- [9] C. Darujati and A. B. Gumelar, "Pemanfaatan teknik supervised untuk klasifikasi teks bahasa indonesia," *Jurnal Bandung Text Mining*, vol. 16, no. 1, pp. 5–1, 2012.
- [10] F. Handayani and F. S. Pribadi, "Implementasi algoritma naive bayes classifier dalam pengklasifikasian teks otomatis pengaduan dan pelaporan masyarakat melalui layanan call center 110," *Jurnal Teknik Elektro*, vol. 7, no. 1, pp. 19–24, 2015.
- [11] A. A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, vol. 2, no. 1, pp. 22–26, 2016.
- [12] M. Awad and R. Khanna, "Support vector machines for classification," in *Efficient Learning Machines*, Springer, 2015, pp. 39–66.