

# Analisis Perancangan Aplikasi Pengklasifikasi Dokumen pada BPKA Deli Serdang Menggunakan Algoritma *Naive Bayes*

Yessika Floriana Manalu, Garuda Ginting

Program Studi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia

Email: yessikafloriana@gmail.com

Submitted 13-06-2021; Accepted 24-06-2021; Published 30-08-2021

## Abstrak

Pengklasifikasian dokumen di Badan Pengelolaan Keuangan dan Aset (BPKA) Kab. Deli Serdang sangat penting dilakukan untuk memudahkan proses disposisi surat bagi seluruh pegawai yang berada dalam instansi tersebut. Namun saat ini disposisi surat di BPKA Kab. Deli Serdang belum efektif dikarenakan masih dilakukan secara manual. Proses disposisi surat yang dilakukan secara manual akan menghabiskan waktu yang cukup lama dan seringkali terjadi kesalahan yang dilakukan oleh pegawai (*human error*) dalam menentukan klasifikasi dokumen karena dibutuhkan ketelitian dalam pengklasifikasian dokumen. Dari permasalahan diatas dapat disimpulkan bahwa Badan Pengelolaan Keuangan dan Aset Kabupaten Deli Serdang sangat membutuhkan suatu sistem pengklasifikasian dokumen. Dalam hal ini, terdapat beberapa metode pengklasifikasian dokumen yaitu *K-Nearest Neighbor (K-NN)*, *Support Vector Machine (SVM)*, *Cosine Similarity*, *Naive Bayes* dan masih banyak lagi metode pengklasifikasian. Dan metode *Naive Bayes* dianggap paling cocok dalam penerapan pengklasifikasian dokumen pada Badan Pengelola Keuangan dan Aset Kabupaten Deli Serdang. Pada penelitian ini, Penulis menggunakan metode Algoritma *Naive Bayes* dalam pengklasifikasian dokumen pada pendisposisian surat masuk di Badan Pengelolaan Keuangan dan Aset Kab. Deli Serdang. Dengan adanya penelitian ini diharapkan dapat membantu Badan Pengelola Keuangan dan Aset Kab Deli Serdang dalam melakukan disposisi surat.

**Kata Kunci:** Klasifikasi; Dokumen; BPKA; *Naive Bayes*; Deli Serdang

## Abstract

Document classification in the Financial and Asset Management Agency (BPKA) Kab. Deli Serdang is very important to do to facilitate the process of disposition of letters for all employees who are in the agency. However, currently the disposition of the letter at BPKA Kab. Deli Serdang has not been effective because it is still done manually. The process of disposition of letters that is done manually will take a long time and often mistakes are made by employees (*human error*) in determining the classification of documents because accuracy is needed in classifying documents. From the above problems, it can be concluded that the Financial and Asset Management Agency of Deli Serdang Regency is in dire need of a document classification system. In this case, there are several document classification methods, namely *K-Nearest Neighbor (K-NN)*, *Support Vector Machine (SVM)*, *Cosine Similarity*, *Naive Bayes* and many other classification methods. And the *Naive Bayes* method is considered the most suitable in the application of document classification to the Financial and Asset Management Agency of Deli Serdang Regency. In this study, the author uses the *Naive Bayes* Algorithm method in classifying documents on the disposition of incoming letters at the District Financial and Asset Management Agency, Deli Serdang. With this research, it is hoped that it can help the Financial and Asset Management Agency of Deli Serdang Regency in conducting the disposition of letters.

**Keywords:** Classification; Documents; BPKA; *Naive Bayes*; Deli Serdang

## 1. PENDAHULUAN

Dewasa ini dunia teknologi informasi dan komunikasi berkembang dengan sangat pesat. Hal ini menyebabkan bertambahnya data atau file elektronik yang digunakan sebagai media informasi yang sah khususnya pada roda pemerintahan di era revolusi industri 4.0. Oleh sebab itu, sebagai salah satu instansi pemerintahan, Kabupaten Deli Serdang dituntut untuk mengikuti perkembangan teknologi. Khususnya dalam hal pertukaran informasi antar instansi daerah sangat penting dilakukan sehingga mempermudah koordinasi sebagai suatu kesatuan pemerintahan. Pertukaran informasi secara digital di pemerintahan Kabupaten Deli Serdang saat ini dilakukan dengan mengirimkan surat melalui email. Dikarenakan banyaknya surat yang masuk setiap hari membuat Sub Bagian Umum kewalahan dalam melakukan disposisi surat.

Pengklasifikasian dokumen di Badan Pengelolaan Keuangan dan Aset (BPKA) Kab. Deli Serdang sangat penting dilakukan untuk memudahkan proses disposisi surat bagi sub bidang terkait yang berada dalam instansi tersebut. Namun saat ini pengelolaan dokumen di BPKA Kab. Deli Serdang belum efektif dikarenakan masih dilakukan secara manual. Dimana Kepala Sub Bagian Umum akan membaca dokumen yang akan didisposisi ke sub bidang terkait. Proses pengklasifikasian dokumen yang dilakukan secara manual akan menghabiskan waktu yang cukup lama karena mengharuskan membaca keseluruhan isi surat sebelum di disposisikan ke Sub Bidang terkait. Hal ini menyebabkan sering terjadinya penyampaian surat yang didisposisi ke Sub Bidang terkait mengalami keterlambatan informasi yang mengakibatkan tenggang waktu yang sedikit dan terburu-buru dalam pengambilan keputusan.

Dari permasalahan diatas dapat disimpulkan bahwa Badan Pengelolaan Keuangan dan Aset Kabupaten Deli Serdang sangat membutuhkan suatu sistem pengklasifikasian dokumen untuk meningkatkan efektivitas dan efisiensi manajemen disposisi surat. Pentingnya klasifikasi dokumen ditunjukkan oleh sejumlah penelitian mengenai topik ini. Salah satunya ialah klasifikasi dokumen berbahasa Indonesia berupa artikel berita menggunakan *Naive Bayes Classifier* [1]. Dalam hal ini, terdapat beberapa metode pengklasifikasian dokumen yaitu *K-Nearest Neighbor (K-NN)*, *Support Vector Machine (SVM)*, *Cosine Similarity*, *Naive Bayes* dan masih banyak lagi metode pengklasifikasian. Dan metode *Naive Bayes* dianggap paling cocok dalam penerapan pengklasifikasian dokumen pada Badan Pengelola Keuangan dan Aset Kabupaten Deli Serdang.

Akhmad dan Heru[2], dengan judul "*Naive Bayes Classification* pada Klasifikasi Dokumen untuk Identifikasi Konten *E-Government*". penelitian terkait selanjutnya adalah penelitian Wayan dan Agus[3], dengan judul "Klasifikasi Artikel

Berita Secara Otomatis Menggunakan *Metode Naive Bayes Classifier* yang Dimodifikasi”. Melihat hasil dari penelitian yang terdahulu maka penulis mengambil metode *Naive Bayes* sebagai metode dalam penelitian pengklasifikasian dokumen pada BPKA.

Algoritma *Naive Bayes* adalah metode pengklasifikasian dokumen dengan menghitung probabilitas dari sebuah dokumen terhadap kata-kata yang sama dalam dokumen lain yang berada dalam kategori tersebut. Algoritma ini dianggap tepat untuk digunakan dalam mengklasifikasikan dokumen karena memiliki komputasi yang sederhana dan akurasi yang tinggi. Pada penelitian ini, penulis menggunakan algoritma *Naive Bayes* untuk mengklasifikasikan dokumen pada Badan Pengelola Keuangan dan Aset Kabupaten Deli Serdang.

Dengan adanya penelitian ini dapat membantu Badan Pengelola Keuangan dan Aset Kabupaten Deli Serdang dalam melakukan disposisi surat sehingga proses disposisi surat terorganisir dan dilakukan dengan cepat.

## 2. METODOLOGI PENELITIAN

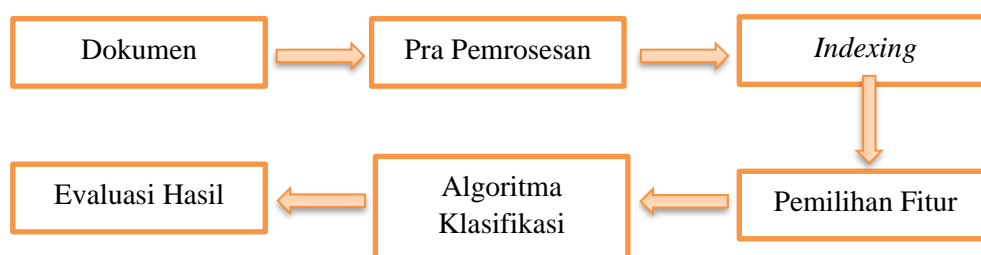
### 2.1 Tahapan Penelitian

Metodologi penelitian yang digunakan adalah metodologi penelitian kualitatif dan kuantitatif. Tahap-tahap metode penelitiannya adalah :

1. Identifikasi masalah  
 Pada tahap ini dilakukan metodologi kualitatif yaitu merumuskan masalah ke beberapa kalimat pertanyaan (*research question*).
2. Studi pustaka  
 Pada tahap ini dilakukan metodologi kualitatif yaitu memahami tentang permasalahan topik dengan mempelajari referensi atau bahan pustaka.
3. Pengolahan dan analisa data  
 Pada tahap ini dilakukan metodologi kuantitatif yaitu mengolah dan menganalisa data yang digunakan pada penelitian.
4. Penyusunan laporan penelitian  
 Pada tahap ini dilakukan metodologi kualitatif yaitu menyusun laporan hasil penelitian yang dikerjakan.
5. Dokumentasi  
 Pada tahap ini dilakukan dokumentasi terhadap penelitian yang dilakukan dengan mengumpulkan data dan informasi kemudian ditelaah secara mendalam. Dokumentasi yang digunakan dalam penelitian ini adalah profil instansi dan PERBUP TUPOKSI Deli Serdang.

### 2.2 Pengklasifikasian Dokumen

Pengklasifikasian teks memiliki banyak metode yaitu metode *K-Nearest Neighbors (KNN)* dan *Chi Square*, *Cosine Similarity*, *Naive Bayes Classifier(NBC)* dan masih banyak lagi . Dalam penelitian ini metode *Naive Bayes* dipilih dalam pengklasifikasian dokumen pada Badan Pengelola Keuangan dan Aset Kab. Deli Serdang[4]. Adapun proses pengklasifikasian teks dapat dilihat pada Gambar 1.



**Gambar 1.** Proses Pengklasifikasian Dokumen

1. Kumpulan Dokumen  
 Merupakan tahap pertama dalam proses pengklasifikasian yang berisi kumpulan dokumen dengan jenis yang berbeda
2. Pra Pemrosesan  
 Pada tahap ini kumpulan dokumen yang ada dipilih beberapa kata yang mewakili representasi isi dari dokumen tersebut. Hal ini bertujuan untuk mengurangi kerumitan dokumen dan membuatnya lebih mudah diklasifikasi.
3. *Indexing* (Pengeindeksan)  
 Pada tahap ini ialah melakukan pembobotan terhadap berapa banyak kemunculan kata dalam dokumen. Dimana setiap kata yang merupakan presentasi dari dokumen dihitung berapa banyak kata tersebut muncul dalam dokumen.
4. Pemilihan Fitur

Setelah tahap pra pemrosesan dan pengindeksan langkah selanjutnya adalah pemilih fitur. Pada tahap ini adalah memilih subset fitur dari dokumen aslinya. Pemilihan fitur dilakukan dengan mencari kata-kata yang memiliki skor kemunculan tertinggi dalam dokumen.

5. Algoritma Klasifikasi  
 Pada tahap ini dipilih algoritma pengklasifikasian dokumen dalam hal ini algoritma yang dipilih adalah algoritma *Naive Bayes Classifier*
6. Evaluasi Hasil  
 Tahap ini adalah tahap terakhir dalam pengklasifikasian dokumen. Evaluasi dilakukan secara ekperimental. Evaluasi hasil adalah tahap dimana mengevaluasi efektivitas pengklasifikasian untuk mengambil keputusan kategorisasi dokumen yang tepat.

## 2.2 Metode Naive Bayes

Rumus *Naive Bayes* adalah rumus yang digunakan sebagai dasar dalam pengklasifikasian teks dengan menggunakan metode *Naive Bayes*. Teorema *Naive Bayes* adalah teorema yang menyatakan peluang kejadian B terjadi dengan syarat kejadian A yang disebut probabilitas bersyarat[5]. Rumus teorema *Bayes* adalah:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(B)P(A|B)}{P(A)} \quad (1)$$

Dimana:

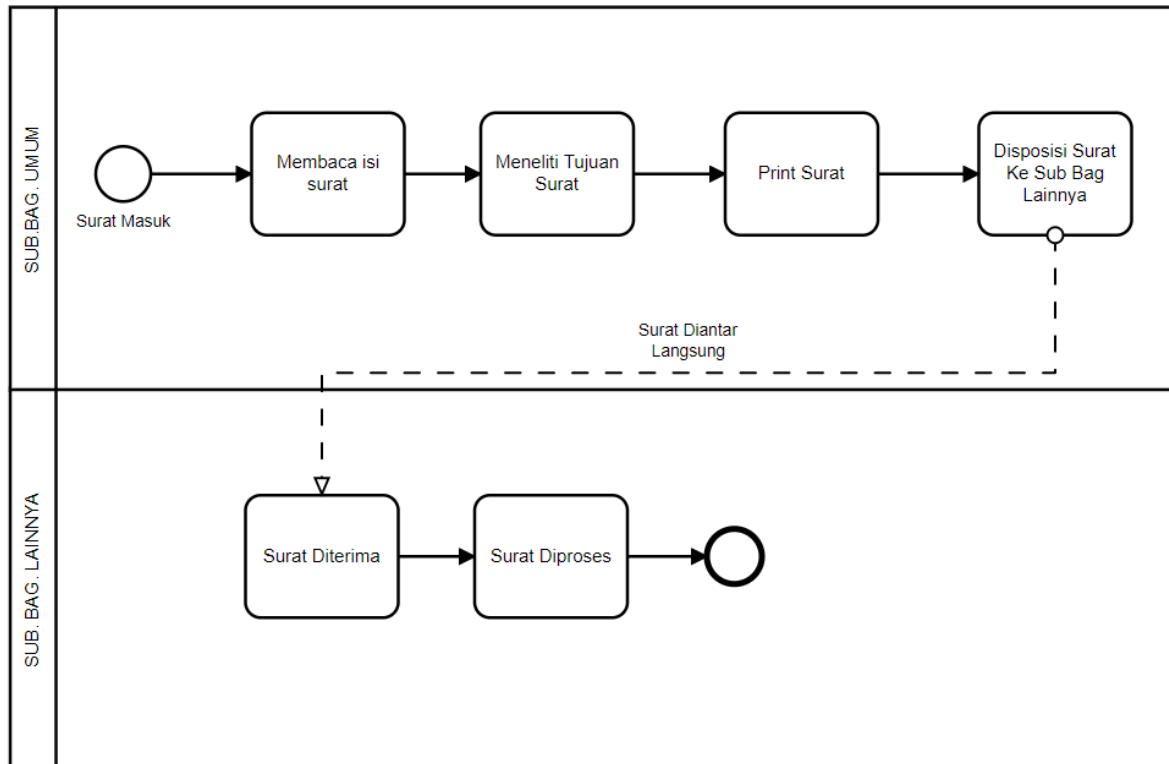
- $P(B|A)$  = peluang kejadian B jika kejadian A terjadi
- $P(A)$  = peluang kejadian A
- $P(A \cap B)$  = peluang kejadian A beririsan dengan kejadian B
- $P(B)$  = peluang kejadian B
- $P(A|B)$  = peluang kejadian A jika kejadian B terjadi

## 3. HASIL DAN PEMBAHASAN

Berdasarkan analisa terhadap proses penelitian yang dilakukan oleh penulis pada Badan Pengelolaan Keuangan dan Aset Kabupaten Deli Serdang terkait prosedur pengklasifikasian dokumen diketahui bahwa proses disposisi yang dilakukan secara manual. Penerima surat dimana Sub Bagian Umum melakukan penelitian isi surat dengan cara membaca keseluruhan isi surat. Setelah meneliti isi surat maka surat tersebut di *print* dan mengantar surat ke Sub Bagian yang dituju sehingga menghabiskan waktu yang cukup lama mengingat banyaknya surat yang di disposisi.

*Naive Bayes Classifier* merupakan sebuah metode klasifikasi dengan teorema *Bayes*. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dr *Naive Bayes Classifier* ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian. *Naive Bayes Classifier* biasa digunakan dalam pengklasifikasian dokumen, membuat diagnosis medis, mendeteksi spam dan lain sebagainya.

Dari permasalahan diatas dapat disimpulkan bahwa Badan Pengelolaan Keuangan dan Aset Kabupaten Deli Serdang sangat membutuhkan suatu sistem pengklasifikasian dokumen. Pada penelitian ini penulis melakukan pengklasifikasian dokumen menggunakan metode *Naive Bayes*. Metode *Naive Bayes* dianggap paling cocok dalam penerapan pengklasifikasian dokumen pada Badan Pengelola Keuangan dan Aset Kabupaten Deli Serdang. Dengan jumlah dokumen yang sangat besar yang ada pada Badan Pengelola Keuangan dan Aset Kab. Deli Serdang klasifikasi dokumen akan sangat membantu dalam disposisi surat ke masing-masing Sub Bidang terkait. *Current System* disposisi surat pada Badan Pengelola Keuangan dan Aset Kab. Deli Serdang dapat dilihat pada Gambar 2.



**Gambar 2.** Current System

Berdasarkan gambar diatas tersebut bahwa Sub Bagian Umum selaku penerima surat akan membaca terlebih dahulu gambaran surat dan ditelaah kemana tujuan surat diberikan. Setelah menentukan tujuan surat, maka Sub Bagian Umum akan mengantarkan surat ke ruangan Sub Bagian yang dituju.

*Naïve Bayes* merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan *teorema Bayes*. Pengklasifikasian dokumen menggunakan algoritma *Naïve Bayes* adalah tahap yang dilakukan oleh penulis untuk menyelesaikan contoh kasus permasalahan pengklasifikasian dokumen pada Badan Pengelola Keuangan dan Aset Kabupaten Deli Serdang. Berdasarkan proses klasifikasi teks yang terdiri dari 6 tahap yaitu:

1. Kumpulan Dokumen

Merupakan tahap pertama dalam proses klasifikasi teks yang berisi kumpulan dokumen dengan jenis berbeda. Dalam hal ini dipilih 3 kategori dokumen yaitu kategori Anggaran, Program dan kategori Kas Daerah. Setiap kategori memiliki beberapa kata yang merepresentasikan isi dari dokumen disebut juga sebagai data training.

**Tabel 1.** Dokumen Latih

No	Anggaran(A)	Program (B)	Kas Daerah (C)
1	kebijakan	umum	tkdd
2	Kertas	barang	dana
3	perubahan	jasa	realisasi
4	rancangan	pengadaan	sp2d
5	anggaran	rencana	pajak
6	Migas	anggaran	ntpn
7	minerba	rancangan	billing
8	Dbh	pelelangan	kas
9	Apbn	kinerja	daerah
10	Apbd	data	penerimaan
11	Dana	laporan	Pengeluaran
12	pendapatan	sakip	Nonfisik
13	belanja	lakip	Laporan
14	realisasi	penilaian	transaksi
15	Hasil	birokrasi	saldo

2. Pra Pemrosesan

Pada tahap ini dipilih sebuah dokumen uji atau disebut dengan *document testing*. Dari keseluruhan isi dokumen maka dipilih beberapa kata yang menggambarkan isi keseluruhan dokumen yang disebut dengan kata kunci ataupun *keyword*. Tujuannya adalah untuk mengurangi kerumitan dokumen dan membuatnya lebih mudah untuk diklasifikasi.

**Tabel 2.** Dokumen uji

No	Doc (D)
1	realisasi
2	tkdd
3	pengeluaran
4	nonfisik
5	sp2d
6	laporan
7	rencana
8	anggaran

3. *Indexing*

Pada tahap ini ialah melakukan pembobotan terhadap berapa banyak kemunculan kata dalam dokumen. Dimana setiap kata yang merupakan presentasi dari dokumen dihitung berapa banyak kata tersebut muncul dalam dokumen.

4. Pemilihan fitur

Setelah tahap pra pemrosesan dan pengindeksan langkah selanjutnya adalah pemiliha fitur. Pada tahap ini adalah memilih subset fitur dari dokumen aslinya. Pemilihan fitur dilakukan dengan mencari kata-kata yang memiliki skor kemunculan tertinggi dalam dokumen.

**Tabel 3.** Kemunculan Kata di Kategori

No	Doc (D)	Doc A (count x,a)	Doc B (count x,b)	Doc C (count x,c)
1	realisasi	1	0	1
2	Tkdd	0	0	0
3	pengeluaran	0	0	1
4	nonfisik	0	0	1
5	sp2d	0	0	0
6	laporan	0	1	1
7	rencana	0	0	0
8	anggaran	0	0	0

5. Algoritma Klasifikasi

Pada tahap ini dipilih algoritma pengklasifikasian dokumen dalam hal ini algoritma yang dipilih adalah algoritma *Naive Bayes Classifier*

6. Evaluasi Hasil

Tahap ini adalah tahap terakhir dalam pengklasifikasian dokumen. Evaluasi dilakukan secara ekperimental. Evaluasi hasil adalah tahap dimana mengevaluasi efektivitas pengklasifikasian untuk mengambil keputusan kategorisasi dokumen yang tepat.

Maka berdasarkan rumus *Naive Bayes* :

$$P(x | c) = \frac{\text{count}(x,c) + 1}{\text{count}(c) + |V|}$$

Dimana :

$$|V| = 41$$

$$\text{count}(c) = 15$$

Maka didapatkan hasil dari  $P(x/c)$  adalah sebagai berikut :

**Tabel 4.** Peluang Kata Dalam Kategori

Doc D	Anggaran(A)	Program (B)	Kas Daerah (C)
realisasi	1,0179	0,0179	1,0179
tkdd	0,0179	0,0179	0,0179
pengeluaran	0,0179	0,0179	1,0179
nonfisik	0,0179	0,0179	1,0179
sp2d	0,0179	0,0179	0,0179
laporan	0,0179	1,0179	1,0179
rencana	0,0179	0,0179	0,0179
anggaran	0,0179	0,0179	0,0179

Diketahui bahwa nilai prior dari setiap kategori adalah sama yaitu  $\frac{1}{3}$  atau sama dengan nilai 0.3333 sehingga :

$$\bar{P}(\text{anggaran/doc}_d) \propto 0.3333 \times 1.0179 \times (0.0179)^7 \approx 1,96449E-13$$

$$\overline{P}(\text{program/doc}_d) \propto 0.3333 \times 1.0179 \times (0.0179)^7 \approx 1,96449\text{E-}13$$

$$\overline{P}(\text{kasdaerah/doc}_d) \propto 0.3333 \times (1.0179)^4 \times (0.0179)^3 \approx 3,63809\text{E-}08$$

Dari perhitungan di atas maka didapatkan hasil klasifikasi Dokumen D merupakan dokumen dengan kategori yang paling mendekati adalah **Kategori Kas Daerah**.

#### 4. KESIMPULAN

Hasil dari penelitian dapat disimpulkan pengklasifikasian dokumen berdasarkan metode *Naive Bayes* merupakan algoritma yang cocok digunakan dalam disposisi surat masuk pada Badan Pengelola Keuangan dan Aset Kab. Deli Serdang. Aplikasi yang dikembangkan dapat mengklasifikasikan dokumen dalam pendisposisian surat di Badan Pengelola Keuangan dan Aset Kab. Deli Serdang sesuai dengan kategori dan bidang dimana surat akan di disposisikan.

#### REFERENCES

- [1] R. Efendi and R. F. Malik, "Klasifikasi dokumen berbahasa indonesia menggunakan naive bayes classifier," vol. I, no. I, pp. 7–13, 2012.
- [2] A. Pandhu and H. Agus, "Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government," *J. Appl. Intell. Syst.*, vol. 1, no. 1, pp. 48–55–55, 2016.
- [3] A. W. Widodo, "Klasifikasi Artikel Berita Menggunakan Naive Bayes Classifier yang Dimodifikasi."
- [4] V. Muntihana, J. T. Informatika, F. Sains, and D. A. N. Teknologi, *Berbasis Web Dan Android Pada Klinik Gigi Lisda*. 2017.
- [5] B. A. B. Ii and K. Pustaka, "Dicetak pada tanggal 2020-03-24 Id Doc: 589c885881944dbf0f494625," pp. 7–15, 2020, [Online]. Available: <http://repository.fkip.unja.ac.id/file?i=3NzOaxySsy641EUFTZwd9Z-g5mrddu-fZGhE6QEaC8c>.
- [6] V. Korde, "Text Classification and Classifiers:A Survey," *Int. J. Artif. Intell. Appl.*, vol. 3, no. 2, pp. 85–99, 2012, doi: 10.5121/ijaia.2012.3208.
- [7] D. P. B. Bestari, R. Saptono, and R. Anggrainingsih, "Academic Articles Classification Using Naive Bayes Classifier (Nbc) Method," *J. Ilm. Teknol. dan Inf.*, vol. 7, no. 2, pp. 74–81, 2018.