

Pengembangan Sistem Rekomendasi Kategori Cerita Pendek Menggunakan Naïve Bayes Classifier

Faisal Rahutomo¹, Arie Rachmad Syulistyo², Mafruroh³

^{1,2,3}Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang
¹faisal@polinema.ac.id, ²arie.rachmad.s@polinema.ac.id, ³mafrurohh@gmail.com

Abstrak— Cerpen atau cerita pendek merupakan cerita yang digemari oleh pembaca di semua kalangan dikarenakan cerita yang disajikan tidak bertele-tele dan membutuhkan waktu singkat untuk membacanya, tetapi sering kali pembaca kebingungan akan memilih cerpen yang sesuai kategorinya. Bahkan terdapat lebih dari satu kategori di dalam artikel tersebut yang semakin membingungkan pembaca memilih cerpen yang sesuai dengan kategorinya. Oleh sebab itu dalam sistem ini cerpen diklasifikasikan sesuai kategorinya menggunakan metode Naïve Bayes Classifier dan hasil klasifikasi berupa angka diubah menjadi text menggunakan irisan himpunan bagian. Berdasarkan pengujian yang sudah peneliti lakukan tingkat akurasi dari pengembangan rekomendasi cerita pendek menggunakan Naïve Bayes Classifier menghasilkan tingkat akurasi 68%. Untuk pengujian yang dilakukan menggunakan cross validation hasil akurasinya sebanyak kelompok 1 menghasilkan nilai akurasi 44%, kelompok 2 menghasilkan nilai akurasi 54%, kelompok 3 menghasilkan nilai akurasi sebanyak 61% dan kelompok 4 menghasilkan nilai akurasi sebanyak 68%.

Kata kunci— klasifikasi cerpen, naïve bayes classifier, text mining, cerita pendek

I. PENDAHULUAN

Cerita pendek saat ini banyak digemari oleh pembaca mulai dari cerita anak-anak, remaja hingga dewasa karena konsep yang diusung oleh penulis. Tetapi pada proses pemilihan cerita pendek ini masih kurang diimbangi dengan sistem yang dapat melakukan filterisasi atau klasifikasi pada cerpen itu sendiri.

Untuk dapat menentukan sebuah cerpen masuk kedalam kategori cerpen tertentu bukanlah hal yang mudah, sudah tentu orang harus membaca keseluruhan atau minimal sebagian isi dari cerpen tersebut kemudian barulah dapat diketahui cerpen tersebut masuk kedalam kategori apa. Dan terdapat juga cerpen yang memiliki lebih dari satu kategori di satu cerita. Hal ini yang menjadikan kesulitan dalam menentukan sebuah cerpen masuk kedalam kategori tertentu, sedangkan terkadang banyak orang yang tidak bisa membaca terlebih dahulu isi dari cerpen tersebut. Permasalahan kadang terjadi pada orang yang ingin membaca cerpen tetapi tidak tahu cerpen tersebut ada dikategori mana. Apalagi terdapat permasalahan mengenai satu artikel terdapat lebih dari satu kategori sehingga pembaca sulit untuk memilih cerpen yang sesuai kategori. Berdasarkan permasalahan tersebut maka perlu sebuah solusi yang dapat mengatasinya sehingga dapat dijadikan sebagai acuan dalam memilih sebuah cerpen.

Sistem Rekomendasi Kategori Cerita Pendek Menggunakan Naïve Bayes Classifier ini diajukan penulis untuk memecahkan permasalahan diatas mengenai cerpen yang memiliki lebih dari satu kategori di satu cerita menggunakan metode Naïve Bayes Classifier dengan perhitungan himpunan bagian. Pemecahan masalah tersebut tidak bisa terselesaikan jika menggunakan metode Naïve Bayes Classifier sehingga membutuhkan perhitungan himpunan bagian.

II. LANDASAN TEORI

Penelitian terkait dengan klasifikasi teks dengan menggunakan *Naïve Bayes Classifier* sudah dilakukan oleh peneliti sebelumnya, seperti yang dilakukan oleh Oman Sumantri mengenai hasil eksperimen yang dilakukan tentang klasifikasi kategori cerita pendek terhadap model yang diusulkan yaitu *Naïve Bayes* [1].

Peneliti yang terkait dengan klasifikasi teks dengan menggunakan *Naïve Bayes Classifier* sudah dilakukan oleh peneliti sebelumnya yaitu oleh peneliti Dwi Puspitasari mengenai analisa sistem klasifikasi judul skripsi [2].

Penelitian terkait juga dilakukan oleh Moch Fadli Shadiqin mengenai mengkategorikan lirik lagu ke dalam empat kelompok usia, yaitu anak, remaja, dewasa dan semua usia. Algoritma yang digunakan adalah *Naïve Bayes Classifier* dengan jumlah dataset lirik sebanyak 400dataset judul [3].

Penelitian lain yang terkait dengan klasifikasi text Bahasa Indonesia dilakukan oleh Darujati dimana dari percobaan yang telah dilakukan menerapkan metode *Naive Bayes Classifier* dalam klasifikasi berita dengan data latih yang besar sebanyak 100artikel [4].

Peneliti lain yang berkaitan dengan klasifikasi menggunakan Naïve Bayes adalah mendeteksi berita hoax berbahasa Indonesia oleh Faisal Rahutomo dimana peneliti ini melakukan experiment terhadap berita dalam jaringan bahasa Indonesia yang valid dan hoax [5].

Peneliti terkait juga dilakukan oleh Erik Hardiyanto mengenai klasifikasi artikel Wikipedia Bahasa Indonesia secara otomatis dengan menggunakan metode *K Nearest Neighbor* [6].

Dari penelitian yang dilakukan sebelumnya ialah melakukan klasifikasi teks hanya pada kelompok yang terpisah sedangkan pada penelitian makalah ini berbeda dari peneliti sebelumnya yaitu dimana melakukan pengklasifikasian teks yang saling

beririsan. Dan pada penelitian yang dilakukan oleh penulis mengenai pengembangan sistem rekomendasi kategori cerita pendek menggunakan Metode *Naive Bayes Classifier*.

A. Klasifikasi

Klasifikasi teks merupakan sebuah teknik dalam text mining yang bertujuan untuk menempatkan teks pada kategori yang sesuai dengan karakteristiknya dari teks tersebut dengan menggunakan aturan-aturan tertentu. Dengan adanya klasifikasi teks, maka dapat memberikan pandangan secara konseptual mengenai cara pengelompokan data yang memiliki peranan penting terhadap dunia nyata [7].

Tujuan dari pengkategorian teks adalah untuk mengklasifikasikan data ke dalam beberapa kategori-kategori tertentu. Tiap dokumen dapat diklasifikasikan dalam beberapa kategori, atau tidak sama sekali. Dengan menggunakan machine learning, pembelajaran dapat dilakukan dengan aturan klasifikasi yang sudah ditentukan dan adanya data latih sebagai acuan pembelajaran sehingga dapat melakukan proses pengklasifikasian secara otomatis nantinya.

B. Text Mining

Text mining merupakan suatu proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan data menggunakan tools analisis yang merupakan kompone-komponen dalam data mining yang salah satunya adalah kategorisasi. Tujuan utama dari text mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan data. Sumber data yang digunakan pada text mining adalah kumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur. Manfaat dari text mining adalah untuk mencari pola dalam teks, yaitu proses penganalisisan teks guna menyarikan informasi yang bermanfaat untuk tujuan tertentu [8].

Secara umum proses dalam text mining dilakukan dalam tiga tahapan yakni pemilihan data sebagai sumber informasi, kemudian dilakukan preprocessing dan analisis, dan yang terakhir adalah hasil dari proses yang menunjukkan interpretasi dari data yang diuji [2].

C. Pre-processing

Tindakan yang dilakukan adalah Case folding merupakan tahapan untuk mengubah semua huruf dalam dokumen menjadi huruf kecil dan Tokenizing digunakan untuk membuang beberapa karakter yang dianggap sebagai tanda baca. Pada tahap Tokenizing penulis menggunakan library PHP-Machine Learning yang berada di laman <https://php-ml.readthedocs.io>. Serta tahap Filtering untuk mengambil katakata penting dari hasil Tokenizing [9]. Stopword digunakan untuk kosakata yang bukan merupakan kata unik dari suatu data. Data yang tidak penting atau kata imbuhan akan dihilangkan dengan stopwords ini [10].

D. Naive Bayes Classifier

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Pengklasifikasian menggunakan *Naive Bayes Classifier* dilakukan proses training dan testing [3].

Bayes sendiri untuk teknik prediksi probabilitas yang berdasar pada penerapan teorema bayes dengan ketidaktergantungan yang kuat (naif) [11]. Yang digunakan oleh Naive Bayes adalah model fitur independen.

$$P(C_j|W_i) = \frac{P(W_i|C_j).P(C_j)}{P(W_i)} \quad (1)$$

Keterangan:

$P(C_j|W_i)$: Peluang kategori j, ketika terdapat kemunculan kata i

$P(W_i|C_j)$: Peluang kata i masuk kedalam kategori j

$P(C_j)$: Peluang kemunculan kategori j

$P(W_i)$: Peluang kemunculan kata

Ada banyak cara untuk menghitung $P(W_i|C_j)$, cara paling sederhana adalah dengan

$$P(W_i|C_j) = \frac{N_{ic}+1}{N_c+V} \quad (2)$$

Keterangan:

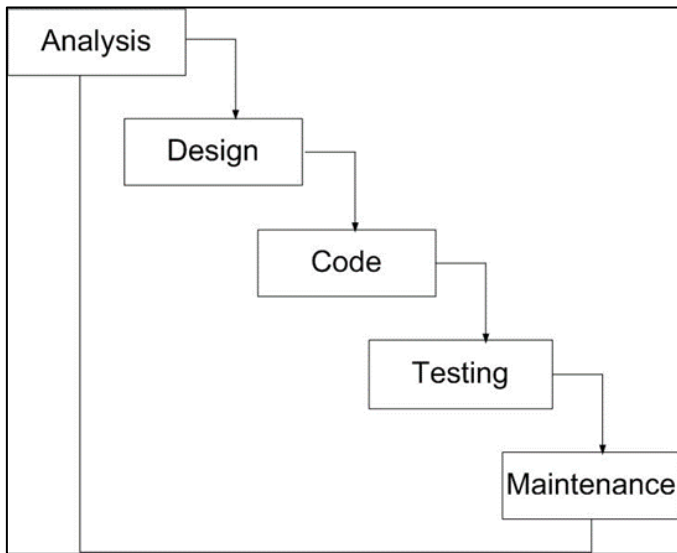
N_{ic} : Jumlah dokumen latih dengan fitur atribut W_i dengan kategori C_j

N_c : Jumlah dokumen dari kategori C_j

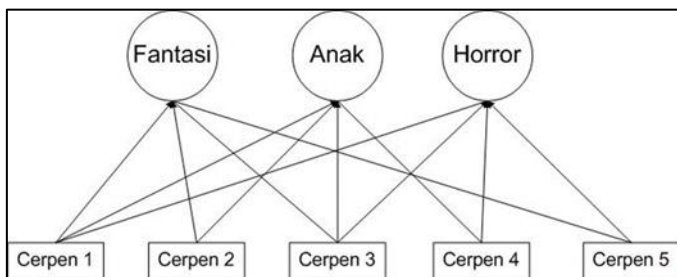
V : Jumlah kategori

III. METODELOGI PENELITIAN

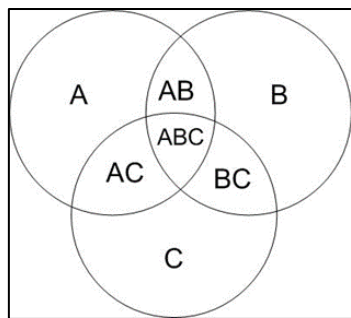
Metode Waterfall digunakan dalam penelitian ini karena setiap proses dalam pengembangan sistem dilakukan secara berurutan dan evaluasi dilakukan pertahap, mulai dari analisa kebutuhan data untuk melengkapi data yang diperlukan dalam pengembangan aplikasi sehingga dapat menghasilkan aplikasi yang baik. Pada penelitian ini, akan digunakan data cerpen dengan jumlah 400 judul dari website cerpenmu.com. Kemudian dilakukan tahap preprocessing pada data tersebut untuk mengekstrak kata-katanya dan didapatkan nilai term frequency dari setiap kata untuk diproses menggunakan metode naive bayes classifier. Proses kedua perancangan sistem untuk membangun implementasi yang menggambarkan bagaimana suatu sistem dibentuk yang berupa penggambaran, perencanaan dan pembuatan sketsa atau pengaturan dari beberapa elemen yang terpisah ke dalam satu kesatuan yang utuh dan berfungsi, menyangkut di dalamnya konfigurasi komponen perangkat lunak dan perangkat keras dari suatu sistem. Ketiga code menerjemahkan hasil proses perancangan menjadi sebuah bentuk program komputer yang dimengerti oleh mesin komputer. Keempat proses testing yaitu untuk mengecek fungsi setiap fitur yang terdapat pada sistem agar setiap fitur yang ada bekerja sesuai perancangan . Terakhir proses maintenance yang bertujuan evaluasi dan pengembangan penelitian serupa dengan performa yang lebih baik. Dapat dilihat pada Gambar 1.



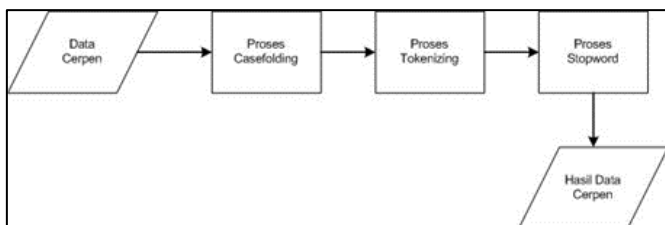
Gambar 1. Tahapan Pengembangan Sistem



Gambar 2. Artikel lebih dari satu kategori



Gambar 3. Model irisan



Gambar 4. Proses preprocessing

A. Pengambilan Data

Perencanaan dan pengumpulan data yang akan diolah adalah berupa data text, yang didapat dari pemilihan data di website cerpenmu.com. Data cerpen yang diambil adalah 3 kategori yaitu, kategori horror, kategori fantasi dan kategori anak dengan total 400 data cerpen. Dari total cerpen yang

diperoleh, akan diolah menggunakan *Naïve Bayes Classifier* dan Himpunan Bagian untuk memecah kategori. Kemudian, diimplementasikan ke sistem berdasarkan hasil dari perancangan. Setelah itu, melakukan pengujian sistem terhadap sistem dengan menganalisis hasil keluaran sistem, apakah sistem tersebut sudah baik atau masih membutuhkan perbaikan. Dapat dilihat pada rancangan Gambar 2 bahwa disetiap satu artikel terdapat lebih dari satu kategori.

B. Pengolahan Data

Data sejumlah 400 cerpen yang telah didapatkan, perlu dilakukan pengolahan agar menjadi data yang mudah digunakan dalam proses klasifikasi. Cerpen akan melalui tahap Irisan himpunan bagian untuk memecah kategori dan proses preprocessing agar menjadi lebih ringkas. Setelah melalui preprocessing, data yang berupa teks akan diubah ke dalam bentuk angka melalui perhitungan tf. Nilai tf ini menjadi nilai matriks yang akan berlaku sebagai masukan untuk algoritma *Naïve Bayes Classifier*.

1) Irisan Himpunan

Pada irisan himpunan bagian ini digunakan untuk memecah kategori dimana di satu artikel atau cerpen memiliki lebih dari satu kategori sehingga dibutuhkan irisan himpunan seperti di Gambar 3 dengan menggunakan persamaan 3.

$$2^n - 1 = 2^3 - 1 = 8 - 1 = 7 \quad (3)$$

Himpunan: {a,b,c} = 3

Himpunan Bagian : {}, {a}{b}{c}, {a,b}{a,c}{b,c}, {a,b,c}

Banyaknya himpunan bagian: 8=23

Jadi untuk memisahkan satu artikel menjadi satu kategori maka diperlukan perhitungan himpunan bagian dan dibuat kelas turunan seperti diatas untuk memecah setiap kategori yang berada di satu artikel.

2) Preprocessing

Pada tahapan ini teks akan diolah dan menghasilkan numeric yang nantinya akan dimasukkan di algoritma *Naïve Bayes*. Tahapannya seperti pada Gambar 4.

TABEL 1. CERPEN YANG MELALUI PREPROCESSING

No	Preprocessing	Hasil
1	Cerpen Awal	Malam ini adalah malam yang menentukan segalanya Aku berdiri di bawah gempuran bola api di Liendholm
2	Casefolding	malam ini adalah malam yang menentukan segalanya aku berdiri di bawah gempuran bola api di liendholm
3	Tokenizing	malam ini adalah malam yang

		menentukan segalanya aku berdiri di bawah gempuran bola api di liendholm
4	Stopword	malam adalah malam menentukan segalanya aku berdiri bawah gempuran bola api liendholm

Proses TF yang berdasar pada Tabel 1 hanya dituliskan beberapa kata dari cerpen sebagai contoh perhitungan TF menggunakan preprocessing. Hasil proses TF termuat pada Tabel 2.

TABEL 2. TERM FREQUENCY

	D1	D2	D3	D4	D5	D6
Malam	1	0	0	0	0	0
Pertempuran	1	0	0	0	0	0
senjata	1	0	0	0	0	0
Mata	0	1	0	0	0	0
Aneh	0	1	0	0	0	0
suara	0	1	0	0	0	0
Faisal	0	0	6	0	0	0
kaget	0	0	1	0	0	0
sahabat	0	0	1	0	0	0
Anak	0	0	0	1	0	0
Sosok	0	0	0	1	0	0
kecil	0	0	0	1	0	0
terdiam	0	0	0	0	1	0
Pojok	0	0	0	0	1	0
Mulut	0	0	0	0	1	0
Hari	0	0	0	0	0	3
Buruk	0	0	0	0	0	1
Perempuan	0	0	0	0	0	1
KATEGORI	F	F	A	A	H	H

3) Algoritma Naïve Bayes

Proses training digunakan untuk menghasilkan model klasifikasi yang nantinya akan digunakan sebagai pedoman dalam klasifikasi. Berikut adalah proses training pada algoritma NBC. Pada proses training, algoritma NBC menggunakan persamaan 1 bertujuan untuk menghitung bobot atau nilai probabilitas setiap kata dalam data training di setiap kategori klasifikasi. Contoh hasil proses training menggunakan hasil dari perhitungan TF termuat pada Tabel 3.

TABEL 3. HASIL PERHITUNGAN DATA TRAINING

Kata	PWord Kategori		
	Fantasi	Anak	Horror
D1	0.12	0	0
D2	0.12	0	0
D3	0	0.32	0
D4	0	0.12	0
D5	0	0	0.12
D6	0	0	0.2
Hasil	0.0144	0.0384	0.024

Dari hasil diatas, terlihat bahwa nilai probabilitas tertinggi ada pada class atau kategori Anak sehingga dapat disimpulkan bahwa hasil klasifikasi yang telah diuji diatas adalah masuk kedalam kategori anak.

4) Pengujian

a) Precision

Precision merupakan nilai kakuraan dari hasil pengklasifikasian seluruh dokumen oleh system, nilai precision dihitung dengan menggunakan persamaan 4 berikut: [3].

$$precision = \frac{\sum data\ relevan\ dan\ terambil}{\sum seluruh\ data\ terambil} \times 100\% \quad (4)$$

b) Recall

Recall merupakan tingkat kesuksesan dari system untuk mengenali sebuah kategori, nilai dari recall dapat dihitung menggunakan persamaan 5 berikut: [3].

$$recall = \frac{\sum data\ relevan\ dan\ terambil}{\sum seluruh\ data\ relevan} \times 100\% \quad (5)$$

c) F-Measure

F-Measure adalah sebuah gambaran dalam pengaruh relative antara precision dan recall. Untuk menghitung nilai F-Measure menggunakan persamaan 6 berikut: [3].

$$F - measure = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (6)$$

d) Accuracy

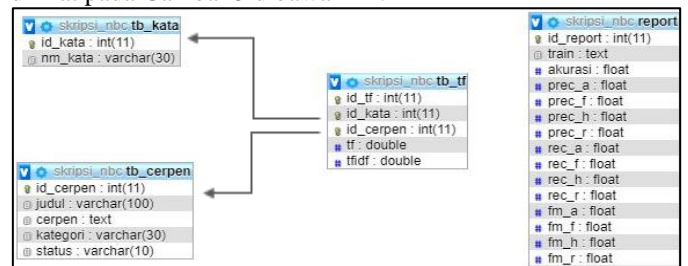
Jadi untuk nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data dengan persamaan 7 berikut: [11].

$$Accuracy = \frac{\sum klasifikasi\ benar}{\sum data\ uji\ coba} \quad (7)$$

IV. IMPLEMENTASI

A. Implementasi Database

Implementasi database untuk sistem yang dibangun ini memiliki 4 tabel berdasarkan perancangan. Database dapat dilihat pada Gambar 5 dibawah ini.



Gambar 5. Physical Data Model

Database yang dibangun memiliki tabel-tabel sebagai berikut:

- a) Tb_cerpen: Tabel cerpen ini berisi data cerpen yang nantinya akan diolah melalui tahap preprocessing. Dalam tb_cerpen ini memiliki kolom id_cerpen, judul, isi cerpen, kategori dan status.
- b) Tb_kata: Tabel kata ini berisi kata-kata yang sudah melalui tahap preprocessing dan hasil dari tokenizing akan masuk ke kolom nm_kata yang nantinya akan diperhitungkan di tf.
- c) Tb_tf: Tabel tf ini berisi hasil tf yang diambil dari tabel kata dan dilakukan pembobotan yang nantinya akan ditampilkan di tfidf
- d) Tb_report: Tabel report berisi hasil dari pengujian dataset dan menampilkan hasil presisi, recall dan akurasinya.

B. Implementasi Diagram

1) Class Diagram

Class Diagram ini adalah model statis yang menggambarkan struktur dan deskripsi serta hubungan antar class yang terdiri dari nama class yang terdiri dari nama class, atribut, dan methods. Class diagram yang dibangun seperti Gambar 6.

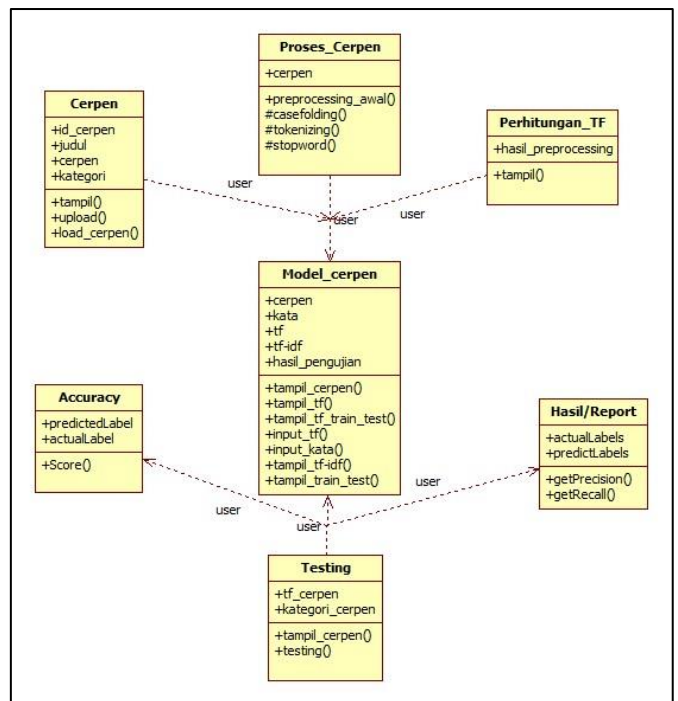
Class diagram yang dibangun memiliki alur class sebagai berikut:

- a) Class cerpen mempunyai relasi terhadap proses_cerpen yang artinya cerpen harus diproses terlebih dahulu oleh class proses_cerpen yang nantinya diolah di model_cerpen.
- b) Class proses_cerpen mempunyai relasi terhadap class cerpen dimana proses_cerpen ini melakukan preprocessing data dari class cerpen dan hasil dari preprocessing akan direlasikan ke model_cerpen.
- c) Perhitungan_tf menampilkan hasil perhitungan preprocessing dari proses_cerpen yang nantinya akan direlasikan ke model_cerpen.
- d) Model_cerpen memiliki relasi terhadap class cerpen, proses_cerpen serta tf yang artinya semua class ini akan diolah di model_cerpen.
- e) Class testing memiliki relasi terhadap class hasil dan akurasi yang artinya melatih data cerpen dan hasilnya akan tampil di class hasil maupun akurasi.
- f) Class hasil akan menampilkan hasil prediksi dari hasil testing yang menampilkan nilai presisi dan recall.
- g) Class akurasi menampilkan prediksi dari data yang sudah dilatih beserta hasil scorenya.

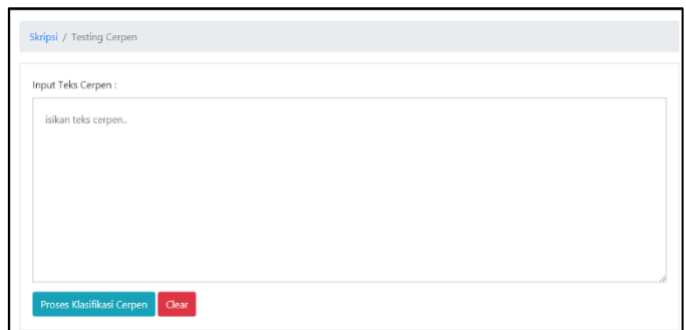
C. Implementasi Sistem

Implementasi sistem untuk sistem Rekomendasi Kategori Cerita Pendek ialah menampilkan tampilan inputan cerpen kedalam sistem dan dilakukannya proses klasifikasi cerpen. Dapat dilihat pada Gambar 7.

Setelah dlakukannya proses klasifikasi cerpen maka akan muncul hasil seperti Gambar 8 dibawah ini.



Gambar 6. Class Diagram



Gambar 7. Tampilan input cerpen



Gambar 8. Hasil rekomendasi cerpen

V. HASIL DAN PEMBAHASAN

Proses pengujian juga dilakukan menggunakan cross validation seperti dibawah ini. Pengujian sistem diperlukan untuk menghitung tingkat akurasi dari keberhasilan sistem.

TABEL 3. PENGUJIAN CROSS VALIDATION 4-FOLD

K	Train	Test	P	R	F	ACC
1	A,B,C	D	51%	70%	56%	44%
2	A,B,D	C	58%	69%	62%	54%
3	A,C,D	B	66%	80%	60%	61%
4	C,D,B	A	68%	76%	64%	69%
Nilai Rata-Rata						57%

Dapat dilihat tingkat akurasi paling tinggi terdapat pada kelompok 4 dengan nilai akurasi 69% dikarenakan kandungan term cerpen pada kelompok 4 lebih baik daripada kelompok 1, 2 dan 3.

Pada Gambar 12 di bawah ini adalah hasil grafik dengan pengujian 4-Fold Validation dimana akurasi semakin membaik di tiap pengujian selanjutnya.

VI. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan pembahasan dan hasil pengujian yang telah dilakukan mengenai klasifikasi cerita pendek menggunakan metode *Naïve Bayes Classifier* penulis berharap setelah sistem memasuki tahap akhir akan memperoleh kesimpulan sebagai berikut:

1. Metode *Naïve Bayes Classifier* dapat diterapkan sebagai metode untuk melakukan klasifikasi cerita pendek dengan hasil nilai akurasinya sebanyak 68%.
2. Pengujian dengan menggunakan jumlah data training yang berbeda menghasilkan nilai akurasi yang berbeda dengan kecenderungan akurasi yang semakin meningkat. Hasil akurasi klasifikasi algoritma *Naïve Bayes Classifier* menggunakan pengujian *cross validation K-fold* dengan data yang merata menghasilkan nilai akurasi sebesar 44%, 54%, 61% dan 68%.

B. Saran

Penelitian mengenai pengembangan sistem rekomendasi kategori cerita pendek dapat dikembangkan dengan beberapa saran sebagai berikut:

1. Mencoba membandingkan dengan metode lain.
2. Penggunaan fitur dan tahap *preprocessing* yang berbeda untuk membandingkan tingkat akurasinya.

DAFTAR PUSTAKA

- [1] O. Somantri, "Text Mining Untuk Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes (NB)," *J. Telemat.*, 2017.
- [2] D. Puspitasari *et al.*, "Analisa sistem klasifikasi judul skripsi menggunakan metode naïve bayes classifier," pp. 43–45, 2013.
- [3] M. F. S. Thirafi and F. Rahutomo, "Implementation of Naïve Bayes Classifier Algorithm to Categorize Indonesian Song Lyrics Based on Age," in *3rd International Conference on Sustainable Information Engineering and Technology, SIET 2018 - Proceedings*, 2019.
- [4] C. Darujati *et al.*, "Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa Indonesia," *J. LINK*, 2012.

- [5] F. Rahutomo, I. Yanuar, R. Pratiwi, and D. M. Ramadhani, "Eksperimen Naïve Bayes Pada Deteksi Berita Hoax Berbahasa Indonesia Naïve Bayes ' s Experiment On Hoax News Detection In Indonesian Language," vol. 23, no. 1, pp. 1–15, 2019.
- [6] F. Rahutomo and E. Hardiyanto, "Studi Awal Klasifikasi Artikel Wikipedia Bahasa Indonesia Dengan Menggunakan Metoda K Nearest Neighbor," vol. 9035, pp. 334–338, 2016.
- [7] B. Sriram, D. Fuhry, E. Demir, and H. Ferhatosmanoglu, "Short Text Classification in Twitter to Improve Information Filtering," pp. 4–5, 2010.
- [8] F. Rahutomo, Z. Hanif, R. Adi, and I. F. Rozi, "Implementasi Text Mining Pada Laman Blog di," pp. 101–109, 2018.
- [9] M. A. Fauzi and T. Afirianto, "Sentiment Analysis Pada Review Aplikasi Mobile Menggunakan Sentiment Analysis Pada Review Aplikasi Mobile Menggunakan Metode Naïve Bayes Dan Query Expansion," no. December, 2016.
- [10] B. Kurniawan, S. Effendi, and O. S. Sitompul, "Klasifikasi Konten Berita Dengan Metode Text Mining," vol. 1, no. 1, pp. 14–19, 2012.
- [11] I. F. Rozi, E. N. Hamdana, M. Balya, and I. Alfahmi, "Pengembangan Aplikasi Analisis Sentimen Twitter (Studi Kasus SAMSAT Kota Malang)," pp. 149–154.