

**SKRIPSI**  
**KLASIFIKASI GENRE CERITA PENDEK BERBAHASA INDONESIA**  
**MENGGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER***



**Disusun Oleh :**

**PETRA WAR A NANDAYANI R**

**DBC 116 058**

**JURUSAN TEKNIK INFORMATIKA**  
**FAKULTAS TEKNIK**  
**UNIVERSITAS PALANGKA RAYA**  
**2020**

**KLASIFIKASI GENRE CERITA PENDEK BERBAHASA INDONESIA  
MENGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER***

**SKRIPSI**

Sebagai salah satu syarat untuk menyelesaikan Program Strata-I pada Jurusan Teknik  
Informatika Fakultas Teknik Universitas Palangka Raya

Oleh

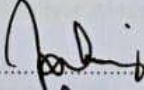



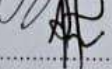
**PETRA WAR A NANDAYANI R**

**DBC 116 058**

Telah dipertahankan didepan tim penguji, pada :

Hari/Tanggal : Jumat, 16 Oktober 2020

Waktu : 13.00 – 14.30 WIB

- |   |   |  |                 |
|---|---|--|-----------------|
| 1. Drs. JADIAMAN PARHUSIP, M.Kom<br>NIP. 19630423 198502 1 001        | : |  | ..... (Ketua)   |
| 2. ADE CHANDRA SAPUTRA, S.Kom., M.Cs<br>NIP. 19870203 201404 1 001    | : |  | ..... (Anggota) |
| 3. NAHUMI NUGRAHANINGSIH, Ph.D<br>NIP. 19791009 200801 2 016          | : |  | ..... (Anggota) |
| 4. SHERLY CHRISTINA, S.Kom., M.Kom<br>NIP. 19810929 200604 2 001      | : |  | ..... (Anggota) |
| 5. ARIESTA LESTARI, S.Kom., M.Cs., Ph.D<br>NIP. 19800322 200501 2 004 | : |  | ..... (Anggota) |

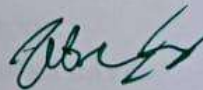
Mengetahui :

Fakultas Teknik  
Universitas Palangka Raya



**Ir. WALUYO NUSWANTORO, M.T.**  
NIP. 19651119 199302 1 001

Jurusan / Program Studi Teknik Informatika  
Fakultas Teknik Universitas Palangka Raya  
Ketua Jurusan,



**ABERTUN SAGIT SAHAY, S.T., M.Eng**  
NIP. 19751212 200312 1 002

**SKRIPSI**

**KLASIFIKASI GENRE CERITA PENDEK BERBAHASA INDONESIA  
MENGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER***

Sebagai salah satu syarat menyelesaikan Program Strata - 1  
pada Jurusan Teknik Informatika Fakultas Teknik Universitas Palangka Raya

**OLEH :**

**PETRA WARA NANDAYANI R**

**NIM. DBC 116 058**

Disetujui untuk diajukan dalam Seminar Akhir Skripsi,

Palangka Raya, Oktober 2020

Pembimbing I

  
**SHERLY CHRISTINA, S.Kom., M.Kom**  
**NIP. 19810929 200604 2 001**

Pembimbing II

  
**ARIESTA LESTARI, S.Kom., M.Kom., Ph.D**  
**NIP. 19800322 200501 2 004**

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS PALANGKA RAYA**

**2020**

iii

## PERNYATAAN

Dengan ini saya menyatakan dengan sebenar - benarnya bahwa dalam Skripsi ini tidak terdapat karya ilmiah yang pernah diajukan untuk memperoleh gelar kesarjanaan disuatu Perguruan Tinggi, serta tidak terdapat karya ilmiah atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali secara tertulis diacu dalam Skripsi ini dan disebutkan dalam Tinjauan Pustaka.

Palangka Raya, Oktober 2020



**Petra Wara Nandayani R**  
DBC 116 058

## RIWAYAT PENYUSUN

### Data Diri

Nama : Petra Wara Nandayani R  
NIM : DBC 116 058  
Fakultas : Teknik  
Jurusan/Program Studi : Teknik Informatika  
Jenjang : Strata 1 ( S-1 )  
Jenis Kelamin : Perempuan  
Tempat, Tanggal Lahir : Warukin, 8 Mei 1999  
Agama : Kristen Protestan  
Status dalam Keluarga : Anak Kandung  
Anak ke - : I (Pertama)  
Alamat : Jl. Dahlia No. 62 PCPR I  
No. Telpon/HP : +6282270249346



### Data Orang Tua

Nama Ayah : Pdt. Yan Ariyus Ramli, S. Th  
Pekerjaan Ayah : Pendeta  
Nama Ibu : Enni Tri Martina, S. Hut  
Pekerjaan Ibu : Guru (Non PNS)  
Alamat Orang Tua : Warukin, RT 01 No. 17 Kec. Tanta, Kab. Tabalong  
No. Telpon/HP : +6281351833325

### Riwayat Pendidikan \*)

SD : SDN Warukin (Tahun Lulus 2010)  
SMP : SMPN 2 Tanta (Tahun Lulus 2013)  
SMA : SMAN 2 Tanjung (Tahun Lulus 2016)

Palangka Raya, Oktober 2020

**Petra Wara Nandayani R**  
**DBC 116 058**

Keterangan:

\*) Nama, Tempat, Tahun Lulus

## HALAMAN PERSEMBAHAN

*Janganlah hendaknya kamu kuatir tentang apapun juga, tetapi nyatakanlah dalam segala hal keinginanmu kepada Allah dalam doa dan permohonan dengan ucapan*

*syukur.*

***Filipi 4 : 6***

Syukur dan terima kasih selalu saya ucapkan kepada Tuhan Yesus Kristus untuk berkat, kasih karunia, dan penyertaan-Nya didalam hidup saya, sehingga saya dapat menyelesaikan Skripsi, baik laporan maupun program, dari awal semester hingga saat ini. Skripsi ini saya persembahkan untuk orang-orang yang telah mendukung saya.

Terima kasih untuk Mamah, Papah, dan Dede Nando yang saya cintai dan sayangi didalam Tuhan Yesus Kristus, yang selalu percaya dan mendoakan, menjadi orang-orang selalu merasa bangga untuk setiap hal sudah saya capai sekecil apapun itu, selalu mendukung dan memberikan kasih sayang dan perhatian kepada saya, serta selalu mengajarkan untuk menyerahkan segala sesuatu yang dikerjakan ke dalam pertolongan tangan kasih Tuhan. Terima kasih kepada Kai dan Mendiang Nenek yang sudah menerima dan mengurus saya selama empat tahun selama di Palangka Raya, walaupun Nenek tidak sempat melihat saya menyelesaikan studi.

Tidak lupa saya berterima kasih pada Ibu Sherly Christina, S.Kom., M.Kom dan Ibu Ariesta Lestari, S.Kom., M.Cs., Ph.D yang telah membimbing dan membantu saya dalam menyelesaikan Skripsi. Serta kepada Pak Drs. Jadianan Parhusip, M.Kom,

Pak Ade Chandra Saputra, S.Kom., M.Cs, dan Ibu Nahumi Nugrahaningsih, Ph.D sebagai dosen penguji yang memberikan saran dan masukan terhadap produk Skripsi saya. Dan kepada Ibu Felicia Sylviana, S.T.,MM, selaku Dosen Pembimbing Akademik saya.

Terima kasih juga untuk teman-teman berharga yang selalu bisa diandalkan. Untuk Slis Ayu Wulan Sari dan Zakharia, yang hampir selalu bersama saya 24 jam sehari 7 hari seminggu. Untuk Muhammad Riski Saputra, teman baik dan teman satu bimbingan Skripsi yang selalu memberikan inspirasi dan yang baik. Untuk Siti Nor Azizah, teman yang selalu memberi nasihat baik walaupun terkadang sedikit melantur. Untuk teman-teman GSC yang *nyeleneh*, tetapi menyenangkan dan selalu menghibur, Ray, Roy, Caca, Andrew, Ghiraldi, Made, Sepri, Joey, Romzy, Apri, Frisnando, Zikri, Teguh, Rio, dan Dandy. Dan untuk teman KKN yang sudah seperti ibu, Naomi Floriyentina, S.Pd. Dan untuk teman-teman dekat lain, dan teman-teman mahasiswa Teknik Informatika UPR angkatan 2016, yang sudah berjuang bersama-sama sejak awal masuk kuliah di Jurusan Teknik Informatika. Juga kepada pihak-pihak yang selalu menanyakan ‘kapan lulus?’, sehingga memotivasi dan membakar semangat saya dalam mengerjakan Skripsi. Dan untuk *Seventeen* yang selalu menghibur lewat lagu-lagu mereka yang menginspirasi dan memberi semangat selama mengerjakan Skripsi.

## **KATA PENGANTAR**

Puji syukur kepada Tuhan yang Maha Esa, karena berkat dan kasih karunia-Nya penulis bisa menyelesaikan Skripsi yang berjudul *Klasifikasi Genre Cerita Pendek Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes Classifier*.

Bersama dengan diangkatnya laporan ini, penulis mengucapkan terima kasih banyak dan penghargaan yang tinggi kepada yang terhormat Dosen Pembimbing I, Ibu Sherly Christina, S.Kom., M.Kom, dan Dosen Pembimbing II, Ibu Ariesta Lestari, S.Kom., M.Cs., Ph.D, yang sudah meluangkan waktu untuk membimbing penulis. Penulis juga mengucapkan terima kasih kepada pihak-pihak, kerabat, serta rekan yang sudah membantu penulis dalam penyusunan Skripsi.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna dari penulisan maupun pemahaman. Oleh karena itu, penulis mengharapkan pendapat, kritik, dan saran yang membangun untuk mengembangkan dan menyempurnakan laporan ini. Semoga laporan ini dapat dipahami pembaca, serta dapat memberikan manfaat dan masukan bagi pembaca.

Palangka Raya, Oktober 2020

Penulis

## **“KLASIFIKASI GENRE CERITA PENDEK BERBAHASA INDONESIA MENGUNAKAN ALGORITMA *NAÏVE BAYES CLASSIFIER*”**

**Petra Wara Nandayani R DBC 116 058**

Jurusan Teknik Informatika Fakultas Teknik Universitas Palangka Raya  
Kampus Tunjung Nyaho Jl. Yos Sudarso Palangka Raya 73112  
email : [ptrwrndaynirml@gmail.com](mailto:ptrwrndaynirml@gmail.com)

### **Abstrak**

Cerita pendek atau cerpen adalah cerita karangan atau karya sastra yang panjangnya tidak lebih dari sepuluh ribu kata, sehingga ceritanya relatif pendek dan habis dibaca sekali duduk. Cerpen memiliki banyak kategori atau genre sesuai dengan isi cerpen. Genre adalah jenis karya sastra. Saat ini sudah banyak situs-situs *online* cerpen yang memfasilitasi penulis cerpen untuk mempublikasikan karyanya. Pada situs-situs *online* yang ada saat ini, pengklasifikasian genre cerpen dilakukan oleh penulis cerpen. Penelitian ini mencoba meningkatkan layanan cerpen *online*, yaitu meningkatkan fungsionalitas sistem agar dapat melakukan pengklasifikasian genre cerpen secara otomatis menggunakan konsep *text mining* dengan algoritma *Naïve Bayes Classifier*.

Metodologi penelitian yang digunakan adalah metode pengumpulan data, metodologi pengembangan perangkat lunak, metode klasifikasi, dan pengujian. Tahapan klasifikasi meliputi ekstraksi data, pemrosesan teks, dan pengujian. Metodologi pengembangan perangkat lunak yang digunakan adalah *Waterfall*, dengan tahapan analisis, desain, pengkodean, dan pengujian. Penelitian ini menggunakan 200 data cerpen, yang dibagi menjadi data latih dan data uji. Sistem klasifikasi dilatih terlebih dahulu untuk membentuk model klasifikasi yang akan digunakan untuk mengklasifikasi data baru. Pengujian menggunakan metode *black box testing* dan metode akurasi Sistem diimplementasikan dalam bentuk *website*, menggunakan bahasa pemrograman *PHP* dan *database MySQL*.

Hasil pengujian dengan akurasi tertinggi diperoleh pada skenario pengujian kelima, yaitu 85%, dengan komposisi data 90% data latih dan 10% data uji. Pengujian menggunakan *black box testing* menunjukkan sistem dapat berjalan sesuai yang tujuan. Berdasarkan hasil implementasi dan pengujian, sistem klasifikasi genre cerpen Berbahasa Indonesia dapat memenuhi tujuan penelitian. Implementasi *website* dapat diakses pada alamat [www.mycerpen.com](http://www.mycerpen.com).

Kata-kata Kunci : **Cerita Pendek, Klasifikasi Genre, *Naïve Bayes Classifier*, *Waterfall*, *Website***

## **“CLASSIFICATION OF INDONESIAN SHORT STORY GENRE USING NAÏVE BAYES CLASSIFIER ALGORITHM”**

**Petra Wara Nandayani R DBC 116 058**

Jurusan Teknik Informatika Fakultas Teknik Universitas Palangka Raya  
Kampus Tunjung Nyaho Jl. Yos Sudarso Palangka Raya 73112  
email : ptrwrndaynirml@gmail.com

### **Abstract**

A short story is a story that is written about less than ten thousand words so that the story is relatively short and no need much time to finish reading a whole story. The short story has many categories or genres according to the contents of the story. Genre is a type of story. Nowadays, many online sites facilitate writers to publish their works. Currently, on online sites, the classification of genres is done by the writer. This research is trying to improve the online short story services, improve the functionality of the system that can automatically classify the short story genre implementing the concept of text mining using Naïve Bayes Classifier algorithm.

The methodology included in the research is the data collecting method, software development methodology, classification method, dan testing. The stages that are implemented in the system are data extracting, text processing, and classification. This research uses Waterfall software development methodology, the stages of the methodology are analysis, design, coding, and testing. This research uses 200 short stories of data, which were divided into training data and test data. First, the classification system needs to be trained to determine the classification model that will be used to classify new data. The testing method is black box testing and accuracy.. The system is a website based, which is built using PHP programming language and MySQL database.

The highest accuracy of the scenario test is 85%, with a composition of 90% training data and 10% test data. The testing of system functionality using black box testing shows the system can run as the research purposes. Based on the results of implementation and testing, the Indonesian short story genre classification system can meet the research purposes. Website implementation of the system can be accessed at [www.mycerpen.com](http://www.mycerpen.com).

**Keywords: Short Story, Genre Classification, Naïve Bayes Classifier, Waterfall, Website**

## DAFTAR ISI

*Halaman*

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>ii</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>iii</b>
<b>HALAMAN PERNYATAAN.....</b>	<b>iv</b>
<b>HALAMAN RIWAYAT PENYUSUN .....</b>	<b>v</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>vi</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>ABSTRAK .....</b>	<b>ix</b>
<b>ABSTRACT.....</b>	<b>x</b>
<b>DAFTAR ISI.....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiv</b>
<b>DAFTAR TABEL .....</b>	<b>xvii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah.....	4
1.4 Tujuan.....	5
1.5 Manfaat.....	5
1.6 Sistematika Penulisan.....	5
1.7 Jadwal Skripsi.....	7

<b>BAB II LANDASAN TEORI .....</b>	<b>8</b>
2.1    Tinjauan Pustaka .....	8
2.2    Cerita Pendek .....	18
2.3 <i>Text Mining</i> .....	21
2.3.1    Tahapan <i>Text Mining</i> .....	21
2.3.2    Klasifikasi .....	24
2.3 <i>Term Frequency-Inverse Document Frequency</i> .....	25
2.4 <i>Naïve Bayes Classifier</i> .....	27
2.4.1    Teorema <i>Bayes</i> .....	27
2.4.2 <i>Naïve Bayes</i> Untuk Klasifikasi.....	28
2.5    Pengujian .....	32
2.5.1 <i>Black Box Testing</i> .....	32
2.5.2    Akurasi .....	33
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>36</b>
3.1    Metode Pengumpulan Data .....	36
3.2    Metodologi Pengembangan Perangkat Lunak.....	36
3.2.1    Analisis.....	37
3.2.1    Desain.....	38
3.2.3    Pengkodean .....	59
3.2.4    Pengujian.....	60
3.2.5    Pemeliharaan .....	60
3.3    Metode Klasifikasi .....	60

3.3.1	Ekstraksi Data .....	61
3.3.2	Pemrosesan Teks .....	62
3.4	Klasifikasi Cerpen .....	87
3.5	Skenario Pengujian.....	96
3.5.1	<i>Black Box Testing</i> .....	96
3.5.2	Akurasi .....	98
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>99</b>
4.1	Implementasi Pengkodean.....	99
4.1.1	Implementasi Desain <i>User Interface</i> .....	99
4.2	Pengujian .....	113
4.2.1	<i>Black Box Testing</i> .....	113
4.2.2	Akurasi .....	126
<b>BAB V PENUTUP.....</b>		<b>150</b>
5.1	Kesimpulan.....	150
5.2	Saran .....	151
<b>DAFTAR PUSTAKA .....</b>		<b>152</b>
<b>LAMPIRAN.....</b>		<b>155</b>

## DAFTAR GAMBAR

Gambar 3.1 Metode <i>Waterfall</i> Pressman (2010) .....	36
Gambar 3.2 <i>Flowchart</i> Sistem .....	37
Gambar 3.3 <i>Use Case</i> .....	39
Gambar 3.4 <i>Activity Diagram</i> Admin .....	42
Gambar 3.5 <i>Activity Diagram</i> Penulis .....	43
Gambar 3.6 <i>Class Diagram</i> .....	45
Gambar 3.7 Desain <i>Database</i> .....	46
Gambar 3.8 Halaman <i>Login</i> .....	47
Gambar 3.9 Beranda.....	48
Gambar 3.10 Halaman Kelola Cerpen Masuk .....	48
Gambar 3.11 Halaman Kelola <i>Stopword</i> .....	49
Gambar 3.12 Halaman Kelola Tanda Baca.....	50
Gambar 3.13 Halaman Kelola <i>Dataset</i> .....	51
Gambar 3.14 Halaman <i>Upload Dataset</i> .....	51
Gambar 3.15 Halaman Pelatihan .....	52
Gambar 3.16 Halaman Pelatihan .....	52
Gambar 3.17 Halaman Hasil Pelatihan .....	53
Gambar 3.18 Halaman Pengujian .....	54
Gambar 3.19 Halaman Hasil Klasifikasi Data Uji .....	55
Gambar 3.20 Halaman Beranda .....	56

Gambar 3.21 Halaman Baca Cerpen.....	57
Gambar 3.22 Halaman Isi Cerpen.....	58
Gambar 3.23 Halaman <i>Post</i> Cerpen.....	59
Gambar 3.24 <i>Flowchart</i> Tahap Klasifikasi.....	61
Gambar 3.25 <i>Flowchart</i> Perhitungan <i>TF-IDF</i> .....	75
Gambar 3.26 <i>Flowchart</i> Perhitungan <i>Naïve Bayes Classifier</i> .....	88
Gambar 4.1 Halaman <i>Login</i> .....	99
Gambar 4.2 Halaman Beranda Admin .....	100
Gambar 4.3 Halaman Kelola Cerpen Masuk .....	100
Gambar 4.4 Halaman Kelola <i>Stopword</i> .....	101
Gambar 4.5 Halaman Kelola Tanda Baca.....	101
Gambar 4.6 Halaman Kelola <i>Dataset</i> .....	102
Gambar 4.7 Pilih Kelas <i>Dataset</i> .....	102
Gambar 4.8 Halaman Kelola <i>Dataset</i> .....	103
Gambar 4.9 Halaman <i>Upload Dataset</i> .....	103
Gambar 4.10 Halaman Pelatihan .....	104
Gambar 4.11 Halaman Pelatihan .....	105
Gambar 4.12 Halaman Hasil Pelatihan .....	106
Gambar 4.13 Halaman Pengujian .....	107
Gambar 4.14 Perhitungan Klasifikasi <i>Naïve Bayes Classifier</i> .....	108
Gambar 4.15 Hasil Klasifikasi <i>Naïve Bayes Classifier</i> .....	108
Gambar 4.16 Halaman Hasil Klasifikasi Data Uji .....	109

Gambar 4.17 Halaman Beranda .....	109
Gambar 4.18 Halaman Baca Cerpen.....	110
Gambar 4.19 Halaman Isi Cerpen.....	111
Gambar 4.20 Halaman <i>Post</i> Cerpen.....	111
Gambar 4.21 Pemberitahuan Genre Cerpen.....	112
Gambar 4.22 Grafik Akurasi Klasifikasi Sistem.....	132
Gambar 4.23 Grafik Akurasi Klasifikasi Sistem.....	138
Gambar 4.24 Waktu Pengujian Data.....	144
Gambar 4.25 Kurva Hasil Seleksi Fitur .....	146
Gambar 4.26 Kurva Hasil Seleksi Fitur .....	148

## DAFTAR TABEL

Tabel 1.1 Jadwal Penyusunan Skripsi.....	7
Tabel 2.1 Perbandingan Tinjauan Pustaka.....	13
Tabel 2.1 Perbandingan Tinjauan Pustaka (lanjutan) .....	14
Tabel 2.1 Perbandingan Tinjauan Pustaka (lanjutan) .....	15
Tabel 2.1 Perbandingan Tinjauan Pustaka (lanjutan) .....	16
Tabel 2.1 Perbandingan Tinjauan Pustaka (lanjutan) .....	17
Tabel 2.2 Matriks Confusion Untuk Klasifikasi 2 Kelas .....	34
Tabel 3.1 Definisi Aktor .....	38
Tabel 3.2 Definisi <i>Use Case</i> .....	38
Tabel 3.2 Definisi <i>Use Case</i> (lanjutan).....	39
Tabel 3.2 Definisi <i>Use Case</i> (lanjutan).....	40
Tabel 3.3 Tabel <i>Dataset</i> .....	61
Tabel 3.4 <i>Case Folding</i> .....	62
Tabel 3.4 <i>Case Folding</i> (lanjutan) .....	63
Tabel 3.4 <i>Case Folding</i> (lanjutan) .....	64
Tabel 3.4 <i>Case Folding</i> (lanjutan) .....	65
Tabel 3.4 <i>Case Folding</i> (lanjutan) .....	66
Tabel 3.4 <i>Tokenizing</i> .....	67
Tabel 3.4 <i>Tokenizing</i> (lanjutan).....	68
Tabel 3.4 <i>Tokenizing</i> (lanjutan).....	69

Tabel 3.4 <i>Tokenizing</i> (lanjutan).....	70
Tabel 3.5 <i>Filtering</i> .....	71
Tabel 3.5 <i>Filtering</i> (lanjutan).....	72
Tabel 3.7 Pembobotan <i>TF-IDF</i> .....	76
Tabel 3.7 Pembobotan <i>TF-IDF</i> (lanjutan) .....	77
Tabel 3.7 Pembobotan <i>TF-IDF</i> (lanjutan) .....	78
Tabel 3.7 Pembobotan <i>TF-IDF</i> (lanjutan) .....	79
Tabel 3.7 Pembobotan <i>TF-IDF</i> (lanjutan) .....	80
Tabel 3.8 Bobot Kata .....	80
Tabel 3.8 Bobot Kata (lanjutan).....	81
Tabel 3.8 Bobot Kata (lanjutan).....	82
Tabel 3.8 Bobot Kata (lanjutan).....	83
Tabel 3.8 Bobot Kata (lanjutan).....	84
Tabel 3.9 Tabel Kata Contoh Klasifikasi .....	88
Tabel 3.10 Matriks Confusion .....	95
Tabel 4.1 Pengujian <i>Form Login Admin</i> .....	113
Tabel 4.2 Pengujian Halaman Kelola Cerpen Masuk .....	114
Tabel 4.2 Pengujian Halaman Kelola Cerpen Masuk (lanjutan).....	115
Tabel 4.3 Pengujian Halaman Kelola <i>Stopword</i> .....	115
Tabel 4.3 Pengujian Halaman Kelola <i>Stopword</i> (lanjutan).....	116
Tabel 4.3 Pengujian Halaman Kelola <i>Stopword</i> (lanjutan).....	117
Tabel 4.4 Pengujian Halaman Kelola Tanda Baca.....	117

Tabel 4.4 Pengujian Halaman Kelola Tanda Baca (lanjutan) .....	118
Tabel 4.4 Pengujian Halaman Kelola Tanda Baca (lanjutan) .....	119
Tabel 4.5 Pengujian Halaman Kelola <i>Dataset</i> .....	120
Tabel 4.5 Pengujian Halaman Kelola <i>Dataset</i> (lanjutan).....	121
Tabel 4.6 Pengujian Halaman Pelatihan .....	121
Tabel 4.7 Pengujian Halaman Hasil Pelatihan.....	121
Tabel 4.7 Pengujian Halaman Hasil Pelatihan (lanjutan) .....	122
Tabel 4.8 Pengujian Halaman Pengujian .....	122
Tabel 4.8 Pengujian Halaman Pengujian (lanjutan).....	123
Tabel 4.9 Pengujian Halaman Hasil Klasifikasi Data Uji.....	123
Tabel 4.9 Pengujian Halaman Hasil Klasifikasi Data Uji (lanjutan) .....	124
Tabel 4.10 Pengujian Halaman Baca Cerpen.....	124
Tabel 4.10 Pengujian Halaman Baca Cerpen (lanjutan) .....	125
Tabel 4.11 Pengujian Halaman <i>Post</i> Cerpen.....	125
Tabel 4.11 Pengujian Halaman <i>Post</i> Cerpen (lanjutan) .....	126
Tabel 4.12 Pembagian Data .....	127
Tabel 4.13 Matriks Confusion Pengujian 50%-50%.....	127
Tabel 4.13 Matriks Confusion Pengujian 50%-50% (lanjutan) .....	128
Tabel 4.14 Matriks Confusion Pengujian 60%-40%.....	128
Tabel 4.14 Matriks Confusion Pengujian 60%-40% (lanjutan) .....	129
Tabel 4.15 Matriks Confusion Pengujian 70%-30%.....	129
Tabel 4.16 Matriks Confusion Pengujian 80%-20%.....	130

Tabel 4.17 Matriks Confusion Pengujian 90%-10%.....	131
Tabel 4.18 Hasil Pengujian .....	132
Tabel 4.19 Matriks Confusion Pengujian 50%-50%.....	133
Tabel 4.20 Matriks Confusion Pengujian 60%-40%.....	134
Tabel 4.21 Matriks Confusion Pengujian 70%-30%.....	135
Tabel 4.22 Matriks Confusion Pengujian 80%-20%.....	135
Tabel 4.22 Matriks Confusion Pengujian 80%-20% (lanjutan) .....	136
Tabel 4.23 Matriks Confusion Pengujian 90%-10%.....	136
Tabel 4.23 Matriks Confusion Pengujian 90%-10% (lanjutan) .....	137
Tabel 4.24 Hasil Pengujian .....	137
Tabel 4.24 Hasil Pengujian .....	138
Tabel 4.25 Kata yang Sering Muncul .....	140
Tabel 4.25 Kata yang Sering Muncul (lanjutan).....	141
Tabel 4.25 Kata yang Sering Muncul (lanjutan).....	142
Tabel 4.26 Hasil Seleksi Fitur.....	145
Tabel 4.27 Hasil Seleksi Fitur.....	146
Tabel 4.27 Hasil Seleksi Fitur (lanjutan) .....	147

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Cerita pendek atau yang biasa disebut cerpen adalah cerita karangan atau karya sastra yang panjangnya tidak lebih dari sepuluh ribu kata, sehingga ceritanya relatif pendek. Cerpen adalah sebuah karangan berbentuk prosa (karya sastra) fiksi yang habis dibaca sekali duduk, artinya adalah tidak membutuhkan waktu yang lama untuk menyelesaikan satu cerita (Mustofa, 2016).

Cerpen dapat dibaca oleh pembaca dari berbagai golongan usia dan latar belakang, mulai dari anak-anak, remaja, dewasa, atau bahkan orang tua. Oleh karena itu, cerpen memiliki banyak kategori atau genre sesuai dengan isi cerpen tersebut (Somantri, 2017). Cerpen sebagai bagian dari karya fiksi memiliki beberapa genre. Genre berasal dari kata bahasa Prancis yang berarti 'jenis'. Jadi, genre sastra berarti jenis karya sastra. Menurut [www.penerbitdeepublish.com](http://www.penerbitdeepublish.com) dalam artikel berjudul "Cara Membuat Buku Fiksi : Lima Genre Yang Wajib Kamu Ketahui!!" oleh Mas Aji Gustiawan (2016), ada lima genre besar fiksi, yaitu romantis, *science fiction*, fantasi, misteri, dan horor.

Biasanya pengkategorian genre cerpen merupakan peranan langsung dari penulis cerpen tersebut. Penulis cerpen menentukan sendiri genre dari cerpen yang ditulisnya saat penulis tersebut akan mempublikasi cerpen karangannya. Misalnya ada paragraf penggalan cerpen sebagai berikut :

- 1) *"Dari balik pohon aku mulai mengambil posisi untuk mendekati mereka. Rasa keingintahuanku yang besar terus memintaku untuk mengetahui*

*apa motivasi mereka melakukan semua ini. Apa yang mereka kerjakan saat ini sungguh-sungguh menarik perhatianku. Sehingga aku memutuskan untuk maju dan berlindung di balik pohon yang lebih besar di arah kanan untuk mendengarkan mereka. Belum saatnya aku muncul, pikirku. Aku harus mendapatkan informasi yang lebih lengkap untuk menangani semua kejanggalan ini.”* (Penjaga Harta Warisan, Oleh Fariz Edgar).

- 2) *“Ada sebuah sinar di ujung sebelah barat, menyita perhatiannya. Sinar putih yang semula kecil, lalu bermetamorfosis menjadi besar lalu membesar hingga mega giga besar. Dan entah mengapa, Ge sangat penasaran. Saat dirinya hendak menanyakan kepada keempat temannya apakah mereka juga melihat cahaya putih atau tidak, mulutnya tidak bisa berkata-kata. Sementara itu, teman-temannya seketika berubah seperti patung. Diam tak bergeming sekalipun. Seakan wantu berhenti.”* (Cahaya di Ujung Sana, oleh Creafids).

Contoh (1) adalah cerpen bergenre misteri, sedangkan contoh (2) adalah cerpen bergenre fantasi. Ketika seseorang membaca kedua contoh cerpen di atas, maka seseorang mungkin dapat langsung membedakan genre keduanya, karena memahami genre cerpen membutuhkan pengetahuan dan pengalaman dari pakar manusia yang memahami dengan benar konten atau gaya bahasa sebuah cerpen.

Saat ini sudah terdapat beberapa situs-situs *online* cerpen yang memfasilitasi penulis dan pembaca untuk mempublikasikan dan mengakses cerpen. Pada sistem-sistem *online* yang ada saat ini, pengklasifikasian genre cerpen dilakukan oleh penulis cerpen. Oleh karena itu, untuk meningkatkan layanan cerpen *online*, maka penelitian ini mencoba meningkatkan fungsionalitas sistem agar dapat melakukan pengklasifikasian genre cerpen secara otomatis.

Penelitian ini menerapkan konsep *text mining* agar sistem dapat mengklasifikasikan genre cerpen, karena data yang digunakan berupa teks cerpen. Sebelum melakukan pengklasifikasian, data cerpen melalui tahap *text*

*preprocessing* dan memberikan pembobotan menggunakan algoritma *TF-IDF*, kemudian pengklasifikasian dilakukan menggunakan algoritma *Naïve Bayes Classifier*. *Naïve Bayes Classifier* adalah algoritma yang memungkinkan sistem untuk melakukan pengklasifikasian otomatis, berdasarkan hasil pembobotan *TF-IDF*. Penelitian ini menggunakan algoritma *Naïve Bayes Classifier*, karena berdasarkan hasil pengujian dari beberapa penelitian, algoritma *Naïve Bayes Classifier* memiliki kinerja yang cukup baik untuk klasifikasi teks. Pada penelitian Nastiti dkk (2019) tentang klasifikasi novel menggunakan *Naïve Bayes Classifier*, menghasilkan *accuracy* 80,5%. Penelitian lainnya, yaitu penelitian Apriliani dkk (2017) didapat *accuracy* 90%.

Berdasarkan latar belakang yang diuraikan di atas, maka tujuan dari penelitian ini adalah mengklasifikasikan genre cerpen ke dalam kelas genre, dengan skenario pengujian performa pengklasifikasian oleh sistem menggunakan metode *accuracy* (Prasetyo, 2014:257). Jadi judul pada penelitian ini adalah "Klasifikasi Genre Cerita Pendek Berbahasa Indonesia Menggunakan Algoritma *Naïve Bayes Classifier*"

## 1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah dari penelitian ini adalah bagaimana mengklasifikasikan genre cerpen berbahasa Indonesia menggunakan algoritma *Naïve Bayes Classifier*.

### 1.3 Batasan Masalah

Batasan masalah untuk membatasi penelitian ini adalah sebagai berikut:

1. Genre cerpen yang ditetapkan dalam penelitian ini hanya empat, yaitu romantis, fantasi, horor, dan misteri, berdasarkan lima genre besar fiksi yang diuraikan pada latar belakang. Genre *science fiction* tidak dimasukkan karena sumber data cerpen *science fiction* berbahasa Indonesia yang kurang memadai.
2. Sistem ini digunakan untuk mengklasifikasi genre cerpen berdasarkan kelas yang sudah ditentukan.
3. Data yang digunakan adalah cerpen berbahasa Indonesia
4. *Dataset* cerpen yang digunakan dalam penelitian ini berjumlah 200 data cerpen, diambil dari *cerpenmu.com*, *lakonhidup.com*, *detectivestoryid.wordpress.com*, dan *fiksilotus.com*, dan *wattpad.com*, dengan jumlah cerpen untuk setiap cerpen sama rata.
5. *Dataset* cerpen yang dikumpulkan adalah cerpen yang sudah dikategorikan atau dilabelkan oleh *website* sumber, sebagai pakar dalam penelitian ini.
6. Pada tahap pemrosesan teks tidak dilakukan *stemming* karena ingin mendapatkan kata asli dengan makna yang sebenarnya.
7. Sistem dirancang menggunakan *HTML*, *CSS*, dan *XAMPP* sebagai *local server*, dan *database* sistem menggunakan *MySQL*. Bahasa pemrograman yang digunakan untuk membangun sistem adalah *PHP*.

8. Algoritma yang digunakan adalah algoritma pembobotan *TF-IDF* dan algoritma klasifikasi *Naïve Bayes Classifier*, serta tidak membandingkan dengan algoritma lain.
9. Pengujian dilakukan secara *online*.

#### **1.4 Tujuan**

Tujuan dari penelitian ini adalah melakukan klasifikasi genre cerpen ke dalam empat kelas genre berbahasa Indonesia, yaitu romantis, fantasi, misteri, dan horor, dengan menerapkan algoritma *Naïve Bayes Classifier*.

#### **1.5 Manfaat**

Manfaat dari penelitian ini adalah tersedianya sistem yang dapat mengklasifikasikan genre cerpen secara otomatis ke dalam kelas genre yang telah ditentukan.

#### **1.6 Sistematika Penulisan**

Sistematika penulisan laporan dalam tugas akhir ini antara lain :

##### **BAB I PENDAHULUAN**

Bab ini diuraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat, metode pengumpulan data, sistematika penulisan dan jadwal kegiatan.

## **BAB II LANDASAN TEORI**

Bab ini menjelaskan berbagai konsep dasar dan teori-teori yang berkaitan dengan topik penelitian. Pada bab ini juga mengemukakan teori-teori yang relevan untuk memecahkan masalah.

## **BAB III METODOLOGI PENELITIAN**

Pada bab ini menjelaskan proses perancangan program yang akan dibuat dalam menyelesaikan proyek.

## **BAB IV HASIL DAN PEMBAHASAN**

Bab ini berisi hasil dan pembahasan dari penelitian yang telah dilakukan.

## **BAB V KESIMPULAN DAN SARAN**

Bab ini terdiri dari kesimpulan dan saran. Kesimpulan memuat kesimpulan-kesimpulan yang merupakan rangkuman dari hasil analisis kinerja pada proyek yang dilakukan. Saran merupakan perbaikan yang diperlukan saat ini atau masa yang akan datang.

## 1.7 Jadwal Skripsi

Tabel 1.1 Jadwal Penyusunan Skripsi

No.	Rencana Kegiatan	Bulan dan Minggu																			
		Maret				April				Mei - Agustus				September				Oktober			
		I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
1	Penyusunan Proposal dan Seminar Proposal	■																			
2	Analisis dan Desain Sistem													■							
3	Pembuatan Program dan Pengujian													■							
4	Pembuatan Laporan													■	■	■	■	■			
5	Seminar Hasil													■	■						
6	Seminar Tugas Akhir																	■	■		

## BAB II

### LANDASAN TEORI

#### 2.1 Tinjauan Pustaka

Terdapat beberapa penelitian terdahulu yang memiliki keterkaitan dengan penelitian yang dilakukan.

Penelitian yang pertama adalah penelitian yang dilakukan oleh Vinna Rajmayanti Setyaning Nastiti, Setio Basuki, dan Hilman (2019) yang berjudul “Klasifikasi Sinopsis Novel Menggunakan Metode *Naïve Bayes Classifier*”. Tujuan dari penelitian ini adalah membangun sebuah sistem yang dapat melakukan klasifikasi otomatis dengan menggunakan sinopsis pada novel dengan metode *Term Frequency Inverse Document Frequency (TF-IDF)* dan *Multinomial Naïve Bayes*. Penelitian ini melakukan klasifikasi sinopsis novel menggunakan metode *Naïve Bayes Classifier* dan *TF-IDF* dengan harapan hasil yang didapatkan dapat membantu pengelola toko. Tahapan yang dilakukan dalam penelitian ini adalah pengumpulan data teks berupa sinopsis novel, *labeling data*, *preprocessing*, *feature selection* dengan *TF-IDF*, pelatihan data dan pengujian data menggunakan *Naïve Bayes*, klasifikasi, dan evaluasi. *Dataset* yang digunakan dalam penelitian ini adalah data teks berupa sinopsis novel, *dataset* dibagi menjadi dua, yaitu 600 data latih dan 200 data uji. Hasil dari penelitian ini menampilkan informasi mengenai prediksi oleh sistem dari masing-masing kelas. Dari pengujian klasifikasi algoritma *Naïve Bayes Classifier* menggunakan metode *accuracy* didapatkan akurasi sebesar 80,5%.

Penelitian yang kedua adalah penelitian yang dilakukan oleh Apriliana, Natalis Ransi, dan Jumadil Nangi (2017) yang berjudul “Implementasi *Text Mining* Klasifikasi Skripsi Menggunakan Metode *Naïve Bayes Classifier*”. Tujuan penelitian ini adalah mengklasifikasikan kategori skripsi ke dalam tiga kategori, yaitu Rekayasa Perangkat Lunak, Komputasi Berbasis Jaringan, dan Komputasi Cerdas Visual. Tahapan yang dilakukan dalam penelitian ini adalah *text pre processing*, yaitu *case folding* dan *tokenizing*, kemudian *feature selection* yaitu melakukan penghilangan *stopword*. Hasil dari penelitian ini berupa sistem yang dapat mengklasifikasikan kategori skripsi dengan menerapkan konsep *text mining* menggunakan algoritma *Naïve Bayes Classifier*. Berdasarkan hasil pengujian klasifikasi menggunakan metode *accuracy* didapatkan akurasi sebesar 90%.

Penelitian yang ketiga adalah penelitian yang dilakukan oleh Admaja Dwi Herlambang dan Satrio Hadi Wijoyo (2018) yang berjudul “Algoritma *Naïve Bayes* Untuk Klasifikasi Sumber Belajar Berbasis Teks Pada Mata Pelajaran Produktif di SMK Rumpun Teknologi Informasi dan Komunikasi”. Tujuan dari penelitian ini adalah mendeskripsikan hasil klasifikasi dan evaluasi kualitas sumber belajar berbasis teks menggunakan algoritma *Naïve Bayes*. Data sumber belajar yang diperoleh sebanyak 235 data yang sudah terklasifikasi berdasarkan masing-masing kelas. Data akan dikelompokkan menjadi sembilan kelas yaitu Animasi 2D, Animasi 3D, Desain Multimedia, Basis Data, Pemodelan Perangkat Lunak, Pemrograman Dasar, Komputer Terapan Jaringan, Komunikasi Data, dan Sistem Operasi. Tahapan yang dilakukan dalam penelitian adalah pengkoleksian data set, pemrosesan awal *text mining*, pembobotan *TF-IDF*, pengklasifikasian *Naïve Bayes*,

dan evaluasi akurasi. Hasil dari penelitian adalah sistem yang dapat mengklasifikasi sumber belajar berdasarkan ciri esensial setiap mata pelajaran. Pengujian menggunakan metode *accuracy* menghasilkan nilai akurasi tertinggi 81,48%, sedangkan nilai akurasi terendah sebesar 79,63%.

Penelitian yang keempat adalah penelitian yang dilakukan oleh Bonifacius Vicky Indriyono, Ema Utami, dan Andi Sunyoto (2015) yang berjudul “Klasifikasi Jenis Buku Berdasarkan Judul dan Sinopsis Menggunakan Metode *Naïve Bayes Classifier* (Studi Kasus: STMIK Kadiri)”. Tujuan penelitian ini adalah membangun sebuah aplikasi yang dapat menentukan atau mengklasifikasi jenis atau kategori buku yang pada awalnya dilakukan secara manual oleh petugas perpustakaan dengan memilih daftar jenis atau kategori buku yang tersedia. Data yang digunakan dalam penelitian ini berjumlah 900 dokumen yang terbagi 100 buku tiap kategori. Aplikasi yang dibangun berbasis *desktop* menggunakan *compiler Delphi*. Penelitian menggunakan metode klasifikasi *text mining Naïve Bayes Classifier*. Dari penelitian ini hasil akurasi tertinggi adalah 90% dan akurasi terendah sebesar 74,2%.

Penelitian yang kelima adalah penelitian yang dilakukan oleh Akhmad Pandhu Wijaya dan Heru Agus Santoso (2016) yang berjudul “*Naïve Bayes Classification* pada Klasifikasi Dokumen Untuk Identifikasi Konten *E-Government*”. Objek dari penelitian ini adalah situs *web* Jawa Tengah dan diklasifikasikan oleh *Naïve Bayes Classification* (NBC), dengan tujuan memfasilitasi klasifikasi dokumen bahasa Indonesia untuk identifikasi konten *e-government*. Penelitian ini menggunakan teknik *text mining* dengan mengusulkan

algoritma klasifikasi *Naïve Bayes Classification* (NBC) dan pembobotan *TF-IDF*. Tahap yang dilakukan sebelum proses klasifikasi adalah *pre processing*, yaitu *tokenizing*, *stopword filtering*, kemudian *word frequency training* dengan menggunakan metode pembobotan *TF-IDF*. Tahap selanjutnya yaitu proses pelatihan, menemukan pola-pola untuk membangun model klasifikasi dengan algoritma *Naïve Bayes Classification* (NBC). Hasil dari klasifikasi dokumen menggunakan algoritma *Naïve Bayes Classifier* (NBC) dan pembobotan *TF-IDF* dalam penelitian ini dengan data latih sebanyak 260 dokumen politik dan 222 dokumen ekonomi menggunakan 40 data uji menunjukkan nilai akurasi yang baik pada keseluruhan klasifikasi, dengan akurasi keseluruhan klasifikasi sebesar 85%.

Penelitian yang keenam adalah penelitian yang dilakukan oleh Enggal Suci Febriani (2017) yang berjudul “Klasifikasi Konten Berita Surat Kabar Berdasarkan Judul dengan *Text Mining* Menggunakan Metode *Naïve Bayes* (Studi Kasus : Radar Kediri)”. Tujuan penelitian ini adalah merancang sebuah sistem yang dapat mengklasifikasikan konten berita surat kabar dengan mengimplementasikan metode *text mining* dan algoritma *Naïve Bayes Classifier* dalam proses klasifikasi konten surat kabar. Judul dan *headline* berita diproses melalui tahapan *text mining*, yaitu *text pre processing* dengan tahapan *case folding*, *tokenizing*, *filtering*, kemudian *text transformation* dengan *stopword removal*, selanjutnya *feature selection* dan *pattern discovery*. Penelitian ini membantu mempermudah user dalam memilih dan mengategorikan konten berita sehingga meminimalkan waktu dan sumber daya manusia dalam pengklasifikasian dan pencarian konten berita serta membantu *user* dalam penyusunan surat kabar harian. Berdasarkan kesimpulan

hasil penelitian ini, direkomendasikan tujuan pokok penggunaan sistem pengklasifikasian ini adalah mengklasifikasikan berita secara otomatis sehingga penyusunan surat kabar harian bisa terselesaikan tepat waktu. Penelitian ini menghasilkan suatu pengklasifikasian konten berita sesuai kategori yang telah ditentukan dengan tingkat akurasi mencapai 79%.



Tabel 2.1 Perbandingan Tinjauan Pustaka

<b>Peneliti</b>	Vinna Rajmayanti Setyaning Nastiti, Setio Basuki, dan Hilman (2019)	Apriliana, Natalis Ransi, dan Jumadil Nangi (2017)	Admaja Dwi Herlambang dan Satrio Hadi Wijoyo (2018)	Bonifacius Vicky Indriyono, Ema Utami, dan Andi Sunyoto (2015)	Akhmad Pandhu Wijaya dan Heru Agus Santoso (2016)	Enggal Suci Febriani (2017)
<b>Judul</b>	Klasifikasi Sinopsis Novel Menggunakan Metode <i>Naïve Bayes Classifier</i>	Implementasi <i>Text Mining</i> Klasifikasi Skripsi Menggunakan Metode <i>Naïve Bayes Classifier</i>	Algoritma <i>Naïve Bayes</i> Untuk Klasifikasi Sumber Belajar Berdasarkan Teks Pada Mata Pelajaran	Klasifikasi Jenis Buku Berdasarkan Judul dan Sinopsis Menggunakan Metode <i>Naïve</i>	<i>Naïve Bayes Classification</i> pada Klasifikasi Dokumen Untuk Identifikasi	Klasifikasi Konten Berita Surat Kabar Berdasarkan Judul dengan <i>Text Mining</i> Menggunakan Metode <i>Naïve</i>

Tabel 2.1 Perbandingan Tinjauan Pustaka (lanjutan)

<b>Judul</b>			Produktif di SMK Rumpun Teknologi Informasi dan Komunikasi	<i>Bayes Classifier</i> (Studi Kasus: STMIK Kadiri)	Konten <i>E-Government</i>	<i>Bayes</i> (Studi Kasus : Radar Kediri)
<b>Metode</b>	<i>Term Frequency Inverse Document Frequency (TF-IDF)</i> dan <i>Multinomial Naïve Bayes</i>	algoritma <i>Naïve Bayes Classifier</i>	algoritma <i>Naïve Bayes Classifier</i>	Algoritma <i>Naïve Bayes Classifier</i>	<i>Naïve Bayes Classifier</i> (NBC) dan pembobotan <i>TF-IDF</i>	algoritma <i>Naïve Bayes Classifier</i>

Tabel 2.1 Perbandingan Tinjauan Pustaka (lanjutan)

<b>Data</b>	Sinopsis novel, terdiri dari 600 data latih dan 200 data uji	Pengujian menggunakan 51 skripsi	Data sumber belajar sebanyak 235	900 dokumen yang terbagi 100 buku tiap kategori	Data latih sebanyak 260 dokumen politik dan 222 dokumen ekonomi, dan 40 data uji	200 data latih dan 80 data uji
<b>Tujuan Penelitian</b>	Membangun sebuah sistem yang dapat melakukan klasifikasi otomatis dengan	Mengklasifikasikan kategori skripsi ke dalam tiga kategori, yaitu Rekayasa Perangkat Lunak, Komputasi	Mendesripsikan hasil klasifikasi dan evaluasi kualitas sumber belajar berbasis teks	Membangun sebuah aplikasi yang dapat menentukan atau mengklasifikasi	Memfasilitasi klasifikasi dokumen bahasa Indonesia untuk identifikasi	Merancang sebuah sistem yang dapat mengklasifikasikan konten berita surat kabar

Tabel 2.1 Perbandingan Tinjauan Pustaka (lanjutan)

<b>Tujuan Penelitian</b>	menggunakan sinopsis pada novel	Berbasis Jaringan, dan Komputasi Cerdas Visual	menggunakan algoritma <i>Naive Bayes</i>	jenis atau kategori buku	konten <i>e-government</i>	
<b>Hasil Penelitian</b>	Menampilkan informasi mengenai prediksi masing-masing kelas	Sistem yang dapat mengklasifikasikan kategori skripsi dengan menerapkan konsep <i>text mining</i>	Sistem yang dapat mengklasifikasi sumber belajar berdasarkan ciri esensial setiap mata pelajaran	Aplikasi yang dapat mengklasifikasi jenis atau kategori buku yang pada awalnya dilakukan secara manual	Sistem yang dapat mengklasifikasi dokumen bahasa Indonesia untuk identifikasi konten <i>e-government</i>	Sistem yang dapat mengklasifikasikan konten berita sesuai kategori yang telah ditentukan secara otomatis sehingga penyusunan surat kabar harian bisa

Tabel 2.1 Perbandingan Tinjauan Pustaka (lanjutan)

				oleh petugas perpustakaan		terselesaikan tepat waktu
<b>Performa Sistem</b>	Evaluasi menggunakan <i>confusion matrix</i> mendapatkan akurasi sebesar 80,5%.	Akurasi hasil klasifikasi 90%.	Nilai tertinggi dengan 55 data latih 81,48% dengan 120 data uji dan nilai akurasi terendah 79,63% dengan 54 data uji.	Hasil akurasi tertinggi adalah 90% dengan 90 data latih dan 10 data uji, dan akurasi terendah sebesar 74,2% dengan 30 data latih dan 70 data uji.	Akurasi keseluruhan klasifikasi sebesar 85%.	Tingkat akurasi klasifikasi mencapai 79%.

Dari beberapa penelitian yang telah dilakukan sebelumnya, maka penelitian ini mengangkat tentang klasifikasi cerita pendek berbahasa Indonesia menggunakan algoritma *Naïve Bayes Classifier*. Perbedaan penelitian ini dari penelitian sebelumnya, yaitu :

1. Objek penelitian ini adalah cerita pendek atau cerpen berbahasa Indonesia.
2. Tujuan dari penelitian ini adalah melakukan klasifikasi genre cerpen ke dalam empat kelas genre berbahasa Indonesia, yaitu romantis, fantasi, misteri, dan horor, dengan menerapkan algoritma pembobotan *TF-IDF (Term Frequency – Inverse Document Frequency)* dan algoritma *Naïve Bayes Classifier*.
3. *Dataset* yang digunakan dalam penelitian ini berjumlah 200 data yang sudah dilabelkan berdasarkan label cerpen pada *website* sumber data, sebagai pakar dalam penelitian ini.
4. Sistem yang dibuat diimplementasikan dalam bentuk *website* dan *online*. Pada *website* terdapat fitur untuk mengunggah cerpen baru, kemudian cerpen baru tersebut akan diklasifikasikan oleh sistem secara otomatis ke dalam kelasnya.

## 2.2 Cerita Pendek

Sesuai dengan namanya, cerita pendek atau yang biasa disebut cerpen, dapat diartikan sebagai cerita berbentuk prosa yang pendek. Ukuran pendek bersifat relatif. Menurut Edgar Allan Poe dalam Suyanto (2012), ukuran pendek di sini adalah selesai dibaca dalam sekali duduk, yakni kira-kira kurang dari satu jam. Jadi, proses membaca cerpen tidak membutuhkan waktu yang lama, dan inti ceritanya dapat langsung

ditemukan. Menurut Sumarjdo dan Saini dalam Indriyani (2015), dapat disimpulkan bahwa secara umum cerpen adalah cerita atau narasi yang fiktif (karangan, tidak benar-benar telah terjadi) serta relatif pendek.

Genre berasal dari kata bahasa Prancis yang berarti 'jenis'. Jadi, genre sastra berarti jenis karya sastra. Genre dapat dipahami sebagai suatu macam atau tipe kesastraan yang memiliki seperangkat karakteristik umum, atau kategori pengelompokan karya sastra yang biasanya berdasarkan gaya, bentuk, atau isi. Hal itu membawa pemahaman bahwa dalam sebuah genre sastra terdapat sejumlah elemen yang memiliki kesamaan sifat, dan elemen-elemen itu menunjukkan perbedaan dengan elemen pada genre lain (Nurgiyantoro dalam Munawir, 2013). Genre merupakan pengkategorian tanpa batas-batas yang jelas yang dibuat untuk mengkategorikan cerita yang ada.

Cerpen dapat dibaca oleh pembaca dari berbagai golongan usia dan latar belakang, mulai dari anak-anak, remaja, dewasa, atau bahkan orang tua. Oleh karena itu, cerpen memiliki banyak kategori atau genre sesuai dengan isi cerpen tersebut (Somantri, 2017). Cerpen sebagai bagian dari karya fiksi memiliki beberapa genre. Menurut [www.penerbitdeepublish.com](http://www.penerbitdeepublish.com) dalam artikel berjudul "Cara Membuat Buku Fiksi : Lima Genre Tulisan Fiksi Yang Wajib Kamu Ketahui !!" (Gustiawan, 2016), ada lima genre besar fiksi, yaitu romantis, fantasi, misteri, horor. Dan *science fiction*. Namun, penulis Indonesia belum banyak menulis cerpen bergenre *science fiction* atau fiksi ilmiah (Nyoma Tusthi Eddy, dalam Merawati, 2015), menyebabkan sumber data cerpen bergenre *science fiction* berbahasa Indonesia menjadi tidak memadai untuk

penelitian ini. Karena itu, diambil hanya empat genre sebagai kelas genre yang akan diklasifikasikan.

1. Genre Romantis

Menurut Kamus Besar Bahasa Indonesia, romantis berarti bersifat seperti dalam cerita roman (percintaan), bersifat mesra, mengasyikkan. Disajikan berdasarkan hubungan romantis antara dua orang. Genre ini dicirikan oleh ketegangan sensual, hasrat, dan idealisme.

2. Genre Fantasi

Genre fantasi menampilkan kisah-kisah yang identik dengan petualangan dunia imajiner dan penuh dengan keajaiban. Biasanya mengandung elemen supranatural, legenda, atau mitos-mitos lama yang tak selalu dapat dijelaskan.

3. Genre Horor

Genre horor menyajikan kisah menakutkan, biasanya melibatkan hantu atau makhluk gaib. Mengisahkan tentang pengejaran, pelarian, atau mengalahkan makhluk gaib.

4. Genre Misteri

Cerita dengan genre misteri adalah cerita dengan masalah utama merupakan teka-teki yang harus dipecahkan, identitas terungkap, dan kejahatan terselesaikan. Dalam cerita terdapat petunjuk yang mengarah pada meningkatnya ketegangan saat sudah mendekati jawaban atas misteri.

## 2.3 *Text Mining*

*Text mining* adalah salah satu bidang dari *data mining*. *Text mining* adalah menambang data yang berupa teks, dengan sumber data yang biasanya berupa dokumen, untuk mencari kata-kata unik yang dapat mewakili isi dari suatu set dokumen dengan label yang sama, sehingga dapat dijadikan model untuk menganalisa hubungan antar dokumen. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen, kumpulan teks yang memiliki format yang tidak terstruktur atau semi terstruktur (Nindito, 2016). *Text mining* adalah proses menemukan tren dalam dokumen yang sebelumnya tidak terungkap, dengan memproses dan menganalisa data dalam jumlah besar. Dalam menganalisa sebagian atau keseluruhan teks yang tidak terstruktur, *text mining* mencoba untuk mengasosiasikan satu bagian teks dengan yang lainnya berdasarkan aturan-aturan tertentu (Akrhamsiagustinah, 2015).

*Text mining* adalah penerapan konsep *data mining* untuk mencari pola dalam teks, menganalisa teks untuk mencari informasi yang bermanfaat untuk tujuan tertentu. Berdasarkan ketidakstrukturan data teks, proses *text mining* memerlukan beberapa tahap awal untuk mempersiapkan teks agar dapat diolah.

### 2.3.1 Tahapan *Text Mining*

Menurut Even-Zohar dalam Nugroho (2011) terdapat empat tahap dalam *text mining*, yaitu pemrosesan awal terhadap teks (*text preprocessing*), transformasi teks

(*text transformation*), seleksi fitur (*feature selection*), dan penemuan pola (*pattern discovery*).

### 2.3.1.1 Text Pre-processing

Tujuan dari *pre-processing* adalah untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Tahapan yang dilakukan *text pre-processing* adalah *case folding* dan *tokenizing*.

- 1) *Case folding* yaitu mengkonversi atau mengubah keseluruhan teks di dalam dokumen menjadi bentuk standar, yaitu huruf kecil atau *lower case*.
- 2) *Tokenizing* adalah tahap pemotongan teks dokumen berdasarkan tiap kata yang menyusunnya.

Contoh dari *tokenizing* :

Saya senang sekali

Dipotong berdasarkan kata penyusunnya menjadi,

Saya || senang || sekali

Hasil dari pemotongan atau pemisahan tersebut dinamakan token.

### 2.3.1.2 Text Transformation

Transformasi teks yaitu tahap untuk mendapatkan representasi dokumen yang diharapkan. Pendekatan representasi dokumen yang biasa digunakan oleh model "*bag of words*" dan model ruang (*vector space model*). Transformasi teks juga melakukan pengubahan kata-kata menjadi bentuk dasarnya dan pengurangan dimensi kata di dalam dokumen. Proses ini dilakukan dengan menerapkan *filtering* menghapus

*stopword (stopword removal)*. *Filtering* adalah tahap mengambil kata-kata penting dari hasil *tokenizing* dan membuang kata-kata yang tidak penting, menggunakan *stopword*. *Stopword* adalah kata-kata yang tidak memiliki pengaruh apapun dalam klasifikasi teks, seperti kata sambung, kata ganti orang, kata penunjuk waktu, dan kata tanya.

### 2.3.1.3 *Feature Selection*

*Feature Selection* adalah tahap lanjutan dari pengurangan dimensi pada tahap transformasi teks. Tahap sebelumnya sudah melakukan penghapusan kata yang tidak deskriptif (*stopword*), tetapi tidak semua kata memiliki arti penting. Untuk mengurangi dimensi, pemilihan kata hanya dilakukan terhadap kata-kata yang relevan, yang merepresentasikan isi dari sebuah dokumen. Dasar dari *feature selection* adalah menghapus kata dengan kemunculan yang terlalu sedikit atau terlalu banyak dalam sebuah dokumen. Salah satu pendekatan yang sering digunakan adalah "*bag of words*", yaitu vektor mewakili frekuensi kata. Algoritma yang digunakan dalam *text mining* melakukan perhitungan pada dokumen dan pada *feature*.

Dalam *feature selection* akan dipilih *term* atau kata "unik", yaitu kata yang dapat mewakili sekumpulan dokumen. Karena kata hasil dari tahap *tokenizing* dan *filtering* tidak mungkin semuanya dapat dijadikan vektor dokumen, dan akan menjadi sangat tidak efektif. Langkah awal untuk menemukan kata yang dapat mewakili dokumen adalah dengan menghitung bobot setiap *term* atau kata. Metode pembobotan kata yang digunakan dalam penelitian ini adalah algoritma pembobotan *TF-IDF*.

#### 2.3.1.4 *Pattern Discovery*

*Pattern discovery* adalah tahap untuk menemukan pola pengetahuan (*knowledge*) dari keseluruhan teks. Tahap ini adalah operasi *text mining*, biasanya menggunakan teknik-teknik *data mining*. Masukan awal dari proses *text mining* adalah data teks dan menghasilkan keluaran berupa penemuan pola atau model sebagai hasil dari interpretasi atau evaluasi. Jika hasil keluaran dari penemuan pola belum sesuai untuk aplikasi, maka evaluasi dilanjutkan dengan melakukan iterasi ke satu atau beberapa tahap sebelumnya.

#### 2.3.2 **Klasifikasi**

Klasifikasi adalah kata serapan dari Bahasa Belanda *classificate* dan dari Bahasa Prancis *classification*, yang merujuk pada sebuah metode untuk menyusun data secara sistematis atau menurut beberapa aturan yang ditetapkan. Bisa dikatakan bahwa klasifikasi adalah pembagian sesuatu menurut kelas-kelas (Lestari, 2016).

Klasifikasi dapat didefinisikan sebagai proses untuk menyatakan suatu objek data sebagai salah satu kategori atau kelas yang telah didefinisikan sebelumnya (Zaki, 2013). Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan kelas data, dengan tujuan untuk memprediksi atau memperkirakan kelas dari suatu objek yang labelnya belum diketahui. Untuk mencapai tujuan tersebut, maka proses klasifikasi membentuk model yang mampu membedakan data ke dalam kelas-kelas data yang berbeda berdasarkan aturan atau fungsi tertentu (Wijaya, 2014).

Pada klasifikasi, untuk membuat model dilakukan proses pelatihan dan *testing*. Proses pembelajaran secara otomatis terhadap suatu himpunan data dapat menghasilkan model klasifikasi (fungsi target) yang memetakan objek data ke salah satu kelas yang telah didefinisikan sebelumnya. Jadi, proses pembelajaran memerlukan masukan (*input*) berupa himpunan data latih (*training set*) yang berlabel (memiliki atribut kelas) dan menghasilkan keluaran (*output*) yang berupa sebuah model klasifikasi (Suyanto, 2017).

Klasifikasi teks merupakan suatu proses pengklasifikasian terhadap suatu data teks dengan membagi menjadi beberapa bagian kelas yang berbeda. Proses klasifikasi tersebut akan melalui beberapa tahapan *pre-processing* untuk mendapatkan data yang dibutuhkan oleh pengguna sehingga proses klasifikasi memiliki tingkat akurasi yang tinggi. Penggunaan klasifikasi pada teks biasanya dapat digunakan pada beberapa data.

### **2.3 Term Frequency-Inverse Document Frequency**

Algoritma *TF-IDF* adalah algoritma untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Algoritma ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan invers frekuensi dokumen yang mengandung kata tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum atau seberapa sering kata tersebut muncul. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi jika frekuensi sebuah kata tinggi di dalam sebuah dokumen, tetapi

frekuensi keseluruhan dokumen yang mengandung kata tersebut rendah pada satu set dokumen (Nurjannah dkk, 2013).

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)} \quad (2.1)$$

$$idf_t = \log\left(\frac{D}{df_t}\right) \quad (2.2)$$

$$W_{d,t} = tf_{d,t} \times IDF_{d,t} \quad (2.3)$$

Keterangan :

- $tf$  = banyak kata yang dicari pada seluruh dokumen
- $\max(tf)$  = jumlah kemunculan terbanyak *term* pada dokumen yang sama
- Nilai  $D$  = total dokumen
- $df_t$  = jumlah dokumen yang mengandung *term*
- $IDF$  = *Inverse Document Frequency*
- $d$  = dokumen ke-d
- $t$  = kata ke-t dari kata kunci
- $W$  = bobot dokumen ke-d terhadap kata ke-t

Dalam *TF-IDF*, kata-kata bisa direpresentasikan dengan sekumpulan *hash value* (angka). Konsep *term frequency* (TF) mengasumsikan sebuah dokumen sebagai sekumpulan kata (*bag of words*). Urutan kemunculan sebuah kata diabaikan, hanya jumlah kemunculan dari kata itu saja yang diperhitungkan. Sedangkan konsep *inverse document frequency* (IDF) adalah untuk mengurangi efek dari kata yang frekuensinya terlalu tinggi dalam kumpulan dokumen. Konsep dasarnya adalah menurunkan bobot

dari kata dengan frekuensi kolektif (frekuensi total kemunculan kata dalam semua dokumen) yang tinggi. Jadi semakin banyak kata tersebut dalam sekumpulan dokumen, maka semakin rendah bobotnya (Ryansyah dkk, 2017).

## 2.4 *Naïve Bayes Classifier*

### 2.4.1 *Teorema Bayes*

*Bayes* adalah teknik pengklasifikasian menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal dengan *Theorema Bayes*. Teorema tersebut dikombinasikan dengan *Naïve* yang mengasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naïve Bayes* mengasumsikan bahwa ada atau tidak ada ciri tertentu dari sebuah kelas dan tidak ada hubungannya dengan ciri dari kelas lainnya (Bustami, 2014). *Bayes* merupakan teknik prediksi berbasis probabilitas sederhana yang berdasar pada penerapan *Teorema Bayes* dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam *Naïve Bayes*, model yang digunakan adalah “model fitur independen” (Harijanto dkk, 2018).

Prediksi *bayes* dengan didasarkan pada *Teorema Bayes* dengan formula umum sebagai berikut :

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \quad (2.4)$$

Penjelasan dari formula tersebut adalah sebagai berikut :

$P(H/X)$  = Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis  $H$  terjadi jika diberikan bukti (*evidence*)  $X$  terjadi.

$P(X/H)$  = Probabilitas sebuah bukti  $X$  terjadi akan mempengaruhi hipotesis  $H$ .

$P(H)$  = Probabilitas awal (*priori*) hipotesis  $X$  terjadi tanpa memandang bukti apapun.

$P(X)$  = Probabilitas awal (*priori*) bukti  $X$  terjadi tanpa memandang hipotesis atau bukti yang lain.

Ide dasar dari aturan *Bayes* adalah bahwa hasil dari hipotesis atau peristiwa  $H$  dapat diperkirakan berdasarkan bukti  $X$  yang diamati. Beberapa hal penting dari aturan *Bayes* tersebut, yaitu :

1. Sebuah probabilitas awal atau *priori*  $H$  atau  $P(H)$  adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir  $H$  atau  $P(H/X)$  adalah probabilitas dari suatu hipotesis setelah bukti diamati.

#### 2.4.2 *Naïve Bayes* Untuk Klasifikasi

Dalam *data mining*,  $X$  adalah sebuah *tuple* atau objek data,  $H$  adalah hipotesis atau dugaan bahwa *tuple*  $X$  adalah kelas  $C$ . Secara spesifik, dalam masalah klasifikasi,  $P(H/X)$  dapat dihitung sebagai probabilitas bahwa hipotesis  $H$  benar untuk *tuple*  $X$  atau dengan kata lain  $P(H/X)$  adalah probabilitas bahwa *tuple*  $X$  berada dalam kelas  $C$ . Sementara itu,  $P(H)$  adalah probabilitas *prior* bahwa hipotesis  $H$  benar untuk setiap

*tuple* tidak peduli nilai-nilai atributnya sedangkan  $P(X)$  adalah probabilitas *prior* dari *tuple*  $X$ .

Untuk menggunakan Teorema *Bayes* dalam klasifikasi data, atau lebih spesifik lagi, untuk melakukan estimasi  $P(H/X)$  untuk menentukan probabilitas bahwa *tuple*  $X$  berada dalam kelas  $C$ , menggunakan metode *Naïve Bayes Classifier*, yang bekerja dengan lima langkah berikut :

1. Misalnya  $D$  adalah himpunan data latih (*training set*) yang berisi sejumlah *tuple* beserta label kelasnya. Setiap *tuple* adalah berdimensi  $n$  yang dinyatakan sebagai  $X = (x_1, x_2, \dots, x_n)$  yang didapat dari  $n$  atribut  $A_1, A_2, \dots, A_n$ .
2. Misalkan terdapat  $m$  kelas, yaitu  $C_1, C_2, \dots, \dots, C_m$ . Untuk sebuah *tuple* masukan  $X$ , *Naïve Bayes Classifier* memprediksi bahwa *tuple*  $X$  termasuk ke dalam kelas  $C_i$  jika dan hanya jika

$$P(C_i|X) > P(C_j|X) \text{ untuk } \leq j \leq m, j \neq i \quad (2.5)$$

Dengan kata lain, *Naïve Bayes Classifier* bekerja dengan cara memaksimalkan (mencari nilai maksimal)  $P(C_i|X)$ . Kelas  $C_i$  yang membuat  $P(C_i|X)$  bernilai maksimal disebut *maximum posterior hypothesis*. Dengan teorema *Bayes*,  $P(C_i|X)$  diestimasi menggunakan formula

$$P(C_i|X) = \frac{P(X|C_i) \times P(C_i)}{P(X)} \quad (2.6)$$

3. Mengingat  $P(X)$  bernilai sama untuk semua kelas (artinya *tuple*  $X$  memiliki probabilitas yang sama untuk masuk ke dalam kelas manapun), maka  $P(X/C_i)P(C_i)$  yang perlu dimaksimalkan. Jika probabilitas *prior* untuk setiap kelas tidak diketahui, maka probabilitas setiap kelas biasanya diasumsikan sama,  $P(C_1) = P(C_2) = \dots = P(C_m)$ . dengan demikian, *Naïve Bayes Classifier* hanya memaksimalkan  $P(X/C_i)P(C_i)$ . Misalnya, jika jumlah *tuple* di kelas  $C_1$  adalah 70% dan jumlah *tuple* di kelas  $C_2$  adalah 30%, maka dapat diberikan  $P(C_1) = 0,7$ , sedangkan  $P(C_2) = 0,3$ .
4. Jika dihadapkan pada himpunan data yang memiliki banyak atribut, kompleksitas perhitungan  $P(X/C_i)$  dapat direduksi dengan asumsi naif tentang independensi bersyarat kelas, yaitu nilai-nilai atribut saling independen, yang berarti antar atribut saling bebas, tidak ada ketergantungan sama sekali. Dengan demikian *Naïve Bayes Classifier* memaksimalkan

$$P(C_i|X) = \prod_{k=1}^n P(x_k|C_i) = P(x_1|C_i) \times \dots \times P(x_n|C_i) \quad (2.7)$$

Berdasarkan tuple-tuple pada himpunan data latih  $P(x_1/C_i), P(x_2/C_i), \dots, P(x_n/C_i)$  berdasarkan tipe atributnya.

Untuk atribut yang bernilai kategorial,  $P(x_k/C_i)$  didefinisikan sebagai jumlah *tuple* di kelas  $C_i$  dalam himpunan data  $D$  yang memiliki nilai  $x_k$  pada atribut  $A_k$  dibagi dengan jumlah semua *tuple* di kelas  $C_i$  dalam  $D$  yang disimbolkan  $|C_{i,D}|$ . Sementara itu, untuk atribut yang bernilai kontinu, yang umumnya diasumsikan memiliki distribusi Gaussian,  $P(x_k/C_i)$  didefinisikan sebagai

$$P(x_i|C_i) = \frac{1}{\sigma_{C_i} \sqrt{2\pi}} e^{-\frac{(x-\mu_{C_i})^2}{2\sigma_{C_i}^2}} \quad (2.8)$$

Dimana  $\mu_{C_i}$  dan  $\sigma_{C_i}$  adalah rata-rata dan deviasi standar dari nilai-nilai pada atribut  $A_k$  untuk kelas  $C_i$ .

5. Untuk memprediksi label kelas dari *tuple*  $X$ , harus menghitung probabilitas  $P(X|C_i)P(C_i)$  untuk setiap kelas  $C_i$ . Selanjutnya hanya perlu memaksimalkan probabilitas tersebut, yaitu mencari kelas  $C_i$  yang menghasilkan probabilitas  $P(X|C_i) P(C_i)$  maksimum sebagai kelas keputusan. Secara matematis, *tuple*  $X$  diberi label kelas  $C_i$  jika dan hanya jika

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \text{ untuk } 1 \leq j \leq m, j \neq i \quad (2.9)$$

Jika perhitungan probabilitas menghasilkan nilai 0 karena nilai atribut yang tidak pernah muncul dalam himpunan data latih, cara yang dapat digunakan adalah menambahkan satu *tuple* pura-pura dengan kelas yang belum pernah muncul untuk setiap atribut yang ada (Suyanto, 2018). Hal itu dilakukan untuk menghindari probabilitas bernilai 0.

*Naïve Bayes Classifier* untuk klasifikasi teks memperhitungkan frekuensi setiap kata yang muncul pada dokumen. Misalnya terdapat dokumen  $d$  dengan himpunan kelas  $c$ . Untuk memperhitungkan kelas dari dokumen  $d$ , maka dapat dihitung dengan rumus :

$$P(c|\text{dokumen } d) = P(c) \times P(x_1|c) \times P(x_2|c) \dots \times P(x_n|c) \quad (2.10)$$

$P(c)$  = Probabilitas prior dari kelas  $c$

- $x_n$  = Kata dokumen  $d$  ke- $n$   
 $P(c \mid \text{dokumen } d)$  = Probabilitas suatu dokumen termasuk kelas  $c$   
 $P(x_n \mid c)$  = Probabilitas kata ke- $n$  dengan diketahui kelas  $c$

Probabilitas *prior* kelas  $c$  ditentukan dengan rumus :

$$P(c) = \frac{N_c}{N} \quad (2.11)$$

- $N_c$  = Jumlah kelas  $c$  pada seluruh dokumen  
 $N$  = Jumlah seluruh dokumen

Sementara rumus *Naïve Bayes Classifier* yang digunakan dengan pembobotan kata *TF-IDF* adalah sebagai berikut :

$$P(x_n \mid c) = \frac{W_{ct} + 1}{(\sum W' \in V W'_{ct}) + B'} \quad (2.12)$$

- $W_{ct}$  = Nilai pembobotan *TFIDF* atau  $W$  dari *term*  $t$  di kelas  $c$   
 $\sum W' \in V W'_{ct}$  = Jumlah total  $W$  dari keseluruhan *term* yang berada di kelas  $c$   
 $B'$  = Jumlah  $W$  kata unik (nilai *idf* tidak dikali dengan *tf*) pada seluruh dokumen

## 2.5 Pengujian

### 2.5.1 *Black Box Testing*

*Black box testing* adalah metode pengujian yang berfokus pada pengujian persyaratan fungsional perangkat lunak, untuk mendapatkan serangkaian kondisi input yang sesuai dengan persyaratan fungsional suatu program. Metode *black box* digunakan untuk mengetahui apakah perangkat lunak sudah berfungsi dengan baik dan benar. Data uji disesuaikan dengan spesifikasi perangkat lunak. Data uji dijadikan *input*, dieksekusi pada perangkat lunak, kemudian keluaran atau *output* dari perangkat lunak akan dicek apakah telah sesuai dengan yang diharapkan (Setiawan, 2011).

*Black box testing* adalah pengujian yang dilakukan hanya mengamati hasil eksekusi melalui data uji dan memeriksa fungsional dari sistem. Dianalogikan seperti melihat suatu kotak hitam, kita hanya dapat melihat penampilan luarnya saja, tanpa tau ada apa dibalik bungkus hitamnya. Sama halnya dengan pengujian *black box*, yaitu dengan mengevaluasi *interface* dan fungsionalitasnya tanpa mengetahui apa sesungguhnya yang terjadi dalam proses detilnya (hanya mengetahui *input* dan *output*)

Terdapat kelebihan dan kekurangan pengujian dengan menggunakan *black box testing* (Jaya, 2018). Kelebihan penggunaan metode *black box testing* adalah pengujian tidak perlu memiliki pengetahuan tentang bahasa pemrograman tertentu, pengujian dilakukan dari sudut pandang pengguna, ini membantu untuk mengungkapkan ambiguitas atau inkonsistensi dalam spesifikasi persyaratan, *programmer* dan *tester* keduanya saling bergantung satu sama lain. Sedangkan kekurangan dari metode *black box testing* adalah uji kasus sulit disain tanpa spesifikasi yang jelas, kemungkinan

memiliki pengulangan tes yang sudah dilakukan oleh programmer, dan beberapa bagian *back end* tidak diuji sama sekali.

### 2.5.2 Akurasi

Sistem yang melakukan klasifikasi diharapkan dapat melakukan klasifikasi semua data set dengan benar. Oleh karena itu, sebuah sistem klasifikasi juga harus diukur kinerjanya dan keakuratannya. Umumnya cara mengukur kinerja klasifikasi menggunakan kinerja klasifikasi menggunakan matriks confusion (Prasetyo, 2014). Matriks confusion merupakan tabel yang mencatat hasil kerja klasifikasi.

**Tabel 2.2 Matriks Confusion Untuk Klasifikasi 2 Kelas**

$f_{ii}$		Kelas hasil klasifikasi ( $j$ )	
		Kelas = 1	Kelas = 0
Kelas asli ( $i$ )	Kelas = 1	$f_{11}$	$f_{10}$
	Kelas = 0	$f_{01}$	$f_{00}$

Tabel 2.2 adalah contoh matriks confusion, yang melakukan klasifikasi masalah biner (dua kelas) untuk dua kelas, misalnya kelas 0 dan 1. Setiap sel  $f_{ij}$  dalam matriks menyatakan jumlah data dari kelas  $i$  yang hasil klasifikasinya masuk ke kelas  $j$ . Misalnya  $f_{11}$  adalah jumlah data dalam kelas 1 yang secara benar diklasifikasi ke kelas 1, dan  $f_{10}$  adalah data dalam kelas 1 yang diklasifikasi secara salah ke dalam kelas 0.

Berdasarkan isi matriks confusion, maka dapat diketahui jumlah data dari masing-masing kelas yang diprediksi secara benar, yaitu  $(f_{11} + f_{00})$ , dan data yang

diklasifikasikan secara salah, yaitu  $(f_{10} + f_{01})$ . Dengan mengetahui jumlah data yang diklasifikasikan secara benar, maka dapat diketahui akurasi hasil klasifikasi, maka.

Untuk menghitung akurasi, dengan formula sebagai berikut :

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}} \quad (2.13)$$

$$\text{Akurasi} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (2.14)$$



## BAB III

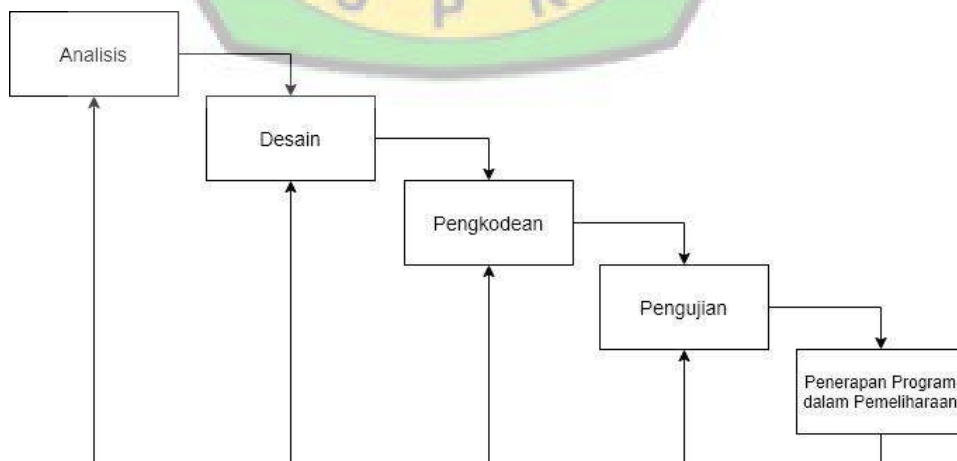
### METODOLOGI PENELITIAN

#### 3.1 Metode Pengumpulan Data

Metode pengumpulan data adalah teknik yang dilakukan untuk mengumpulkan atau memperoleh data (Hendryadi, 2014), yang dibutuhkan dalam penelitian, dan bagaimana data tersebut dikumpulkan. Metode pengumpulan data yang dilakukan dalam penelitian ini adalah studi kepustakaan dan literatur sejenis untuk mencari referensi dan teori yang berhubungan dengan penelitian untuk membuat pemecahan masalah.

#### 3.2 Metodologi Pengembangan Perangkat Lunak

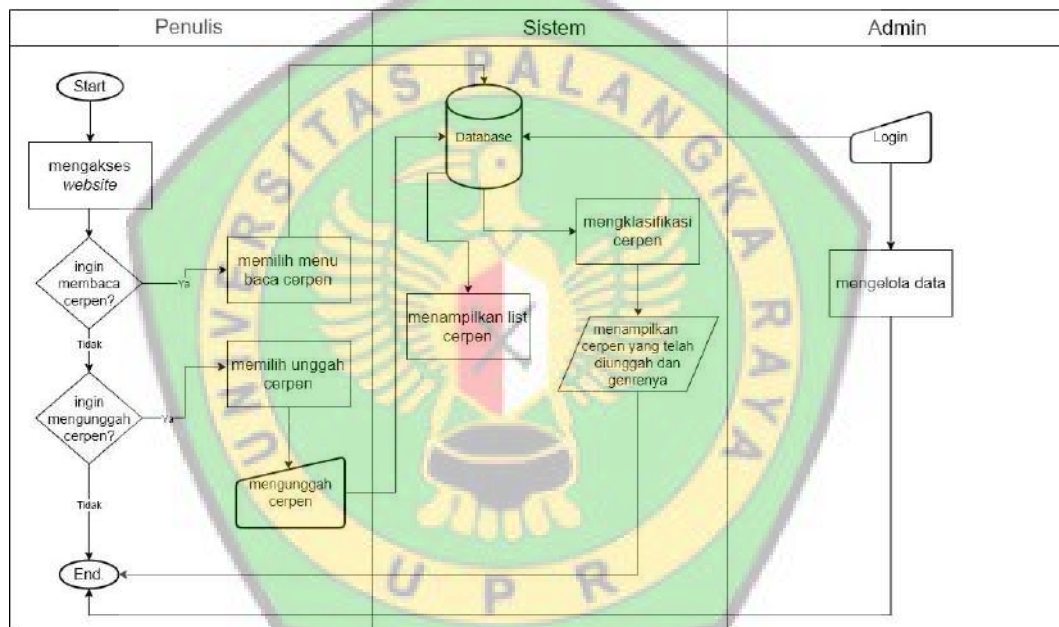
Dalam perancangan sistem pada penelitian ini, dilakukan langkah-langkah dalam pembangunan sistem menggunakan metode *waterfall* dengan tahapan-tahapan berikut:



**Gambar 3.1 Metode *Waterfall* Pressman (2010)**

### 3.2.1 Analisis

Tahap ini adalah tahap mengumpulkan kebutuhan dari sistem yang akan dibangun, memahami sistem, dan batasan sistem. Tahap ini dilakukan untuk mencari pemecahan masalah dan bagaimana membangun sebuah sistem yang dapat mengklasifikasi genre cerpen secara otomatis. Analisis sistem dibuat menggunakan menggunakan *flowchart*. *Flowchart* mendeskripsikan sistem yang dibuat dan menggambarkan bagaimana alur sistem.



**Gambar 3.2 Flowchart Sistem**

*Flowchart* di atas menggambarkan proses yang dalam sistem. Saat pengguna mengakses *website*, akan tampil halaman awal *website*, dengan penulis dan admin sebagai pengguna sistem. Jika penulis ingin membaca cerpen, maka penulis dapat mengakses daftar cerpen berdasarkan genrenya. Ketika penulis memilih salah satu cerpen, maka *website* akan menampilkan isi cerpen tersebut. Pada *website*, penulis juga dapat memilih unggah cerpen baru. Penulis dapat

langsung mengunggah cerpen ke *website*. Cerpen baru yang diunggah penulis akan diklasifikasikan secara otomatis oleh sistem.

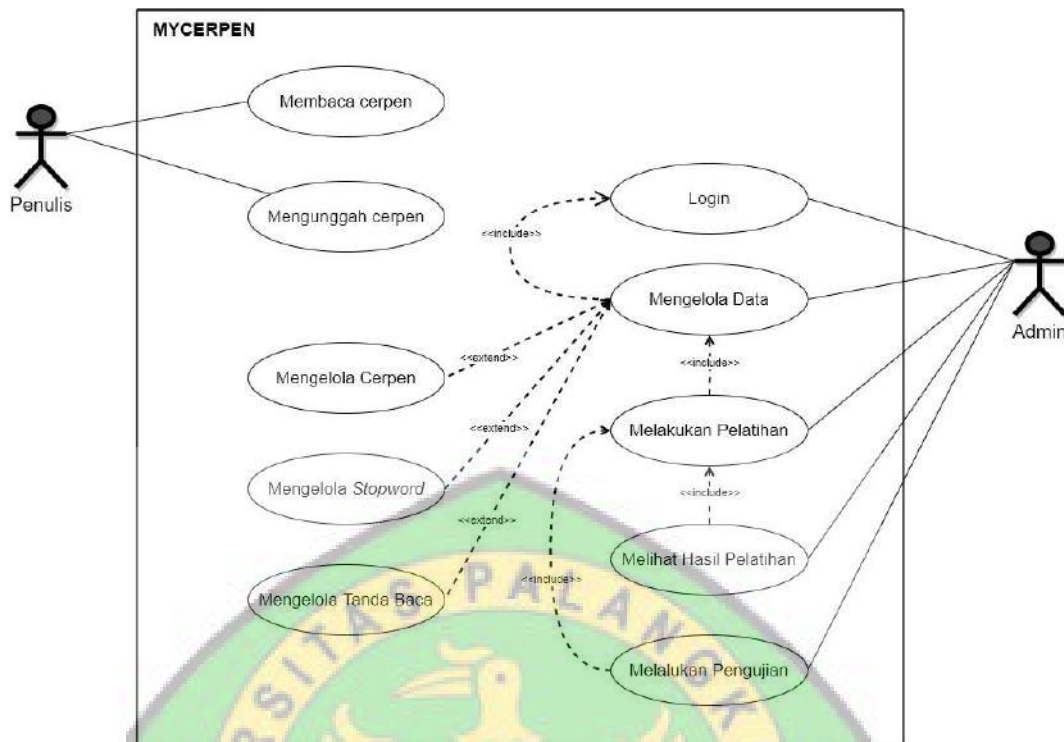
Dalam sistem yang dibuat, admin dapat mengelola data yang ada dalam *database*. Data-data yang dikelola admin adalah cerpen masuk, *stopword*, tanda baca, dan *dataset* pelatihan.

### 3.2.1 Desain

Tahap ini adalah tahap perancangan sistem dari hasil analisis kebutuhan yang didapat pada tahap analisis. Pada tahap ini dilakukan pemodelan sistem menggunakan *Unified Modelling Language* (UML), yaitu bahasa visual yang dilakukan untuk pemodelan pemrograman berorientasi objek.

#### 3.2.2.1 Use Case Diagram

*Use case diagram* menggambarkan fungsionalitas proses yang diharapkan terjadi dari sebuah sistem. *Use case* mempresentasikan sebuah interaksi antar aktor dengan sistem. *Use case* adalah pemodelan untuk kelakuan (*behavior*) sistem yang akan dibuat, mendeskripsikan interaksi aktor dengan sistem. *Use case* digunakan untuk mengetahui fungsi apa saja yang ada di dalam sistem dan siapa saja yang dapat menggunakan fungsi-fungsi tersebut (Rosa dan Shalahuddin, 2016).



**Gambar 3.3 Use Case**

Untuk mendefinisikan skenario dari penggunaan aplikasi ini, maka kegiatan pengguna dibagi menjadi beberapa kegiatan yang dijabarkan dalam tabel-tabel berikut.

**Tabel 3.1 Definisi Aktor**

No	Aktor	Deskripsi
1	Penulis dan Admin	Aktor yang akan menggunakan sistem

**Tabel 3.2 Definisi Use Case**

No	Use Case	Deskripsi
1	Penulis membaca cerpen	<i>Pre-condition :</i>

Tabel 3.2 Definisi *Use Case* (lanjutan)

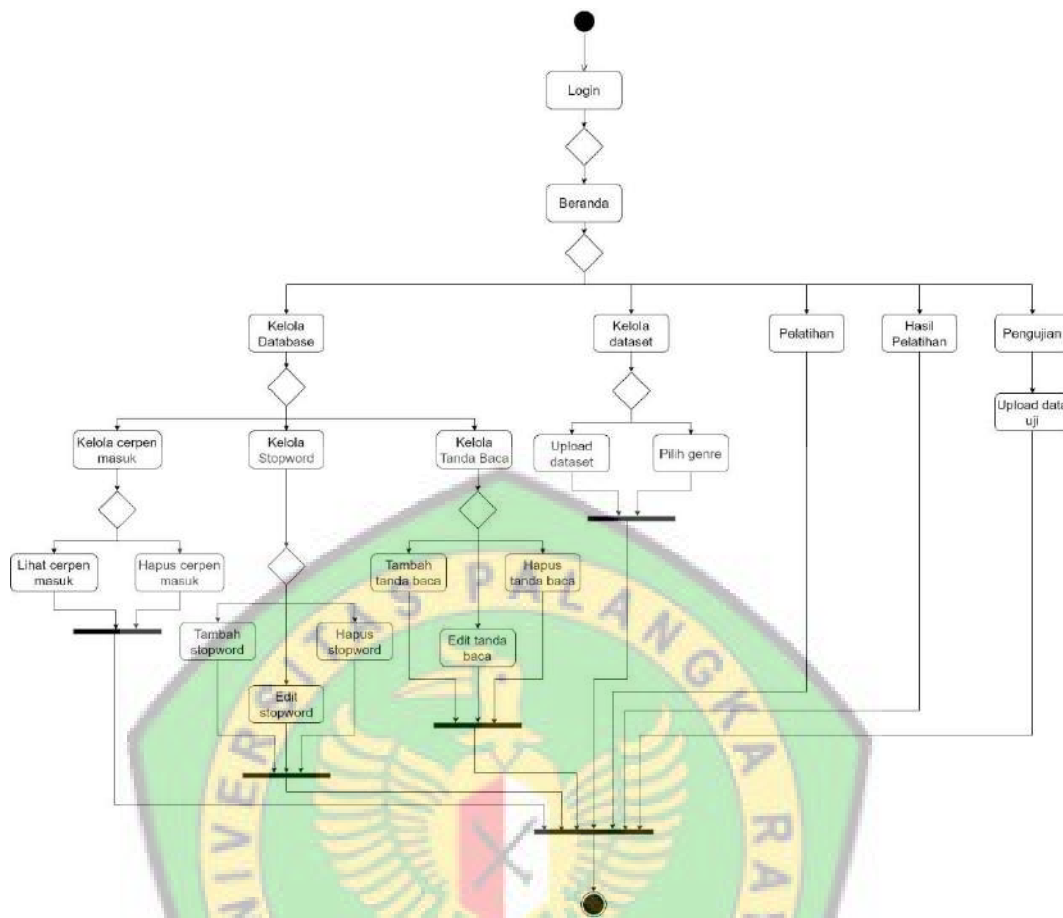
No	<i>Use Case</i>	Deskripsi
		<p>Penulis mengakses <i>website</i> dan pengguna memilih baca cerpen</p> <p><i>Process :</i></p> <p>Menampilkan halaman daftar cerpen</p>
2	Penulis mengunggah cerpen	<p><i>Pre-condition :</i></p> <p>Penulis mengakses <i>website</i> dan pengguna memilih unggah cerpen</p> <p><i>Process :</i></p> <p>Menampilkan halaman unggah cerpen</p>
3	Admin <i>login</i>	<p><i>Pre-condition :</i></p> <p>Admin mengakses <i>website</i>, admin harus <i>login</i> terlebih dahulu</p> <p><i>Process :</i></p> <p>Menampilkan halaman <i>login</i> admin</p>
4	Admin mengelola data	<p><i>Pre-condition :</i></p> <p>Admin mengelola <i>database</i></p> <p><i>Process :</i></p> <p>Admin mengelola <i>dataset</i>, data <i>stopword</i>, dan data tanda baca. Admin dapat mengelola cerpen yang diunggah Penulis</p>

Tabel 3.2 Definisi *Use Case* (lanjutan)

No	<i>Use Case</i>	Deskripsi
5	Admin melakukan pelatihan	<p><i>Pre-condition :</i></p> <p>Admin telah mengelola <i>database</i></p> <p><i>Process :</i></p> <p>Admin melakukan pelatihan menggunakan <i>dataset</i></p>
6	Admin melihat hasil pelatihan	<p><i>Pre-condition :</i></p> <p>Admin telah menyelesaikan pelatihan</p> <p><i>Process :</i></p> <p>Admin memilih melihat hasil klasifikasi</p>
7	Admin melakukan pengujian	<p><i>Pre-condition :</i></p> <p>Admin telah melakukan pelatihan</p> <p><i>Process :</i></p> <p>Admin melakukan pengujian terhadap hasil pelatihan, yaitu melakukan klasifikasi terhadap data baru</p>

### 3.2.2.2 Activity Diagram

*Activity diagram* atau diagram aktivitas adalah menggambarkan *workflow* (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis atau menu yang ada pada perangkat lunak. Diagram aktivitas menggambarkan aktivitas sistem (Rosa dan Shalahuddin, 2016).



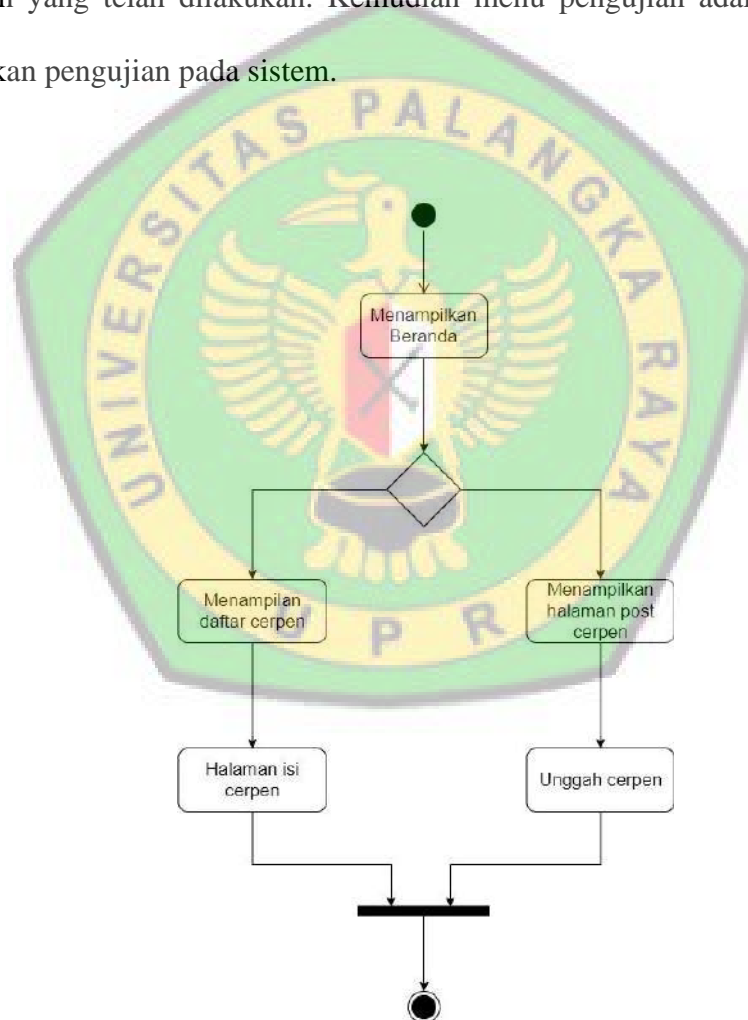
**Gambar 3.4 Activity Diagram Admin**

Gambar 3.4 adalah gambar urutan aktivitas sistem untuk pengguna admin, yaitu pengelompokan tampilan dari sistem atau *user interface* admin. Diagram aktivitas di atas menggambarkan rancangan menu yang ditampilkan untuk admin, terdapat menu kelola *database*, kelola *dataset*, pelatihan, hasil pelatihan dan pengujian.

Menu kelola *database* terbagi menjadi tiga sub menu, yaitu kelola cerpen masuk, kelola *stopword*, dan kelola tanda baca. Dalam sub menu kelola cerpen masuk, terdapat pilihan yang dapat dilakukan, yaitu lihat cerpen masuk hapus cerpen masuk. Dalam sub menu kelola *stopword* terdapat tiga pilihan yang dapat

dilakukan, yaitu tambah *stopword*, edit *stopword*, dan hapus *stopword*, begitu pula dengan sub menu kelola tanda baca yang terdapat tiga pilihan, yaitu tambah, edit, dan hapus tanda baca.

Pada kelola *dataset*, terdapat *upload dataset* dan pilih genre. Sedangkan menu pelatihan adalah dimana sistem melakukan pelatihan terhadap *dataset* yang telah dikelola pada kelola *dataset*, dan menu hasil pelatihan untuk melihat hasil dari pelatihan yang telah dilakukan. Kemudian menu pengujian adalah menu untuk melakukan pengujian pada sistem.



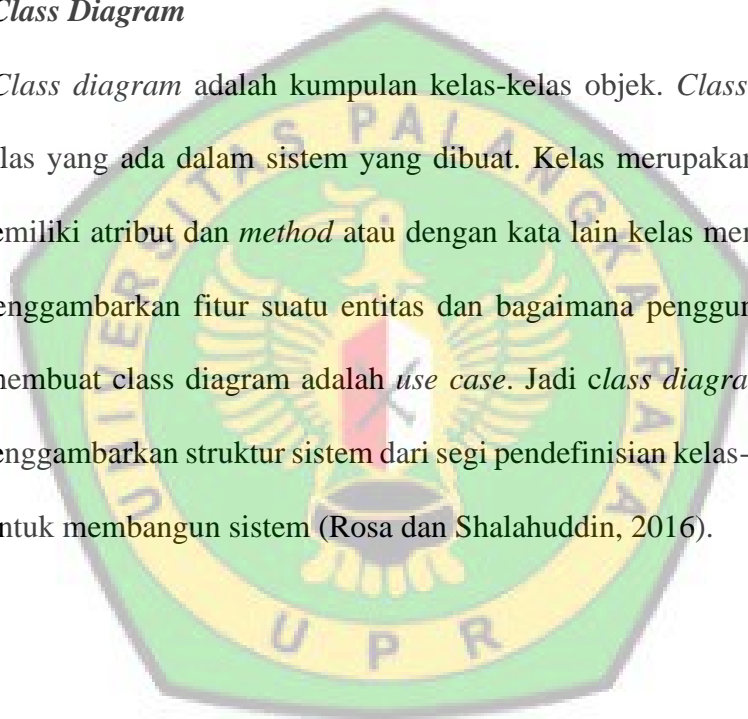
**Gambar 3.5 Activity Diagram Penulis**

Gambar 3.5 adalah gambar urutan aktivitas sistem untuk penulis. Pada saat pengguna penulis pertama kali mengakses *website*, sistem akan menampilkan

halaman utama, yaitu beranda. Kemudian terdapat menu daftar cerpen dimana sistem akan menampilkan daftar cerpen yang tersedia, jika memilih salah satu cerpen, sistem akan menampilkan isi cerpen tersebut. Sedangkan pada menu *post* cerpen, sistem akan menampilkan halaman yang disediakan untuk penulis mengunggah cerpen.

### 3.2.2.3 Class Diagram

*Class diagram* adalah kumpulan kelas-kelas objek. *Class diagram* berisi kelas-kelas yang ada dalam sistem yang dibuat. Kelas merupakan satu set objek yang memiliki atribut dan *method* atau dengan kata lain kelas memiliki informasi yang menggambarkan fitur suatu entitas dan bagaimana penggunaannya. Acuan dalam membuat class diagram adalah *use case*. Jadi *class diagram* atau diagram kelas menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem (Rosa dan Shalahuddin, 2016).

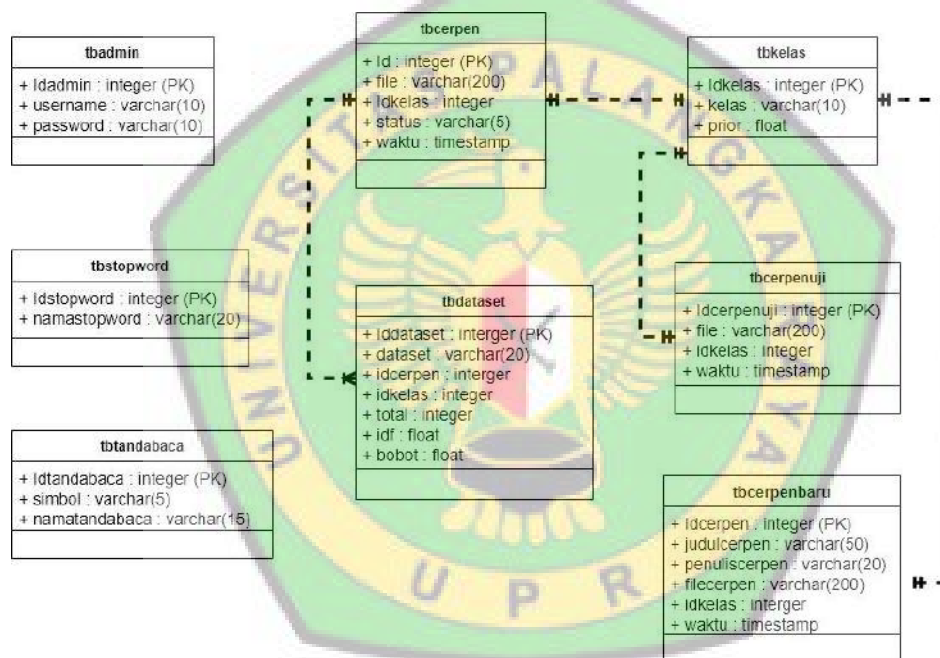




Gambar 3.6 adalah diagram kelas dari sistem yang akan dibuat. Pada gambar terdapat diagram *UI* untuk menampilkan sistem ke pengguna, dan diagram Tabel untuk data-data yang digunakan, dan kelas untuk proses.

### 3.2.2.4 Desain *Database*

Tabel untuk data-data yang digunakan, dan kelas untuk proses. Desain *database* yang digunakan dalam sistem, sebagai berikut.



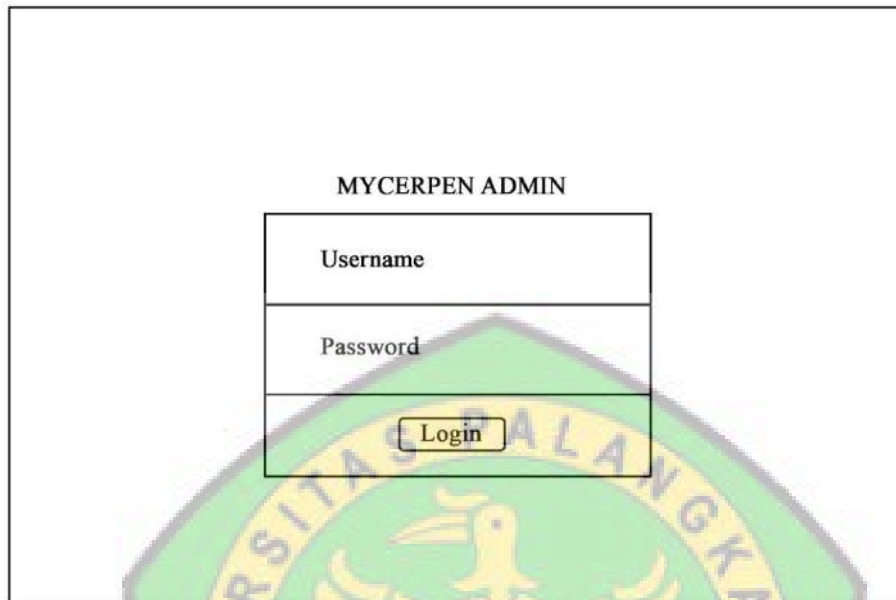
**Gambar 3.7 Desain *Database***

### 3.2.2.5 Desain *User Interface*

Desain *user interface* dibuat berdasarkan kebutuhan dan kenyamanan penulis dalam menggunakan *website*. Desain *user interface* menangani interaksi pengguna dengan sistem. Berikut adalah desai *user interface* yang dibuat.

## 1. Halaman Admin

### i. Halaman *Login*



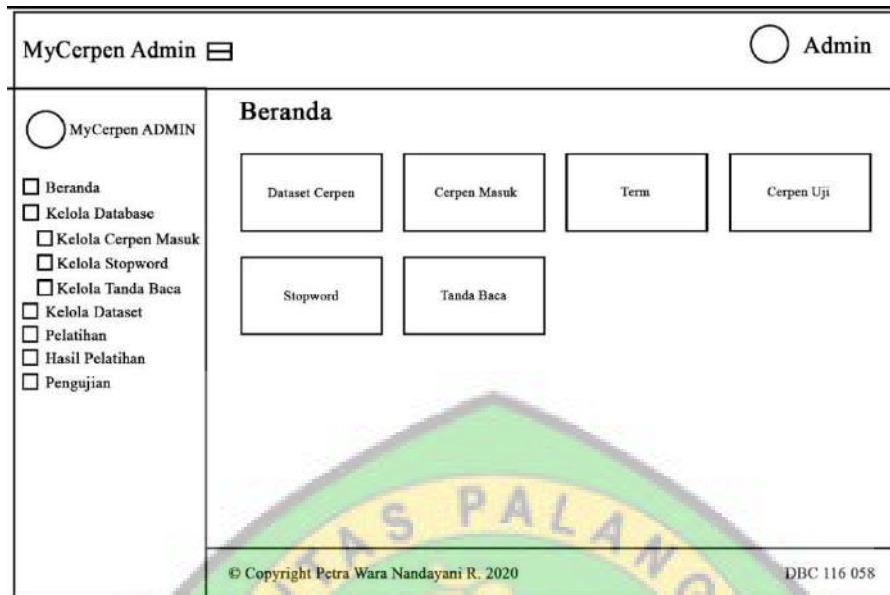
MYCERPEN ADMIN

Username
Password
<input type="button" value="Login"/>

**Gambar 3.8 Halaman *Login***

Pada halaman *login* admin akan memasukkan *username* dan *password* untuk masuk pada halaman admin. *Username* dan *password* yang dimasukkan adalah data yang sesuai dengan data yang sudah terdaftar.

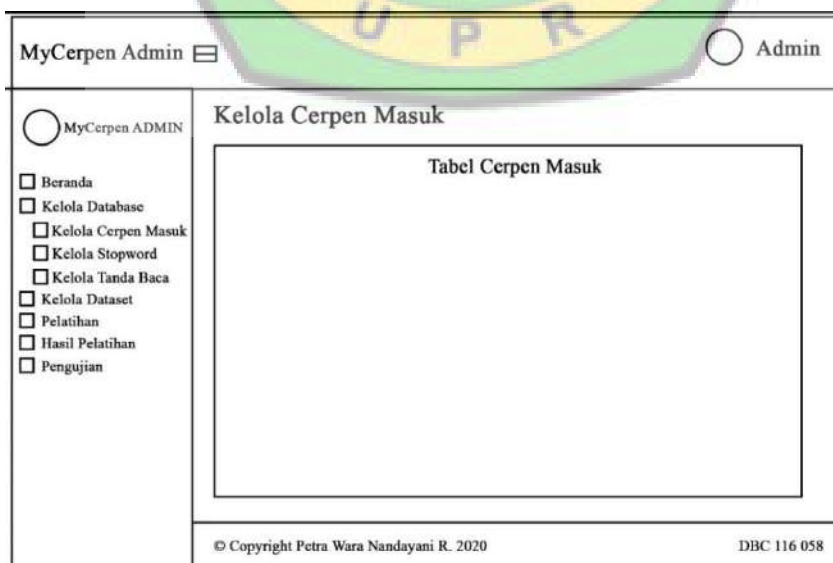
## ii. Beranda



**Gambar 3.9 Beranda**

Halaman ini adalah halaman utama yang akan tampil setelah admin berhasil *login*.

## iii. Halaman Kelola Cerpen Masuk



**Gambar 3.10 Halaman Kelola Cerpen Masuk**

Halaman kelola cerpen masuk adalah bagian dari menu kelola *database*. Pada halaman kelola cerpen masuk, admin dapat mengelola cerpen yang diunggah penulis. Admin dapat melihat dan menghapus cerpen.

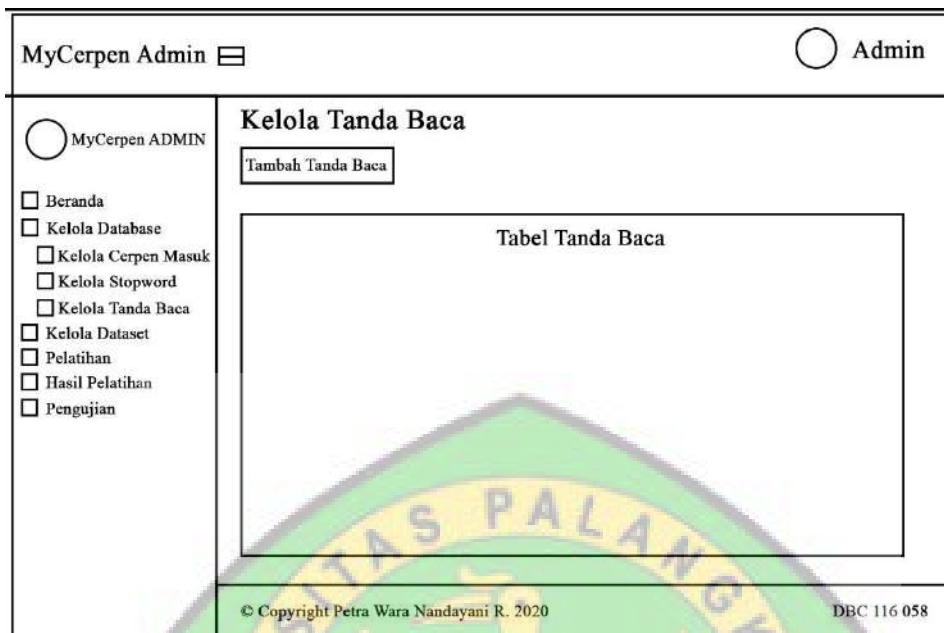
iv. Halaman Kelola *Stopword*



**Gambar 3.11 Halaman Kelola *Stopword***

Halaman kelola *stopword* adalah bagian dari menu kelola *database*. Pada halaman kelola *stopword*, admin dapat mengelola *stopword*. Admin dapat menambah, mengedit, dan menghapus *stopword*.

## v. Halaman Kelola Tanda Baca



**Gambar 3.12 Halaman Kelola Tanda Baca**

Halaman kelola tanda baca adalah bagian dari menu kelola *database*. Pada halaman kelola tanda baca, admin dapat mengelola tanda baca. Admin dapat menambah, mengedit, dan menghapus tanda baca.

vi. Halaman Kelola *Dataset*

MyCerpen Admin Admin

MyCerpen ADMIN

- Beranda
- Kelola Database
  - Kelola Cerpen Masuk
  - Kelola Stopword
  - Kelola Tanda Baca
- Kelola Dataset
- Pelatihan
- Hasil Pelatihan
- Pengujian

**Kelola Dataset Cerpen**

Upload Dataset

Tabel Dataset Cerpen

© Copyright Petra Wara Nandayani R. 2020 DBC 116 058

**Gambar 3.13 Halaman Kelola *Dataset***

Pada halaman kelola *dataset* cerpen, admin dapat melihat dan menambah *dataset* yang akan digunakan untuk pelatihan sistem.

vii. Halaman *Upload Dataset*

MyCerpen Admin Admin

MyCerpen ADMIN

- Beranda
- Kelola Database
  - Kelola Cerpen Masuk
  - Kelola Stopword
  - Kelola Tanda Baca
- Kelola Dataset
- Pelatihan
- Hasil Pelatihan
- Pengujian

**Upload Dataset Cerpen**

Choose File No file chosen

- Pilih Kelas - ▾

Simpan

© Copyright Petra Wara Nandayani R. 2020 DBC 116 058

**Gambar 3.14 Halaman *Upload Dataset***

Halaman *upload dataset* adalah halaman untuk mengunggah *dataset* cerpen yang akan digunakan untuk pelatihan sistem.

### viii. Halaman Pelatihan

MyCerpén Admin 		Admin 
<input type="radio"/> MyCerpén ADMIN <input type="checkbox"/> Beranda <input type="checkbox"/> Kelola Database <input type="checkbox"/> Kelola Cerpen Masuk <input type="checkbox"/> Kelola Stopword <input type="checkbox"/> Kelola Tanda Baca <input type="checkbox"/> Kelola Dataset <input type="checkbox"/> Pelatihan <input type="checkbox"/> Hasil Pelatihan <input type="checkbox"/> Pengujian	<b>Pelatihan</b> Proses Case Folding <input type="button" value="Mulai Pelatihan"/>	
		
© Copyright Petra Wara Nandayani R. 2020		DBC 116 058

**Gambar 3.15 Halaman Pelatihan**

MyCerpén Admin 		Admin 
<input type="radio"/> MyCerpén ADMIN <input type="checkbox"/> Beranda <input type="checkbox"/> Kelola Database <input type="checkbox"/> Kelola Cerpen Masuk <input type="checkbox"/> Kelola Stopword <input type="checkbox"/> Kelola Tanda Baca <input type="checkbox"/> Kelola Dataset <input type="checkbox"/> Pelatihan <input type="checkbox"/> Hasil Pelatihan <input type="checkbox"/> Pengujian	<b>Pelatihan</b> Proses Case Folding <input type="button" value="Mulai Pelatihan"/>	
	<input type="text" value="Progress bar"/> <input type="text" value="Proses Case Folding"/> <input type="text" value="Proses Tokenizing"/> <input type="text" value="Proses Filtering"/>	
© Copyright Petra Wara Nandayani R. 2020		DBC 116 058

**Gambar 3.16 Halaman Pelatihan**

Halaman pelatihan adalah halaman untuk proses pelatihan. Pada halaman pelatihan menampilkan hasil dari proses *case folding*, *tokenizing*, *filtering*, dan pembobotan *TF-IDF*.



ix. Halaman Hasil Pelatihan

The screenshot shows the 'MyCerpén Admin' web interface. At the top left, it says 'MyCerpén Admin' with a hamburger menu icon. At the top right, there is a user profile icon and the name 'Admin'. On the left side, there is a navigation menu with the following items: 'MyCerpén ADMIN', 'Beranda', 'Kelola Database' (with sub-items 'Kelola Cerpen Masuk', 'Kelola Stopword', 'Kelola Tanda Baca'), 'Kelola Dataset', 'Pelatihan', 'Hasil Pelatihan', and 'Pengujian'. The main content area is titled 'Hasil Pelatihan'. It contains four boxes, each labeled 'Genre Probabilitas Prior :'. Below these boxes is a large empty rectangular area labeled 'Tabel Kata'. At the bottom of the page, there is a copyright notice '© Copyright Petra Wara Nandayani R. 2020' and the ID 'DBC 116 058'. A large watermark of the Universitas Palangka Raya logo is overlaid on the page.

**Gambar 3.17 Halaman Hasil Pelatihan**

Pada halaman hasil pelatihan, admin dapat melihat hasil dari pelatihan *dataset*. Halaman ini menampilkan probabilitas prior dari setiap kelas dan sebuah tabel yang berisi pembobotan kata.

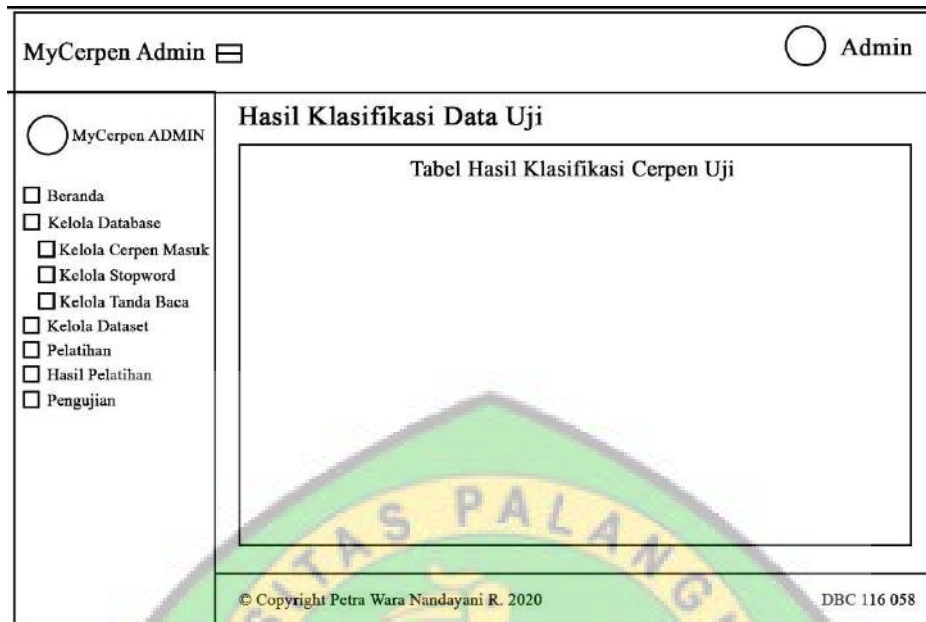
## x. Halaman Pengujian

MyCerpen Admin 		 Admin		
<input type="radio"/> MyCerpen ADMIN <input type="checkbox"/> Beranda <input type="checkbox"/> Kelola Database <input type="checkbox"/> Kelola Cerpen Masuk <input type="checkbox"/> Kelola Stopword <input type="checkbox"/> Kelola Tanda Baca <input type="checkbox"/> Kelola Dataset <input type="checkbox"/> Pelatihan <input type="checkbox"/> Hasil Pelatihan <input type="checkbox"/> Pengujian	<b>Pengujian</b>			
	<input type="button" value="Choose File"/>	<input type="text"/>	<input type="button" value="Uji"/>	<input type="button" value="Hasil Klasifikasi Data Uji"/>
	Proses Case Folding			
	Proses Tokenizing			
	Proses Filtering			
Naive Bayes Classifier				
© Copyright Petra Wara Nandayani R. 2020		DBC 116 058		

**Gambar 3.18 Halaman Pengujian**

Halaman pengujian adalah halaman untuk proses pengujian. Pada halaman ini, admin dapat menambahkan data yang akan diuji. Hasil klasifikasi akan ditampilkan pada halaman pengujian.

## xi. Halaman Hasil Klasifikasi Data Uji



**Gambar 3.19 Halaman Klasifikasi Data Uji**

Halaman hasil klasifikasi data uji adalah halaman untuk menampilkan hasil klasifikasi dari data uji, yaitu menampilkan data uji dan genre hasil klasifikasinya.

## 2. Halaman Penulis

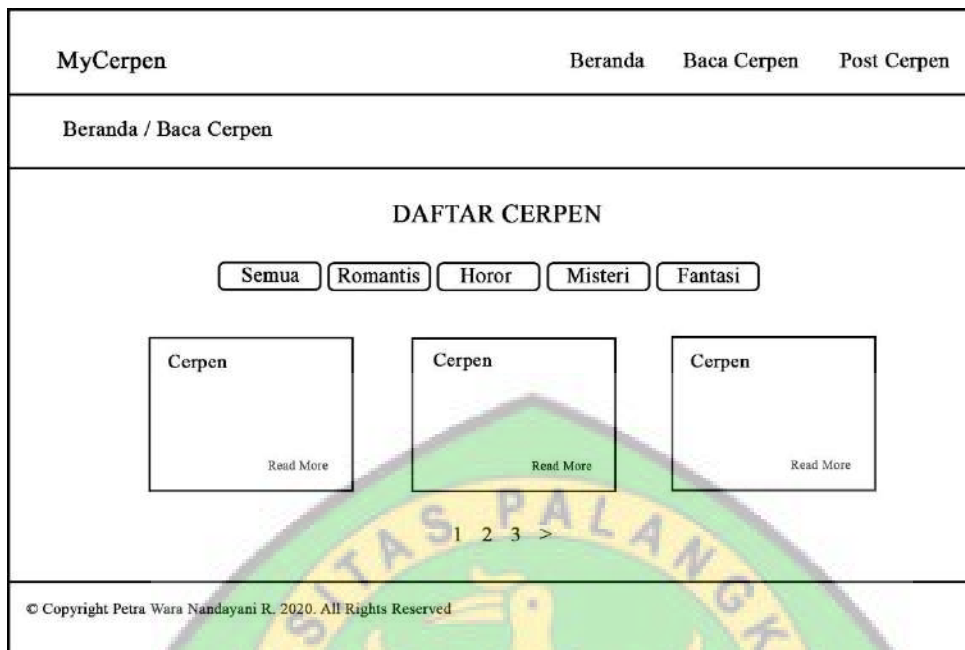
### i. Halaman Beranda



**Gambar 3.20 Halaman Beranda**

Halaman ini adalah halaman yang akan tampil ketika penulis pertama kali mengakses *website*.

## ii. Halaman Baca Cerpen



**Gambar 3.21 Halaman Baca Cerpen**

Halaman ini halaman yang menampilkan daftar cerpen. Penulis dapat memilih untuk melihat daftar cerpen sesuai genrenya.

## iii. Halaman Isi Cerpen

MyCerpén	<a href="#">Beranda</a> <a href="#">Baca Cerpen</a> <a href="#">Post Cerpen</a>
Beranda / Baca Cerpen Judul Cerpen	
<b>Judul Cerpen</b> isi cerpen ..... ..... ..... ..... ..... ..... ..... .....	<b>Cerpen Terbaru</b> Judul cerpen Judul cerpen Judul cerpen Judul cerpen
© Copyright Petra Wara Nandayani R. 2020. All Rights Reserved	

**Gambar 3.22 Halaman Isi Cerpen**

Halaman ini adalah halaman yang akan tampil saat penulis memilih sebuah cerpen untuk dibaca. Pada halaman ini akan ditampilkan isi cerpen yang dipilih.

#### iv. Halaman Post Cerpen

MyCerpen	Beranda	Baca Cerpen	Post Cerpen
Beranda / Post Cerpen			
<div style="border: 1px solid black; padding: 5px;"> <p>Upload Cerpen</p> <p>Judul <input type="text"/></p> <p>Penulis <input type="text"/></p> <p>Choose File <input type="button" value="No file chosen"/></p> <p><input type="button" value="Post Cerpen"/></p> </div>			
© Copyright Petra Wara Nandayani R. 2020. All Rights Reserved			

**Gambar 3.23 Halaman Post Cerpen**

Halaman ini adalah halaman yang dapat digunakan oleh penulis untuk mengunggah cerpen baru dan sistem akan mengklasifikasikan cerpen tersebut ke dalam empat kelas genre yang ada.

### 3.2.3 Pengkodean

Tahap ini adalah tahap mengimplementasikan rancangan sistem yang sudah dibuat, dan algoritma yang digunakan, ke dalam kode program untuk membangun sistem. Program ditulis menggunakan bahasa pemrograman *PHP* dan *database* menggunakan *MySQL*.

### 3.2.4 Pengujian

Tahap *testing* adalah tahap pengujian sistem secara keseluruhan. Terdapat dua jenis pengujian yang dilakukan dalam penelitian ini. Pengujian yang pertama adalah pengujian terhadap fungsionalitas sistem, untuk mengetahui apakah sistem dapat berjalan dan berfungsi dengan baik sebagaimana mestinya. Pengujian ini menggunakan metode *Black Box*.

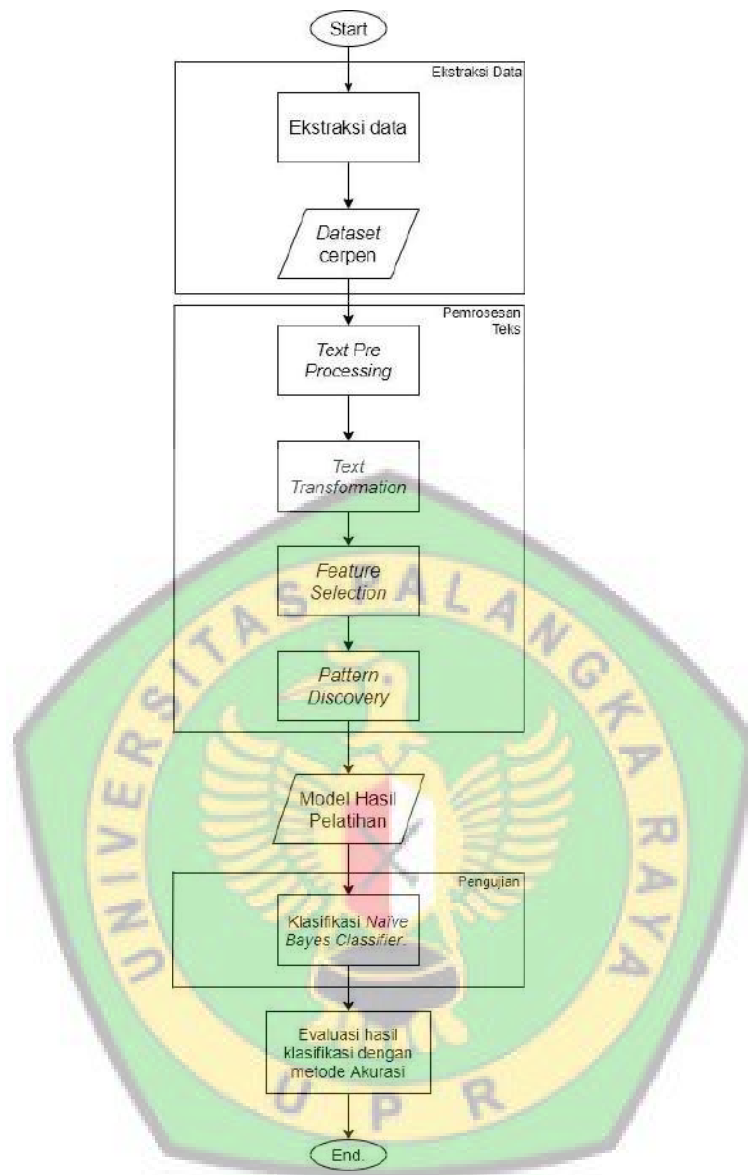
Pengujian yang kedua adalah pengujian terhadap performa klasifikasi menggunakan metode akurasi (*accuracy*) yang dibantu dengan matriks confusion (Prasetyo, 2014), yaitu untuk seberapa besar akurasi pengklasifikasian genre cerpen yang dilakukan oleh sistem dengan menghitung jumlah data yang berhasil diklasifikasikan dengan benar dibagi jumlah klasifikasi yang dilakukan, dengan rumus (2.10) dan (2.11).

### 3.2.5 Pemeliharaan

Tahap ini adalah tahap pemeliharaan sistem untuk menjaga proses operasional sistem, termasuk memperbaiki kesalahan yang ditemukan pada tahap sebelumnya. Tetapi dalam penelitian ini, hanya sampai tahap pengujian.

## 3.3 Metode Klasifikasi

Untuk melakukan klasifikasi genre cerpen dalam penelitian ini melalui beberapa tahap, yaitu ekstraksi data, pemrosesan teks yang meliputi tahap *text pre processing*, *text transformation*, *feature selection*, dan *pattern discovery*, dan pengujian.



**Gambar 3.24 Flowchart Tahap Klasifikasi**

### 3.3.1 Ekstraksi Data

Ekstraksi data adalah proses mengambil data dari sumber data, untuk digunakan dalam pemrosesan. *Dataset* cerpen yang digunakan dalam penelitian ini berjumlah 200 cerpen yang diambil dengan cara manual dan diperoleh dari beberapa situs publikasi cerpen, yaitu *cerpenmu.com*, *lakonhidup.com*,

*detectivestoryid.wordpress.com*, dan *www.fiksilotus.com*, dan *wattpad.com*. *Dataset* cerpen yang dikumpulkan adalah cerpen yang sudah dikategorikan atau dilabelkan oleh *website* sumber, sebagai pakar dalam penelitian ini. Berikut adalah pembagian *dataset* untuk setiap genre.

**Tabel 3.3 Tabel *Dataset***

<b>Genre</b>	<b>Jumlah <i>Dataset</i></b>
<b>Romantis</b>	50
<b>Horor</b>	50
<b>Misteri</b>	50
<b>Fantasi</b>	50

### 3.3.2 Pemrosesan Teks

Tahap pemrosesan dalam *text mining* yang dilakukan, yaitu pemrosesan awal terhadap teks (*text preprocessing*), transformasi teks (*text transformation*), seleksi fitur (*feature selection*), dan penemuan pola (*pattern discovery*) (Nugroho, 2011).

#### 3.3.2.1 *Text Pre-Processing*

*Text pre-processing* dilakukan untuk mengubah teks yang tidak terstruktur menjadi teks yang terstruktur sehingga teks dapat diproses. Tahapan dari *text pre-processing* adalah sebagai berikut:

i. *Case Folding*

*Case folding* dilakukan untuk mengkonversi atau mengubah keseluruhan teks di dalam dokumen menjadi bentuk standar, yaitu huruf kecil atau *lower case*.

Tahap *case folding* juga menghilangkan tanda baca dalam teks. Misalnya terdapat dokumen data latih. Berikut adalah contoh tahapan *case folding*.

**Tabel 3.4 Case Folding**

	<b>Kalimat Cerpen</b>	<b>Genre</b>
<b>D1</b>	<p>Lalu tiba-tiba saja aku merasa takut. Sangat takut. Aku merinding dan gemetar ketakutan. Aku meringkuk, menarik selimut hingga menutupi kepala. Aku juga merasakan keringat dinginku merembes pelan dari sekujur tubuh. Yang membuatku lebih takut lagi, aku mendengar suara serak, mirip bisikan dan erangan, memanggil-manggil namaku. Suara siapa itu?</p> <p><b>Setelah :</b></p> <p>lalu tiba tiba saja aku merasa takut. sangat takut aku merinding dan gemetar ketakutan aku meringkuk menarik selimut hingga menutupi kepala. aku juga merasakan keringat dinginku merembes pelan dari sekujur tubuh. yang membuatku lebih takut lagi aku mendengar suara serak, mirip bisikan dan erangan, memanggil manggil namaku suara siapa itu</p>	Horor

Tabel 3.4 Case Folding (lanjutan)

	Kalimat Cerpen	Genre
D2	<p>Kedua lelaki itu pun mulai panik dan takut. Bagaimana tidak takut? Di tempat itu, tak ada satu pun orang yang bisa dimintai bantuan. Sudah gelap, sepi lagi. Pasti banyak hantunya, nih. Pikir Roni sedikit gemetar.</p> <p><b>Setelah :</b></p> <p>kedua lelaki itu pun mulai panik dan takut. bagaimana tidak takut di tempat itu, tak ada satu pun orang yang bisa dimintai bantuan sudah gelapsepi lagi pasti banyak hantunya nih pikirnya sedikit gemetar</p>	Horor
D3	<p>“Apakah cinta datang disaat kenyamanan mulai hadir?”.</p> <p>“Cinta, kenyamanan?, saat aku melihatmu pertama kali, Aku langsung jatuh cinta padamu tanpa aku mendapatkan kenyamanan dahulu” ujarku.</p> <p><b>Setelah :</b></p> <p>apakah cinta datang disaat kenyamanan mulai hadir cinta, kenyamanan saat aku melihatmu pertama kali aku langsung jatuh cinta padamu tanpa aku mendapatkan kenyamanan dahulu ujarku</p>	Romantis

Tabel 3.4 Case Folding (lanjutan)

	Kalimat Cerpen	Genre
<b>D4</b>	<p>Sungguh malu rasanya ketika melihat tatapan matamu, aku tahu ada sesuatu yang berbeda di mata itu. Aku benar-benar yakin bahwa tatapan itu adalah benar, ya, kau jatuh cinta padaku. Dan aku pun begitu, aku jatuh cinta padamu.</p> <p><b>Setelah :</b></p> <p>sungguh malu rasanya ketika melihat tatapan matamu aku tahu ada sesuatu yang berbeda di mata itu. aku benar-benar yakin bahwa tatapan itu adalah benar ya kau jatuh cinta padaku dan aku pun begitu aku jatuh cinta padamu</p>	Romantis
<b>D5</b>	<p>Mereka memburuku. Aku dituduh membunuh diriku sendiri. Sungguh tuduhan tak masuk akal. Tapi mereka yakin ini memang pembunuhan yang sangat rapi: aku memotong-motong tubuhku sendiri, dan memasukkannya ke dalam kopor yang kini selalu ku tenteng ke mana pun pergi.</p> <p><b>Setelah :</b></p> <p>mereka memburuku aku dituduh membunuh diriku sendiri sungguh tuduhan tak masuk akal tapi mereka yakin ini memang pembunuhan yang sangat rapi aku memotong motong tubuhku sendiri dan memasukkannya ke dalam kopor yang kini selalu ku tenteng ke mana pun pergi</p>	Misteri

Tabel 3.4 *Case Folding* (lanjutan)

	Kalimat Cerpen	Genre
<b>D6</b>	<p>Kasus kali ini adalah kasus pembunuhan berantai di sebuah gereja kuno. “Inikah rasanya menjadi detektif?, menyelesaikan masalah, berkeliling kota, wah serunya...” gumanku dalam hati.</p> <p><b>Setelah :</b></p> <p>kasus kali ini adalah kasus pembunuhan berantai di sebuah gereja kuno inikah rasanya menjadi detektif menyelesaikan masalah berkeliling kota wah serunya gumamku dalam hati</p>	Misteri
<b>D7</b>	<p>Aku merasakan ada yang aneh. Aku seperti terkena hipnotis cahaya itu. Seperti ilusi, cahaya itu mendatangkan jutaan bahkan milyaran hasrat kenangan yang menggoda.</p> <p><b>Setelah :</b></p> <p>aku merasakan ada yang aneh aku seperti terkena hipnotis cahaya itu seperti ilusi cahaya itu mendatangkan jutaan bahkan milyaran hasrat kenangan yang menggoda</p>	Fantasi

Tabel 3.4 *Case Folding* (lanjutan)

	Kalimat Cerpen	Genre
<b>D8</b>	<p>Ada rintangan yang bertarung dengan raksasa. Setiap orang mendapatkan kekuatan dan mereka berhasil. Banyak rintangan yang sulit dihadapi. Lalu muncullah kakek misterius dan berkata “kalian akan bertarung dengan raja jika kalian berhasil kalian bisa keluar dari dunia mimpi ini, tapi jika kalian kalah kalian akan tinggal di sini selama-lamanya”.</p> <p><b>Setelah :</b></p> <p>ada rintangan yang bertarung dengan raksasa setiap orang mendapatkan kekuatan dan mereka berhasil banyak rintangan yang sulit dihadapi lalu muncullah kakek misterius dan berkata kalian akan bertarung dengan raja jika kalian berhasil kalian bisa keluar dari dunia mimpi ini tapi jika kalian kalah kalian akan tinggal di sini selama lamanya</p>	Fantasi

ii. *Tokenizing*

Pada tahap *tokenizing* dilakukan pemotongan *string* terhadap teks cerpen, berdasarkan setiap kata yang menyusunnya. Hasil dari pemotongan atau pemisahan tersebut dinamakan token. Berikut adalah hasil tahap *tokenizing* dari contoh dokumen pada tabel 3.4.

Tabel 3.5 *Tokenizing*

Token			
Horor	Romantis	Misteri	Fantasi
lalu	apakah	mereka	aku
tiba	cinta	memburuku	merasakan
tiba	datang	aku	ada
saja	disaat	dituduh	yang
aku	kenyamanan	membunuh	aneh
merasa	mulai	diriku	aku
takut	hadir	sendiri	seperti
sangat	cinta	sungguh	terkena
takut	kenyaman	tuduhan	hipnotis
aku	saat	tak	cahaya
merinding	aku	masuk	itu
dan	melihatmu	akal	seperti
gemetar	pertama	tapi	ilusi
ketakutan	kali	mereka	cahaya
aku	aku	yakin	itu
meringkuk	langsung	ini	mendatangkan
menarik	jatuh	memang	jutaan
selimut	cinta	pembunuhan	bahkan
hingga	padamu	yang	milyaran

Tabel 3.5 *Tokenizing* (lanjutan)

Token			
Horor	Horor	Horor	Horor
menutupi	tanpa	sangat	hasrat
kepala	mendapatkan	rapi	kenangan
aku	kenyamanan	aku	yang
juga	dahulu	memotong	menggoda
merasakan	ujarku	motong	ada
keringat	sungguh	tubuhku	rintangan
dinginku	malu	sendiri	yang
merembes	rasanya	dan	bertarung
pelan	ketika	memasukkannya	dengan
dari	melihat	ke	raksasa
sejujur	tatapan	dalam	setiap
tubuh	matamu	kopor	orang
kedua	aku	yang	mendapatkan
lelaki	tahu	kini	kekuatan
itu	ada	selalu	dan
pun	sesuatu	ku	meraka
mulai	yang	tentang	berhasil
panik	berbeda	ke	banyak
dan	di	mana	rintangan

Tabel 3.5 *Tokenizing* (lanjutan)

Token			
Horor	Horor	Horor	Horor
takut	mata	pun	yang
bagaimana	itu	pergi	sulit
tidak	aku	kasus	dihadapi
takut	benar	kali	lalu
di	benar	ini	munculah
tempat	yakin	adalah	kakek
itu	bahwa	kasus	misterius
tak	tatapan	pembunuhan	dan
ada	itu	berantai	berkata
satu	adalah	di	kalian
pun	benar	sebuah	akan
orang	ya	gereja	bertarung
yang	kau	kuno	dengan
bisa	jatuh	inikah	raja
dimintai	cinta	rasanya	jika
bantuan	padaku	menjadi	kalian
sudah	dan	detektif	berhasil
gelap	aku	menyelesaikan	kalian
sepi	pun	masalah	bisa

Tabel 3.5 *Tokenizing* (lanjutan)

Token			
Horor	Horor	Horor	Horor
lagi	begitu	berkeliling	keluar
pasti	aku	kota	dari
banyak	jatuh	wah	dunia
hantunya	cinta	serunya	ini
nih	padamu	gumamku	tapi
pikirnya		dalam	jika
sedikit		hati	kalian
gemetar			kalah
			kalian
			akan
			tinggal
			di
			sini
			selama
			lamanya

### 3.3.2.2 *Text Transformation*

Pada tahap *text transformasion* dilakukan *filtering*. *Filtering* adalah tahap mengambil kata-kata penting dari hasil *tokenizing* dan membuang kata-kata yang tidak penting, menggunakan *stopword*. *Stopword* adalah kata-kata yang tidak

memiliki pengaruh apapun dalam klasifikasi teks, seperti kata sambung, kata ganti orang, kata penunjuk waktu, dan kata tanya (Rahutomo dan Ariadi, 2019). *Stopword* yang digunakan dalam penelitian ini diambil dari [github.com/masdevid/ID-Stopwords](https://github.com/masdevid/ID-Stopwords) (Haryalesmana, 2016), dengan tambahan *stopword* yang diperlukan dalam penelitian ini, terdapat pada lampiran. Berikut hasil dari *filtering* token pada Tabel 3.4.

**Tabel 3.6 Filtering**

<b>Token</b>			
<b>Horor</b>	<b>Romantis</b>	<b>Misteri</b>	<b>Fantasi</b>
takut	cinta	memburuku	merasakan
takut	kenyamanan	dituduh	aneh
merinding	hadir	membunuh	terkena
gemetar	cinta	tuduhan	hipnotis
ketakutan	kenyamanan	akal	cahaya
meringkuk	melihatmu	pembunuhan	ilusi
menarik	jatuh	rapi	cahaya
selimut	cinta	tubuhku	mendatangkan
menutupi	mendapatkan	memotong	jutaan
merasakan	kenyamanan	motong	milyaran
keringat	malu	memasukkannya	hasrat
dinginku	rasanya	kopor	kenangan
merembes	melihat	tentang	menggoda

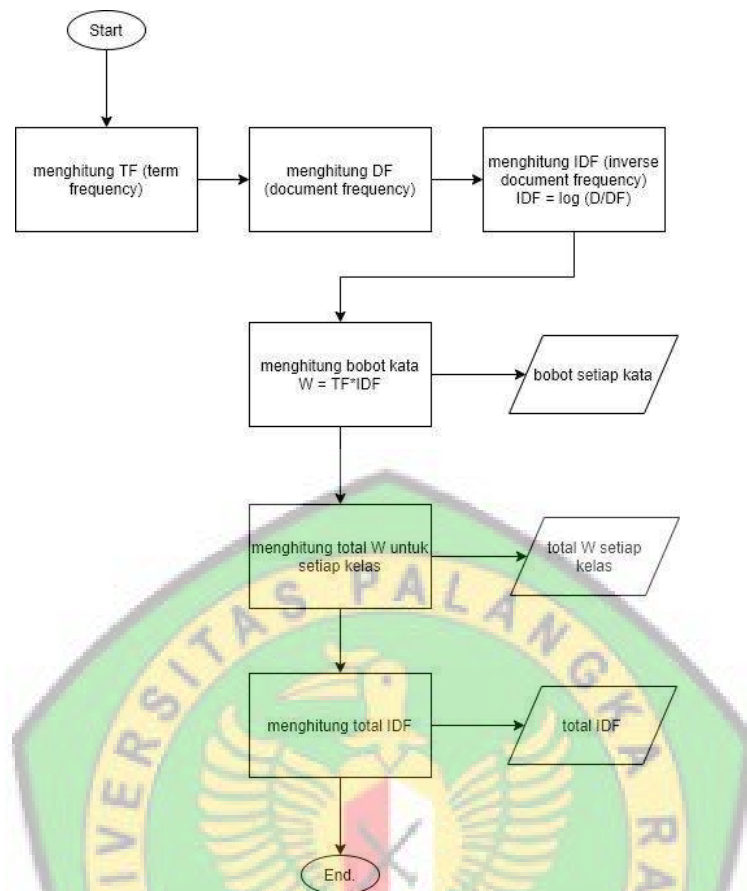
Tabel 3.6 *Filtering* (lanjutan)

Token			
Horor	Horor	Horor	Horor
sejujur	tatapan	kasus	rintangan
tubuh	matamu	kasus	bertarung
panik	mata	pembunuhan	raksasa
takut	tatapan	berantai	mendapatkan
takut	jatuh	gereja	kekuatan
dimintai	cinta	kuno	berhasil
bantuan	jatuh	rasanya	rintangan
gelap	cinta	detektif	dihadapi
sepi		menyelesaikan	muncullah
hantunya		berkeliling	misterius
pikirnya		kota	berkata
gemetar		gumamku	bertarung
		hati	raja
			berhasil
			keluar
			dunia
			kalah
			tinggal

### 3.3.2.3 Feature Selection

Tahap sebelumnya sudah melakukan penghapusan kata yang tidak deskriptif (*stopword*), karena tidak semua kata memiliki arti penting. Dalam *feature selection* akan diketahui *term* atau kata “unik”, yaitu kata yang dapat mewakili sekumpulan cerpen untuk setiap kelasnya.

Dari hasil *text preprocessing* dan *text transformation* telah didapatkan kata-kata yang dianggap dapat merepresentasikan kelasnya dari setiap dokumen. Namun bisa jadi kata tersebut juga muncul di dalam dokumen dalam kelas lainnya, hal itu menyebabkan kata tersebut tidak lagi bisa mewakili atau merepresentasikan kelasnya. Untuk mengatasi hal tersebut, maka dilakukan perhitungan bobot setiap *term* atau kata, dengan mengubah teks cerpen menjadi representasi numerik, dan mendapatkan nilai unik dari teks cerpen tersebut. Pada penelitian ini *feature selection* dilakukan dengan menggunakan algoritma pembobotan *TF-IDF*. *TF-IDF* adalah algoritma untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Dengan menerapkan algoritma *TF-IDF*, sistem akan menghitung frekuensi kemunculan sebuah kata dalam sebuah teks cerpen, dan invers frekuensi cerpen yang mengandung kata tersebut. Di bawah ini adalah tahapan perhitungan *TF-IDF*.



**Gambar 3.25 Flowchart Perhitungan TF-IDF**

Pertama, yang dilakukan adalah menghitung *term frequency* (*TF*), yaitu berapa kali atau seberapa sering sebuah kata muncul pada sebuah cerpen dengan kelas tertentu menggunakan perhitungan rumus (2.1). Kemudian menghitung *document frequency* (*DF*), yaitu jumlah dokumen yang mengandung kata tersebut. Setelah itu menghitung *inverse document frequency* (*IDF*) dengan rumus (2.2) pada Bab II, yaitu invers dari jumlah keseluruhan dokumen dibagi nilai *DF*. *IDF* menunjukkan hubungan jumlah kata dalam keseluruhan *dataset* cerpen. Jika sebuah kata memiliki frekuensi kemunculan yang tinggi dalam keseluruhan *dataset* cerpen, maka semakin rendah bobot kata tersebut. Jika bobot sebuah kata tinggi, maka kata tersebut adalah kata unik yang dapat mewakili atau merepresentasikan kelasnya.

Bobot tersebut kemudian dipakai dalam perhitungan probabilitas dalam klasifikasi *Naïve Bayes Classifier*. Berikut adalah tabel hasil pembobotan kata dari Tabel 3.5.



Tabel 3.7 Pembobotan *TF-IDF*

<i>Term</i>	Tf								df	D/df (D=8)	idf = log(D/df)
	D1	D2	D3	D4	D5	D6	D7	D8			
takut	2	2	0	0	0	0	0	0	2	4	0.602059991
merinding	1	0	0	0	0	0	0	0	1	8	0.903089987
gemetar	1	1	0	0	0	0	0	0	2	4	0.602059991
ketakutan	1	0	0	0	0	0	0	0	1	8	0.903089987
meringkuk	1	0	0	0	0	0	0	0	1	8	0.903089987
menarik	1	0	0	0	0	0	0	0	1	8	0.903089987
selimut	1	0	0	0	0	0	0	0	1	8	0.903089987
menutupi	1	0	0	0	0	0	0	0	1	8	0.903089987
merasakan	1	0	0	0	0	0	1	0	2	4	0.602059991
keringat	1	0	0	0	0	0	0	0	1	8	0.903089987
dinginku	1	0	0	0	0	0	0	0	1	8	0.903089987
merembes	1	0	0	0	0	0	0	0	1	8	0.903089987
sejujur	1	0	0	0	0	0	0	0	1	8	0.903089987
tubuh	1	0	0	0	0	0	0	0	1	8	0.903089987
panik	0	1	0	0	0	0	0	0	1	8	0.903089987
dimintai	0	1	0	0	0	0	0	0	1	8	0.903089987
bantuan	0	1	0	0	0	0	0	0	1	8	0.903089987
gelap	0	1	0	0	0	0	0	0	1	8	0.903089987
sepi	0	1	0	0	0	0	0	0	1	8	0.903089987
hantunya	0	1	0	0	0	0	0	0	1	8	0.903089987

Tabel 3.7 Pembobotan *TF-IDF* (lanjutan)

<i>Term</i>	Tf								df	D/df (D=8)	idf = log(D/df)
	D1	D2	D3	D4	D5	D6	D7	D8			
pikirnya	0	1	0	0	0	0	0	0	1	8	0.903089987
cinta	0	0	3	2	0	0	0	0	2	4	0.602059991
kenyamanan	0	0	3	0	0	0	0	0	1	8	0.903089987
hadir	0	0	1	0	0	0	0	0	1	8	0.903089987
melihatmu	0	0	1	0	0	0	0	0	1	8	0.903089987
jatuh	0	0	1	2	0	0	0	0	2	4	0.602059991
mendapatkan	0	0	1	0	0	0	0	1	2	4	0.602059991
malu	0	0	0	1	0	0	0	0	1	8	0.903089987
rasanya	0	0	0	1	0	1	0	0	2	4	0.602059991
melihat	0	0	0	1	0	0	0	0	1	8	0.903089987
tatapan	0	0	0	2	0	0	0	0	1	8	0.903089987
matamu	0	0	0	1	0	0	0	0	1	8	0.903089987
mata	0	0	0	1	0	0	0	0	1	8	0.903089987
memburuku	0	0	0	0	1	0	0	0	1	8	0.903089987
dituduh	0	0	0	0	1	0	0	0	1	8	0.903089987
membunuh	0	0	0	0	1	0	0	0	1	8	0.903089987
tuduhan	0	0	0	0	1	0	0	0	1	8	0.903089987
akal	0	0	0	0	1	0	0	0	1	8	0.903089987
pembunuhan	0	0	0	0	1	1	0	0	2	4	0.602059991
rapi	0	0	0	0	1	0	0	0	1	8	0.903089987

Tabel 3.7 Pembobotan *TF-IDF* (lanjutan)

<i>Term</i>	Tf								df	D/df (D=8)	idf = log(D/df)
	D1	D2	D3	D4	D5	D6	D7	D8			
tubuhku	0	0	0	0	1	0	0	0	1	8	0.903089987
memotong	0	0	0	0	1	0	0	0	1	8	0.903089987
motong	0	0	0	0	1	0	0	0	1	8	0.903089987
memasukkannya	0	0	0	0	1	0	0	0	1	8	0.903089987
kopor	0	0	0	0	1	0	0	0	1	8	0.903089987
tentang	0	0	0	0	1	0	0	0	1	8	0.903089987
kasus	0	0	0	0	0	2	0	0	1	8	0.903089987
berantai	0	0	0	0	0	1	0	0	1	8	0.903089987
gereja	0	0	0	0	0	1	0	0	1	8	0.903089987
kuno	0	0	0	0	0	1	0	0	1	8	0.903089987
detektif	0	0	0	0	0	1	0	0	1	8	0.903089987
menyelesaikan	0	0	0	0	0	1	0	0	1	8	0.903089987
berkeliling	0	0	0	0	0	1	0	0	1	8	0.903089987
kota	0	0	0	0	0	1	0	0	1	8	0.903089987
gumamku	0	0	0	0	0	1	0	0	1	8	0.903089987
hati	0	0	0	0	0	1	0	0	1	8	0.903089987
aneh	0	0	0	0	0	0	1	0	1	8	0.903089987
terkena	0	0	0	0	0	0	1	0	1	8	0.903089987
hipnotis	0	0	0	0	0	0	1	0	1	8	0.903089987
cahaya	0	0	0	0	0	0	2	0	1	8	0.903089987

Tabel 3.7 Pembobotan *TF-IDF* (lanjutan)

<i>Term</i>	Tf								df	D/df (D=8)	idf = log(D/df)
	D1	D2	D3	D4	D5	D6	D7	D8			
ilusi	0	0	0	0	0	0	1	0	1	8	0.903089987
mendatangkan	0	0	0	0	0	0	1	0	1	8	0.903089987
jutaan	0	0	0	0	0	0	1	0	1	8	0.903089987
milyaran	0	0	0	0	0	0	1	0	1	8	0.903089987
hasrat	0	0	0	0	0	0	1	0	1	8	0.903089987
kenangan	0	0	0	0	0	0	1	0	1	8	0.903089987
menggoda	0	0	0	0	0	0	1	0	1	8	0.903089987
rintangan	0	0	0	0	0	0	0	2	1	8	0.903089987
bertarung	0	0	0	0	0	0	0	2	1	8	0.903089987
raksasa	0	0	0	0	0	0	0	1	1	8	0.903089987
kekuatan	0	0	0	0	0	0	0	1	1	8	0.903089987
berhasil	0	0	0	0	0	0	0	2	1	8	0.903089987
dihadapi	0	0	0	0	0	0	0	1	1	8	0.903089987
muncullah	0	0	0	0	0	0	0	1	1	8	0.903089987
misterius	0	0	0	0	0	0	0	1	1	8	0.903089987
berkata	0	0	0	0	0	0	0	1	1	8	0.903089987
bertarung	0	0	0	0	0	0	0	1	1	8	0.903089987
raja	0	0	0	0	0	0	0	1	1	8	0.903089987
keluar	0	0	0	0	0	0	0	1	1	8	0.903089987
dunia	0	0	0	0	0	0	0	1	1	8	0.903089987



Tabel 3.8 Bobot Kata (lanjutan)

<i>Term</i>	$W = (tf*idf)$							
	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>	<b>D5</b>	<b>D6</b>	<b>D7</b>	<b>D8</b>
merembes	0.90309	0	0	0	0	0	0	0
sejujur	0.90309	0	0	0	0	0	0	0
tubuh	0.90309	0	0	0	0	0	0	0
panik	0	0.90309	0	0	0	0	0	0
dimintai	0	0.90309	0	0	0	0	0	0
bantuan	0	0.90309	0	0	0	0	0	0
gelap	0	0.90309	0	0	0	0	0	0
sepi	0	0.90309	0	0	0	0	0	0
hantunya	0	0.90309	0	0	0	0	0	0
pikirnya	0	0.90309	0	0	0	0	0	0
cinta	0	0	1.80618	1.20412	0	0	0	0
kenyamanan	0	0	2.70927	0	0	0	0	0
hadir	0	0	0.90309	0	0	0	0	0
melihatmu	0	0	0.90309	0	0	0	0	0
jatuh	0	0	0.60206	1.20412	0	0	0	0
mendapatkan	0	0	0.60206	0	0	0	0	0.60206
malu	0	0	0	0.90309	0	0	0	0
rasanya	0	0	0	0.60206	0	0.60206	0	0
melihat	0	0	0	0.90309	0	0	0	0
tatapan	0	0	0	1.80618	0	0	0	0

Tabel 3.8 Bobot Kata (lanjutan)

<i>Term</i>	$W = (tf*idf)$							
	D1	D2	D3	D4	D5	D6	D7	D8
matamu	0	0	0	0.90309	0	0	0	0
mata	0	0	0	0.90309	0	0	0	0
memburuku	0	0	0	0	0.90309	0	0	0
dituduh	0	0	0	0	0.90309	0	0	0
membunuh	0	0	0	0	0.90309	0	0	0
tuduhan	0	0	0	0	0.90309	0	0	0
akal	0	0	0	0	0.90309	0	0	0
pembunuhan	0	0	0	0	0.60206	0.60206	0	0
rapi	0	0	0	0	0.90309	0	0	0
tubuhku	0	0	0	0	0.90309	0	0	0
memotong	0	0	0	0	0.90309	0	0	0
motong	0	0	0	0	0.90309	0	0	0
memasukkannya	0	0	0	0	0.90309	0	0	0
kopor	0	0	0	0	0.90309	0	0	0
tentang	0	0	0	0	0.90309	0	0	0
kasus	0	0	0	0	0	1.80618	0	0
berantai	0	0	0	0	0	0.90309	0	0
gereja	0	0	0	0	0	0.90309	0	0
kuno	0	0	0	0	0	0.90309	0	0
detektif	0	0	0	0	0	0.90309	0	0



Tabel 3.8 Bobot Kata (lanjutan)

<i>Term</i>	$W = (tf*idf)$							
	<b>D1</b>	<b>D2</b>	<b>D3</b>	<b>D4</b>	<b>D5</b>	<b>D6</b>	<b>D7</b>	<b>D8</b>
berhasil	0	0	0	0	0	0	0	1.80618
dihadapi	0	0	0	0	0	0	0	0.90309
muncullah	0	0	0	0	0	0	0	0.90309
misterius	0	0	0	0	0	0	0	0.90309
berkata	0	0	0	0	0	0	0	0.90309
bertarung	0	0	0	0	0	0	0	0.90309
raja	0	0	0	0	0	0	0	0.90309
keluar	0	0	0	0	0	0	0	0.90309
dunia	0	0	0	0	0	0	0	0.90309
kalah	0	0	0	0	0	0	0	0.90309
tinggal	0	0	0	0	0	0	0	0.90309
<b>Kelas</b>	<b>Horor</b>	<b>Horor</b>	<b>Romantis</b>	<b>Romantis</b>	<b>Misteri</b>	<b>Misteri</b>	<b>Fantasi</b>	<b>Fantasi</b>
<b>Jumlah</b>	<b>12.34223</b>	<b>8.12781</b>	<b>7.52575</b>	<b>8.42884</b>	<b>11.43914</b>	<b>11.13811</b>	<b>11.43914</b>	<b>16.85768</b>

Tabel 3.7 berisi jumlah  $TF$ ,  $DF$  dan hasil perhitungan  $IDF$ .  $TF$  adalah frekuensi kemunculan sebuah kata dalam sebuah dokumen. Sedangkan  $DF$  adalah jumlah dokumen yang mengandung kata tersebut. Kemudian nilai  $IDF$  kata didapat dari hasil invers jumlah dokumen keseluruhan dibagi nilai  $DF$ .

Tabel 3.8 berisi nilai  $W$  atau bobot dari setiap kata. Nilai bobot didapatkan dari hasil perhitungan  $TF$  dikali  $IDF$ .

#### 3.3.2.4 *Pattern Discovery*

Tahap *pattern discovery* adalah tahap pelatihan sistem yang nantinya akan menghasilkan model klasifikasi. Algoritma pelatihan dan klasifikasi yang digunakan adalah algoritma *Naïve Bayes Classifier*. Data pelatihan berupa set fitur yang didapat dari nilai pembobotan kata yang dihasilkan dari perhitungan algoritma  $TF-IDF$  dengan kelas genre cerpen yang sudah ditentukan. Pelatihan ini nantinya akan menghasilkan model klasifikasi sebagai pengetahuan untuk sistem mengklasifikasikan genre cerpen.

Data cerpen yang digunakan sebagai data *training* adalah data yang sudah berlabel sesuai dengan kelas genrenya. Identifikasi cerpen mengacu pada isi cerita yang dibagi menjadi empat kelas, yaitu romantis, fantasi, horor, dan misteri.

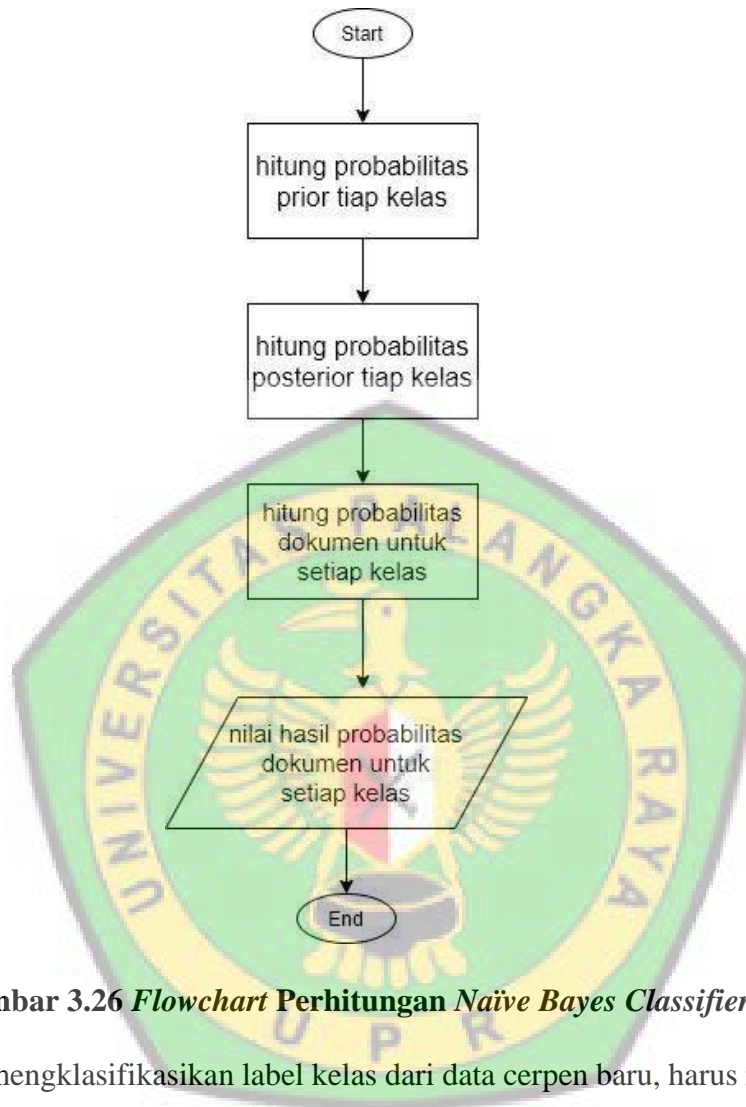
*Training* untuk membentuk model menggunakan algoritma *Naïve Bayes Classifier* adalah dengan menghitung probabilitas kemunculan kata untuk setiap kata menggunakan nilai bobot yang didapatkan dari perhitungan  $TF-IDF$ . Penghitungan probabilitas prior untuk setiap kelas menggunakan rumus (2.11) pada. Proses pelatihan

akan menghasilkan model klasifikasi yang dapat digunakan sebagai pengetahuan untuk mengklasifikasikan data baru dalam sistem. Kata yang telah melalui tahap *case folding*, *tokenizing*, dan *filtering*, dihitung bobotnya menggunakan algoritma pembobotan *TF-IDF*, disimpan kedalam *databaset* sebagai kamus kata. Kamus kata ini yang akan menjadi model klasifikasi yang akan digunakan dalam perhitungan klasifikasi *Naïve Bayes Classifier*. Kemudian tahap selanjutnya adalah perhitungan probabilitas prior untuk setiap kelas dan probabilitas posterior setiap kata untuk semua kelas, untuk klasifikasi *Naïve Bayes Classifier*.

### 3.4 Klasifikasi Cerpen

Proses klasifikasi cerpen oleh sistem menggunakan perhitungan dengan algoritma *Naïve Bayes Classifier*. Proses perhitungan melibatkan bobot kata, jumlah cerpen dalam tiap kelas dan jumlah keseluruhan data *training*.

Misalnya terdapat *dataset* cerpen yang digunakan untuk *training*. *Dataset* cerpen tersebut sudah diklasifikasikan secara manual ke dalam empat kelas genre, yaitu romantis, fantasi, horor, dan misteri. *Dataset* cerpen kemudian akan melalui tahap pemrosesan teks, yaitu *text preprocessing*, *text transformation*, *feature selection*, dan *pattern discovery*. Berdasarkan pembobotan kata menggunakan algoritma *TF-IDF*, bobot kata hasil pelatihan yang akan digunakan dalam perhitungan algoritma *Naïve Bayes* dalam proses klasifikasi.



**Gambar 3.26** *Flowchart Perhitungan Naïve Bayes Classifier*

Dalam mengklasifikasikan label kelas dari data cerpen baru, harus menghitung probabilitas  $P(x_n / c)P(c)$  untuk setiap kelas, menggunakan rumus (2.10). Misalnya terdapat data baru D9 sebagai berikut :

*“Aku hanya terdiam mencerna kata-katamu. Lalu kamu melanjutkan “keberadaanmu membuatku percaya, bahwa meski tanpa suara, meski tanpa ikatan, cinta masih bisa*

*tumbuh dan bertahan meski bertahun-tahun terabaikan. Cinta tak perlu diumumkan, cukup kita rasakan.”*

Setelah melalui tahap *case folding*, *tokenizing*, dan *filtering*, maka dari data baru D9 di atas akan didapat kata sebagai berikut :

**Tabel 3.10 Tabel Kata Contoh Klasifikasi**

terdiam	ikatan
mencerna	cinta
katamu	tumbuh
melanjutkan	bertahan
keberadaanmu	terabaikan
membuatku	cinta
percaya	diumumkan
suara	rasakan

Pada contoh data latih terdapat delapan dokumen dan empat kelas, yaitu horor, romantis, misteri, dan fantasi, dengan jumlah dua dokumen untuk masing-masing kelas. Maka perhitungan probabilitas untuk setiap kelas sebagai berikut.

1.  $P(\text{horor}) = 2/8 = 0.25$
2.  $P(\text{romantis}) = 2/8 = 0.25$
3.  $P(\text{misteri}) = 2/8 = 0.25$
4.  $P(\text{fantasi}) = 2/8 = 0.25$

Perhitungan probabilitas posterior, yaitu probabilitas kemunculan kata untuk setiap kelas menggunakan rumus (2.12). Bobot kata ( $W$ ) dan  $IDF$  digunakan dalam perhitungan probabilitas posterior untuk setiap kata dalam dokumen uji yang sudah melalui tahap *case folding*, *tokenizing*, dan *filtering*, yaitu menggunakan total bobot setiap kata pada masing-masing kelas, total  $IDF$ , dan total bobot setiap kelas. Berikut hasil perhitungan probabilitas posterior untuk contoh data uji. Bobot kata yang digunakan adalah bobot kata pada tabel 3.8.

1. Probabilitas *term* di kelas romantis

$$\begin{aligned}
 P(\text{terdiam} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142 \\
 P(\text{mencerna} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142 \\
 P(\text{katamu} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142 \\
 P(\text{melanjutkan} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142 \\
 P(\text{keberadaanmu} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142 \\
 P(\text{membuatku} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142 \\
 P(\text{suara} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142 \\
 P(\text{ikatan} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142 \\
 P(\text{cinta} \mid \text{romantis}) &= \frac{3.0103 + 1}{15,95459 + 71,64514} = 0,04578 \\
 P(\text{tumbuh} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142 \\
 P(\text{bertahan} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142 \\
 P(\text{terabaikan} \mid \text{romantis}) &= \frac{0 + 1}{15,95459 + 71,64514} = 0,01142
 \end{aligned}$$

$$P(\text{cinta} \mid \text{romantis}) = \frac{3.0103 + 1}{15,95459 + 71,64514} = 0,04578$$

$$P(\text{diumumkan} \mid \text{romantis}) = \frac{0 + 1}{15,95459 + 71,64514} = 0,01142$$

$$P(\text{rasakan} \mid \text{romantis}) = \frac{0 + 1}{15,95459 + 71,64514} = 0,01142$$

2. Probabilitas *term* di kelas horor

$$P(\text{terdiam} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{mencerna} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{katamu} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{melanjutkan} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{keberadaanmu} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{membuatku} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{suara} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{ikatan} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{cinta} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{tumbuh} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{bertahan} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{terabaikan} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{cinta} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{diumumkan} \mid \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

$$P(\text{rasakan} | \text{horor}) = \frac{0 + 1}{20,47004 + 71,64514} = 0,01086$$

3. Probabilitas *term* di kelas misteri

$$P(\text{terdiam} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{mencerna} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{katamu} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{melanjutkan} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{keberadaanmu} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{membuatku} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{suara} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{ikatan} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{cinta} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{tumbuh} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{bertahan} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{terabaikan} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{cinta} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{diumumkan} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

$$P(\text{rasakan} | \text{misteri}) = \frac{0 + 1}{22,57725 + 71,64514} = 0,01084$$

4. Probabilitas *term* di kelas fantasi

$$\begin{aligned}
 P(\text{terdiam} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{mencerna} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{katamu} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{melanjutkan} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{keberadaanmu} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{membuatku} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{suara} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{ikatan} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{cinta} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{tumbuh} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{bertahan} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{terabaikan} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{cinta} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{diumumkan} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005 \\
 P(\text{rasakan} \mid \text{fantasi}) &= \frac{0 + 1}{28,29682 + 71,64514} = 0,010005
 \end{aligned}$$

Untuk dapat menentukan kelas dari dokumen D9, harus menghitung probabilitas dari setiap kata dalam dokumen uji di atas untuk setiap kelasnya berdasarkan probabilitas yang dihasilkan pada proses *training*, yaitu kelas genre

romantis, horor, misteri dan fantasi, menggunakan rumus (2.10). Berikut perhitungan probabilitas kata.

$$1. P(d9 | romantis) = P(romantis) \times P(terdiam | romantis) \times P(mencerna | romantis) \times P(katamu | romantis) \times P(melanjutkan | romantis) \times P(keberadaan | romantis) \times P(membuatku | romantis) \times P(percaya | romantis) \times P(suara | romantis) \times P(ikatan | romantis) \times P(cinta | romantis) \times P(tumbuh | romantis) \times P(terabaikan | romantis) \times P(cinta | romantis) \times P(diumumkan | romantis) \times P(rasakan | romantis)$$

$$P(d9 | romantis) = 0,25 \times 0,01142 \times 0,01142 \times 0,01142 \times 0,01142 \times 0,01142 \times 0,01142 \times 0,01142 \times 0,01142 \times 0,01142 \times 0,04578 \times 0,01142 \times 0,01142 \times 0,04578 \times 0,01142 \times 0,01142$$

$$P(d9 | romantis) = 2,9 \times 10^{-29}$$

$$2. P(d9 | horor) = P(horor) \times P(terdiam | horor) \times P(mencerna | horor) \times P(katamu | horor) \times P(melanjutkan | horor) \times P(keberadaan | horor) \times P(membuatku | horor) \times P(percaya | horor) \times P(suara | horor) \times P(ikatan | horor) \times P(cinta | horor) \times P(tumbuh | horor) \times P(terabaikan | horor) \times P(cinta | horor) \times P(diumumkan | horor) \times P(rasakan | horor)$$

$$P(d9 | horor) = 0,25 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086 \times 0,01086$$

$$P(d9 | horor) = 8,6 \times 10^{-31}$$

$$3. P(d9 | misteri) = P(misteri) \times P(terdiam | misteri) \times P(mencerna | misteri) \times P(katamu | misteri) \times P(melanjutkan | misteri) \times P(keberadaan | misteri) \times P(membuatku | misteri) \times P(percaya | misteri) \times P(suara | misteri) \times P(ikatan | misteri) \times P(cinta | misteri) \times P(tumbuh | misteri) \times P(terabaikan | misteri) \times P(cinta | misteri) \times P(diumumkan | misteri) \times P(rasakan | misteri)$$

$$P(d9 | misteri) = 0,25 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084 \times 0,01084$$

$$P(d9 | misteri) = 8,3 \times 10^{-31}$$

$$4. P(d9 | fantasi) = P(fantasi) \times P(terdiam | fantasi) \times P(mencerna | fantasi) \times P(katamu | fantasi) \times P(melanjutkan | fantasi) \times P(keberadaan | fantasi) \times P(membuatku | fantasi) \times P(percaya | fantasi) \times P(suara | fantasi) \times P(ikatan | fantasi) \times P(cinta | fantasi) \times P(tumbuh | fantasi) \times P(terabaikan | fantasi) \times P(cinta | fantasi) \times P(diumumkan | fantasi) \times P(rasakan | fantasi)$$

$$P(d9 | fantasi) = 0,25 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005 \times 0,010005$$

$$P(d9 | fantasi) = 2,5 \times 10^{-31}$$

Dari perhitungan di atas nilai probabilitas untuk dokumen D9 yang terbesar adalah nilai probabilitas untuk kelas genre romantis, maka dapat diketahui bahwa dokumen D9 bergenre romantis.

### 3.5 Skenario Pengujian

Terdapat dua tahap pengujian yang akan dilakukan dalam penelitian ini, yaitu pengujian performa klasifikasi dengan metode akurasi (*accuracy*) dan pengujian fungsionalitas sistem menggunakan metode *black box testing*. Berikut adalah deskripsi skenario pengujian yang akan dilakukan.

#### 3.5.1 *Black Box Testing*

Pengujian terhadap fungsionalitas sistem dilakukan menggunakan metode *black box*. Pengujian dilakukan dengan mengevaluasi fungsionalitas sistem melalui *input* dan *output* (Setiawan, 2011). Sistem ini dirancang untuk dapat menampilkan cerpen yang telah diklasifikasi berdasarkan genrenya dan sistem dapat menjadi tempat untuk membaca cerpen. Dalam sistem dapat mengunggah cerpen baru, kemudian sistem secara otomatis akan mengklasifikasikan cerpen baru ke dalam kelas genrenya.

Dalam skenario pengujian sistem menggunakan *black box testing* yang akan dilakukan, yaitu dengan menguji beberapa fungsi di dalam sistem sebagai berikut :

- 1) Sistem dapat menampilkan cerpen berdasarkan genrenya,
- 2) Sistem dapat menampilkan isi cerpen, sehingga fungsi sistem sebagai tempat untuk membaca cerpen dapat berjalan dengan baik.
- 3) Fungsi *upload* atau unggah cerpen dalam sistem berfungsi dengan baik.

### 3.5.2 Akurasi

Pengujian menggunakan metode akurasi adalah pengujian untuk mengukur kinerja dan keakuratan sistem dalam mengklasifikasikan genre cerpen sesuai kelas genrenya. Kinerja klasifikasi dicatat menggunakan tabel matriks confusion (Prasetyo, 2014). Tabel matriks confusion berisi jumlah cerpen yang yang diklasifikasi secara benar ke dalam kelasnya dan jumlah cerpen yang klasifikasi secara salah ke dalam kelas lain.

**Tabel 3.11 Matriks Confusion**

Kelas		Kelas hasil klasifikasi			
		Romatis ( <i>r</i> )	Horor ( <i>h</i> )	Misteri ( <i>m</i> )	Fantasi ( <i>p</i> )
Kelas asli	Romatis ( <i>r</i> )	$f_{rr}$	$f_{rh}$	$f_{rm}$	$f_{rp}$
	Horor ( <i>h</i> )	$f_{hr}$	$f_{hh}$	$f_{hm}$	$f_{hp}$
	Misteri ( <i>m</i> )	$f_{mr}$	$f_{mh}$	$f_{mm}$	$f_{mp}$
	Fantasi ( <i>p</i> )	$f_{pr}$	$f_{ph}$	$f_{pm}$	$f_{pp}$

Dari tabel 3.1, misalnya  $f_{rr}$  adalah jumlah data cerpen romantis yang dipetakan dengan benar ke dalam kelas genre romantis,  $f_{rp}$  adalah jumlah data cerpen romantis yang dipetakan secara salah ke dalam kelas genre fantasi,  $f_{rh}$  adalah jumlah data cerpen romantis yang dipetakan secara salah ke dalam kelas genre horor, dan  $f_{rm}$  adalah jumlah data cerpen romantis yang dipetakan secara salah ke dalam kelas genre misteri, begitu seterusnya untuk data lainnya dalam tabel.

Berdasarkan isi matriks confusion, jumlah data cerpen yang diklasifikasikan secara benar ke dalam kelasnya adalah  $(f_{rr} + f_{pp} + f_{hh} + f_{mm})$ . Maka akurasi pengklafisikasian yang dilakukan oleh sistem dapat dihitung sebagai berikut :

(1) *Akurasi* =

$$\frac{f_{rr} + f_{pp} + f_{hh} + f_{mm}}{f_{rr} + f_{rp} + f_{rh} + f_{rm} + f_{pr} + f_{pp} + f_{ph} + f_{pm} + f_{hr} + f_{hp} + f_{hh} + f_{hm} + f_{mr} + f_{mp} + f_{mh} + f_{mm}}$$

Pengujian akan dilakukan beberapa kali dengan persentase data yang berbeda untuk menemukan pola data yang data yang optimal untuk membentuk model klasifikasi.



## BAB IV

### HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan hasil penelitian yang dilakukan. Tahapan yang dibahas, yaitu implementasi pengkodean dan pengujian sistem. Alamat *website* yang sudah *dihosting* untuk penelitian ini dapat diakses pada *www.mycerpen.com*.

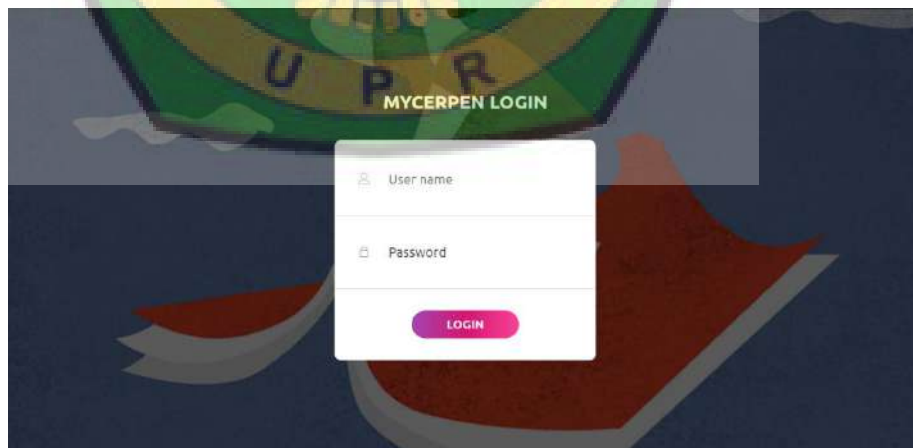
#### 4.1 Implementasi Pengkodean

Tahap implementasi adalah tahap mengimplementasikan rancangan sistem yang sudah dibuat, ke dalam kode program untuk membangun sistem. Program ditulis menggunakan bahasa pemrograman *PHP* dan *database* menggunakan *MySQL*.

##### 4.1.1 Implementasi Desain *User Interface*

##### 4.1.1.1 Implementasi Halaman Admin

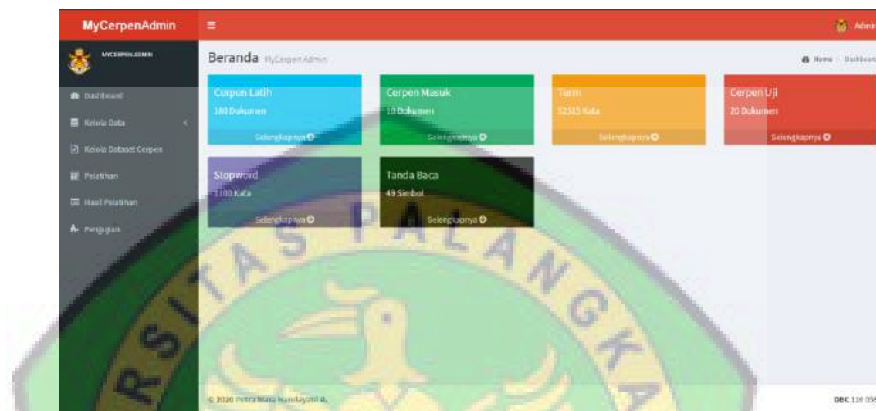
1. Implementasi Halaman *Login*



**Gambar 4.1 Halaman Login**

Halaman ini adalah tampilan halaman *login* admin. Admin harus memasukkan *username* dan *password* untuk dapat masuk ke halaman admin.

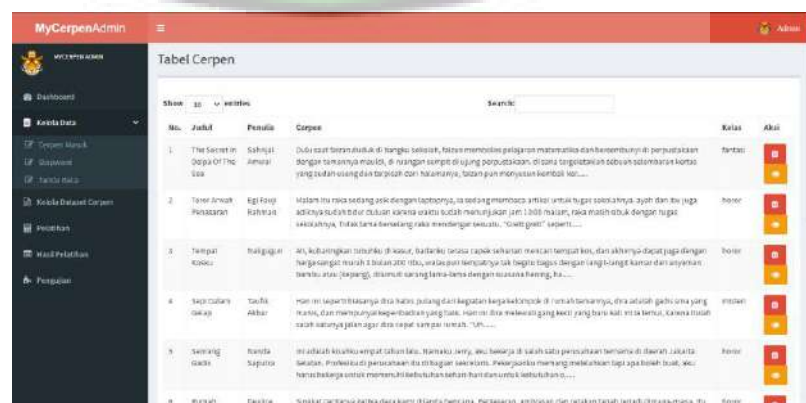
## 2. Implementasi Beranda



**Gambar 4.2 Halaman Beranda Admin**

Halaman ini adalah halaman yang akan tampil pertama kali saat admin berhasil *login*.

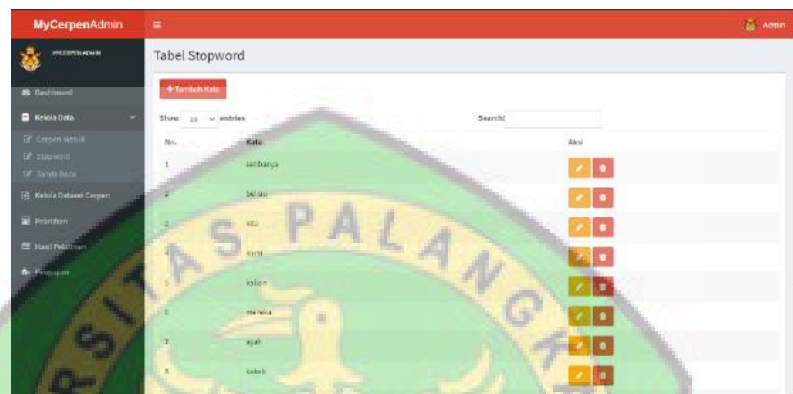
## 3. Implementasi Halaman Kelola Cerpen Masuk



**Gambar 4.3 Halaman Kelola Cerpen Masuk**

Halaman kelola cerpen masuk berisikan data cerpen masuk. Admin dapat mengelola cerpen masuk, admin dapat melihat dan menghapus cerpen masuk.

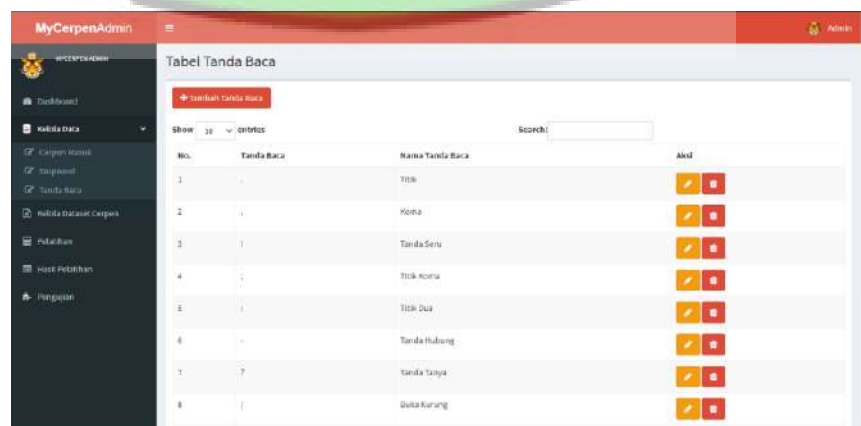
#### 4. Implementasi Halaman Kelola *Stopword*



**Gambar 4.4 Halaman Kelola *Stopword***

Halaman kelola *stopword* berisi daftar *stopword*. Pada halaman admin dapat menghapus, mengubah, dan menambah *stopword*.

#### 5. Implementasi Halaman Kelola Tanda Baca



**Gambar 4.5 Halaman Kelola Tanda Baca**



ditambahkan pada halaman ini, status latih *dataset* adalah ‘belum’. Pada halaman ini, admin dapat melabelkan *dataset*, sebelum *dataset* digunakan dalam pelatihan. Setelah pelatihan dilakukan, status latih *dataset* pada halaman kelola *dataset* cerpen akan berubah menjadi ‘sudah’.

No.	Judul	Cerpen	Kelas	Status Pelatihan	Aksi
1	Waktu Itu	DiTikidemi detik berlalu begitu lautan... Lanun panjang jam penuh di pengalangan bergelut... (text truncated)	romantis	sudah	[Edit] [Delete]
2	Tak Perinci	"Halo, namaiku Tony, saat ini aku berumur 20 tahun, dan aku belajar di salah satu Perusahaan swasta, aku monggo bagikan keseluruhan waktuku dengan keluarga, bekerja dan belajar, oh ya dan aku belajar bahasa Inggris, bahasa seperti bahasa sebelumnya, dan hari ini non sebelumnya. Katakanlah waktu dan waktu dia tidak benar, semoga berhasil..."	romantis	sudah	[Edit] [Delete]
3	Ngaji Cinta Ulang Tahun Pramuka	Demu gagas gemilla indah dalam gemilla api utangan yang menyala... (text truncated)	romantis	sudah	[Edit] [Delete]
4	Siapa yang menyanyi	Aku terbiasa saat ini sudah mempunyai lagu yang sudah terdapat dan sebuah lagu yang di telinganku... (text truncated)	romantis	sudah	[Edit] [Delete]

Gambar 4.8 Halaman Kelola *Dataset*

## 7. Implementasi Halaman *Upload Dataset*

Upload Dataset Cerpen

File Cerpen

\*) Minimal file harus 1000 karakter

\*\* Maksimal multiple upload 20 file

[Choose Files] No file chosen

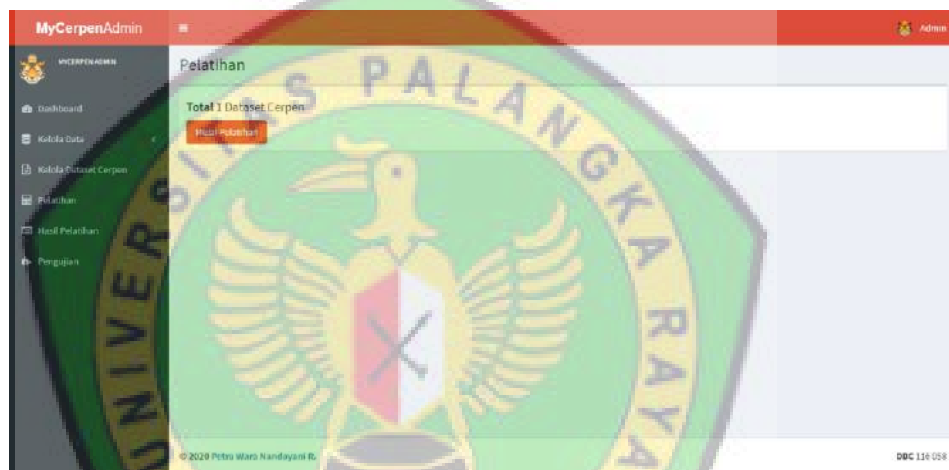
Pilih kelas: [dropdown menu]

[Simpan]

Gambar 4.9 Halaman *Upload Dataset*

Halaman ini adalah halaman yang digunakan untuk mengunggah *dataset* cerpen untuk pelatihan. Data yang diunggah berupa file dengan ekstensi *.txt* dan dapat mengunggah sampai 20 file sekaligus. Saat akan mengunggah *dataset*, admin memilih genre dari *dataset* tersebut.

#### 8. Implementasi Halaman Pelatihan

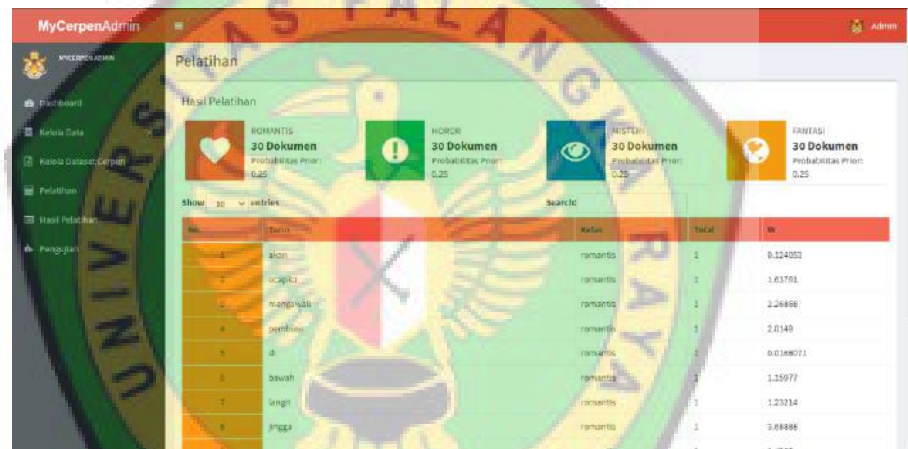


**Gambar 4.10 Halaman Pelatihan**



Halaman pelatihan adalah halaman yang menampilkan proses pelatihan *dataset*. Data yang digunakan adalah *dataset* cerpen dari halaman kelola *dataset*, yaitu *dataset* dengan status latihnya ‘belum’. Halaman pelatihan menampilkan hasil dari *case folding*, *tokenizing*, *filtering*, dan pembobotan *TF-IDF*.

### 9. Implementasi Halaman Hasil Pelatihan



**Gambar 4.12 Halaman Hasil Pelatihan**

Halaman ini adalah halaman yang berisi kamus kata yang dihasilkan dari pelatihan. Pada halaman ini ditampilkan probabilitas prior dari setiap kelas, kata hasil pelatihan, dan bobot katanya.

## 10. Implementasi Halaman Pengujian

The screenshot shows the 'MyCerpAdmin' interface with a sidebar menu on the left containing items like Dashboard, Kelola Data, Cerpem Masuk, Tindakan Baca, Kelola Dataset Cerpem, Pelatihan, Hasil Pelatihan, and Pengujian. The main content area is titled 'Pengujian' and includes a file upload section with 'Choose File' and 'No file chosen' buttons, and a 'Hasil Klaborasi Data Upl' button.

**Case Folding**  
 Show 10 entries | Search:   
 No. Case Folding  
 1 entah mengapa buku selalu melarangu untuk bermain di rumah kosong di seberang rumah kami. ibuku bilang bahwa di dalam rumah itu, terdapat monster penculik anak kecil yang suamuran dengkerku. tapi aku tidak percaya akan hal itu. aku melihat rumah itu dari jendela kamarku, tampak luar rumah tersebut terlihat seperti rumah biasa pada umumnya, dengan kondisi halaman depan agak berantakan oleh debu-dun kering. aku yakin buku hanya menakui supaya aku tak takut rumah kosong hannya. di saat ditunggu tidak ada di rumah, aku pun mencoba masuk ke rumah kosong itu sendirian. aku berjalan perlahan masuk ke dalam sebuah ruangan, seperti ruang tamu, aku melihat ada banyak bercak darah yang menghalasi dinding dengan diturangi pencahayaan redup, namun aku mengatakannya dan tetap melangkahkan kakiku ke ruangan lain. di ruang lain pun semakin banyak bercak darah di dinding bahkan di lantai berceceran darah, aku menemukan sebuah box menukang sedang di setiap pojokan ruangan, aku mencoba membuka box itu dan ternyata isinya organ-organ dalam manusia. astaga, untuk apa ini? aku bertanya ke ruangan berikutnya, masih dalam keadaan yang sama di sisi lain bercak darah, aku melihat pintu ruangan tersebut sedikit terbuka, aku melihat ruang di belakang sedang bermain di atas kasur dan aku mendengar suara tawa mereka bersama beberapa anak kecil. aku memperhatikan dari untuk membuka pintu kamar itu, betapa terkejutnya aku melihat itu dan ayahnya bermain bersama 2 anak kecil dengan tawa mereka yang riang. aku pun bertanya kepada mereka tentang hal itu. ibuku mengajaku keluar kamar bilapun buku mencarakan tentang ini kepadaku. aku tak mengerti, kata ibuku anak-anak kecil itu setiap sumber uang bagi keluarga kami. sumber uang? rupanya keluarga mereka menyembunyikan hal ini kepadaku. aku kira aku bukan anak fugganya mereka. ternyata ini adalah sebuah ide mereka untuk mencari penghasilan keluarga. semenjak hari itu, aku sering mengajak teman-temanku untuk bermain bersama. ya, tentu saja membantu orangtuaku, ini lah aku, seorang anak yang hidup di tengah keluarga dari hasil merampas organ teman-temanku. apapun akan aku lakukan untuk membantu orangtuaku.

**Tokenizing**  
 Show 10 entries | Search:   
 No. Tokenizing  
 1 mereka membunuh aku dibunuh membunuh diriku sendiri sungguh tuduhan tak masuk akal tapi mereka yakin ini memang pembunuhan yang sangat rapi aku memotong motong tubuhku sendiri dan memasukkannya ke dalam kopor yang satu-satunya kuperting ke mana pun pergi saja itu aku tak ekspektasikan mereka membunuhku melainkan diriku mereka ingin memastikan bahwa kopor yang satu ini memang berisi potongan tubuhku sendiri aku ingin menjelaskan tapi aku yakin mereka tak akan pernah percaya ceritaku suatu malam aku pulang sudah malam karena hujan turun bersama seorang kawan yang sehabis bertaruh tabung tak berhasil mendapat muncu ke kantor terus terting dia memang pacarku dia masih datang meski sehabis sudah ia baru saja seminggu jadi pada suaminya mati terbunuh itu yaitu ia dibunuh seseorang yang dulu pernah mengabdikan karena ia tahu suaminya suka menyiksa ketanya ia juga marah karena menegakkan dirinya lalu pingin suaminya basuh ia rumah pertama apa yang akan ia lakukan di siang tak tau dan hanya menyangka menyandakan kobatnya ke bahwa aku ke pergi dan aku pulang sampai di rumah aku mendengar suara sedang tidur bersama seorang laki-laki yang mengerjainya dia laki-laki yang suka menyiksa istrinya aku benar benar tak paham bagaimana suaminya terbuak menghantui laki-laki sebalik yang sering dengat laki-laki yang bahkan pada istri sendiri pun suka menukui aku tak ingin ini menjadi gosip merendah penduduk kota kubesakan semuanya dan aku pergi membawa kopor yang kini selalu kutentang ya aku mendatangi rumah perempuan yang tadi sore menemuku aku yakin ia masih mencintai tapi rumahnya terkonci betangga bilang perempuan itu sudah mati dibunuh sekitar sebulan lalu karena cekok dengan seorang perempuan yang ketahuan tidur dengan suaminya entah apa yang diceritakan istriku pada posai kini mereka membunuhku mereka ingin tahu di mana aku menyembunyikan potongan potongan tubuhku yang sampai saat ini tak pernah ditemukan lagi bistro februan.

Proses [memisahkan] setiap kata => mereka | membunuhku | aku | dibunuh | membunuh | diriku | sendiri | sungguh | tuduhan | tak | masuk | akal | tapi | mereka | yakin | ini | memang | pembunuhan | yang | sangat | rapi | aku | memotong | motong | tubuhku | sendiri | dan | memasukkannya | ke | dalam | kopor | yang | satu | satunya | kuperting | ke | mana | pun | pergi | saja | itu | aku | tak | ekspektasikan | mereka | membunuhku | melainkan | diriku | mereka | ingin | memastikan | bahwa | kopor | yang | satu | ini | memang | berisi | potongan | tubuhku | sendiri | aku | ingin | menjelaskan | tapi | aku | yakin | mereka | tak | akan | pernah | percaya | ceritaku | suatu | malam | aku | pulang | sudah | malam | karena | hujan | turun | bersama | seorang | kawan | yang | sehabis | bertaruh | tabung | tak | berhasil | mendapat | muncu | ke | kantor | terus | terting | dia | memang | pacarku | dia | masih | datang | meski | sehabis | sudah | ia | baru | saja | seminggu | jadi | pada | suaminya | mati | terbunuh | itu | yaitu | ia | dibunuh | seseorang | yang | dulu | pernah | mengabdikan | karena | ia | tahu | suaminya | suka | menyiksa | ketanya | ia | juga | marah | karena | menegakkan | dirinya | lalu | pingin | suaminya | basuh | ia | rumah | pertama | apa | yang | akan | ia | lakukan | di | siang | tak | tau | dan | hanya | menyangka | menyandakan | kobatnya | ke | bahwa | aku | ke | pergi | dan | aku | pulang | sampai | di | rumah | aku | mendengar | suara | sedang | tidur | bersama | seorang | laki | laki | yang | mengerjainya | dia | laki | laki | yang | suka | menyiksa | istrinya | aku | benar | benar | tak | paham | bagaimana | suaminya | terbuak | menghantui | laki | laki | sebalik | yang | sering | dengat | laki | laki | yang | bahkan | pada | istri | sendiri | pun | suka | menukui | aku | tak | ingin | ini | menjadi | gosip | merendah | penduduk | kota | kubesakan | semuanya | dan | aku | pergi | membawa | kopor | yang | kini | selalu | kutentang | ya | aku | mendatangi | rumah | perempuan | yang | tadi | sore | menemuku | aku | yakin | ia | masih | mencintai | tapi | rumahnya | terkonci | betangga | bilang | perempuan | itu | sudah | mati | dibunuh | sekitar | sebulan | lalu | karena | cekok | dengan | seorang | perempuan | yang | ketahuan | tidur | dengan | suaminya | entah | apa | yang | diceritakan | istriku | pada | posai | kini | mereka | membunuhku | mereka | ingin | tahu | di | mana | aku | menyembunyikan | potongan | potongan | tubuhku | yang | sampai | saat | ini | tak | pernah | ditemukan | lagi | bistro | februan

Showing 1 to 1 of 1 entries

Previous 1 Next

**Filtering**  
 Show 10 entries | Search:   
 No. Hasil Filtering  
 1 entah | ku | melarangu | bermain | di | rumah | kosong | di | seberang | rumah | ku | bilang | di | rumah | itu | monster | penculik | anak | kecil | yang | suamuran | denganku | tidak | percaya | akan | hal | itu | melihat | rumah | itu | jela | kamarku | tampak | luar | rumah | tersebut | terlihat | rumah | pada | umumnya | dengan | kondisi | halaman | depan | agak | berantakan | oleh | debu-dun | kering | yakin | ku | melihat | keluar | rumah | ke | sekitar | hannya | di | ku | tidak | ada | di | rumah | mencoba | masuk | ke | rumah | kosong | itu | an | berjalan | perlahan | masuk | ke | sebuah | ruangan | ruang | tamu | melihat | ada | banyak | bercak | darah | yang | menghalasi | dinding | dengan | diturangi | pencahayaan | redup | namun | mengatakannya | tetap | melangkahkan | kakiku | ke | ruangan | di | ruang | semakin | banyak | bercak | darah | di | dinding | di | lantai | berceceran | darah | menemukan | sebuah | box | berukuran | di | setiap | pojokan | ruangan | mencoba | membuka | box | itu | ternyata | isinya | organ | manusia | astaga | apa | itu | bertanya | ke | ruangan | itu | masih | keadaan | yang | sama | di | sisi | lain | bercak | darah | melihat | pintu | ruangan | tersebut | sedikit | terbuka | melihat | dewasa | bermain | di | atas | kasur | mengaj | suara | tawa | bersama | anak | kecil | membenarkan | dan | membuka | pintu | kamar | itu | terkejutnya | melihat | ku | bermain | bersama | anak | kecil | dengan | tawa | yang | riang | bertanya | tentang | hal | ku | mengajaku | keluar | kamar | itu | ku | menceri | tentang | mengerti | kata | ku | anak | kecil | itu | sumber | uang | bagi | keluarga | sumber | uang | rupanya | menyembunyikan | hal | ini | kepadaku | aku | kira | bukan | anak | fugganya | mereka | ternyata | ini | adalah | sebuah | ide | mereka | mencari | penghasilan | keluarga | semenjak | hari | itu | sering | mengajak | teman | temanku | bermain | bersama | ya | tentu | saja | membantu | ku | lah | anak | yang | hidup | di | tengah | keluarga | hasil | merampas | organ | teman | temanku | organ | membantu | ku |

Showing 1 to 1 of 1 entries

Previous 1 Next

Gambar 4.13 Halaman Pengujian

Perhitungan NBC

Show: 10 entries Search:

No.	Rumus	Perhitungan				
1	• Probabilitas Setiap Kelas Genre $P(c) = N_c / N$	<table border="0"> <tr> <td>Genre Romantis</td> <td>Genre Horor</td> </tr> <tr> <td>• <math>P(\text{Romantis}) = 45/180 = 0.25</math> • Probabilitas sebuah cerpen bergenre: romantis = 25%</td> <td>• <math>P(\text{Horor}) = 45/180 = 0.25</math> • Probabilitas sebuah cerpen bergenre: horor = 25%</td> </tr> </table>	Genre Romantis	Genre Horor	• $P(\text{Romantis}) = 45/180 = 0.25$ • Probabilitas sebuah cerpen bergenre: romantis = 25%	• $P(\text{Horor}) = 45/180 = 0.25$ • Probabilitas sebuah cerpen bergenre: horor = 25%
Genre Romantis	Genre Horor					
• $P(\text{Romantis}) = 45/180 = 0.25$ • Probabilitas sebuah cerpen bergenre: romantis = 25%	• $P(\text{Horor}) = 45/180 = 0.25$ • Probabilitas sebuah cerpen bergenre: horor = 25%					
2	• Probabilitas Kata Pada Setiap Kelas Genre $P(X c) = (W_{ct} + 1) / ((\sum W_{c'v} + B))$	<table border="0"> <tr> <td>Genre Romantis</td> <td>Genre Horor</td> </tr> <tr> <td>           P(ortah Romantis) =  <math>(22.6317 + 1) / (50308.602 + 188456.0096) = 8.6634E-5</math>             P(iu Romantis) =  <math>(16.4151 + 1) / (52066.602 + 188456.0096) = 7.11304E-5</math>             P(mearang Romantis) =  <math>(3.1239 + 1) / (50308.602 + 188456.0096) = 1.8630E-5</math>             P(ermis Romantis) =         </td> <td>           P(ortah Horor) =  <math>(12.3758 + 1) / (38192.9487 + 188456.0096) = 5.9915E-5</math>             P(iu Horor) =  <math>(15.6835 + 1) / (38192.9487 + 188456.0096) = 7.43007E-5</math>             P(mearang Horor) =  <math>(3.1239 + 1) / (38192.9487 + 188456.0096) = 1.81819E-5</math>             P(ermis Horor) =         </td> </tr> </table>	Genre Romantis	Genre Horor	P(ortah Romantis) = $(22.6317 + 1) / (50308.602 + 188456.0096) = 8.6634E-5$  P(iu Romantis) = $(16.4151 + 1) / (52066.602 + 188456.0096) = 7.11304E-5$  P(mearang Romantis) = $(3.1239 + 1) / (50308.602 + 188456.0096) = 1.8630E-5$  P(ermis Romantis) =	P(ortah Horor) = $(12.3758 + 1) / (38192.9487 + 188456.0096) = 5.9915E-5$  P(iu Horor) = $(15.6835 + 1) / (38192.9487 + 188456.0096) = 7.43007E-5$  P(mearang Horor) = $(3.1239 + 1) / (38192.9487 + 188456.0096) = 1.81819E-5$  P(ermis Horor) =
Genre Romantis	Genre Horor					
P(ortah Romantis) = $(22.6317 + 1) / (50308.602 + 188456.0096) = 8.6634E-5$  P(iu Romantis) = $(16.4151 + 1) / (52066.602 + 188456.0096) = 7.11304E-5$  P(mearang Romantis) = $(3.1239 + 1) / (50308.602 + 188456.0096) = 1.8630E-5$  P(ermis Romantis) =	P(ortah Horor) = $(12.3758 + 1) / (38192.9487 + 188456.0096) = 5.9915E-5$  P(iu Horor) = $(15.6835 + 1) / (38192.9487 + 188456.0096) = 7.43007E-5$  P(mearang Horor) = $(3.1239 + 1) / (38192.9487 + 188456.0096) = 1.81819E-5$  P(ermis Horor) =					

Gambar 4.14 Perhitungan Klasifikasi *Naïve Bayes Classifier*

4

Menyimpulkan Genre Cerpen Uji

Kesimpulan: Karena nilai probabilitas cerpen uji (doc) lebih besar pada genre Horor dibandingkan pada genre Rc.

**Horor**

$P(\text{doc} | \text{Romantis}) >$   
 $P(\text{doc} | \text{Horor}) >$   
 $P(\text{doc} | \text{Misteri}) >$   
 $P(\text{doc} | \text{Fantas}) = \text{Romantis}$

$P(\text{doc} | \text{Romantis}) <$   
 $P(\text{doc} | \text{Horor}) >$   
 $P(\text{doc} | \text{Misteri}) >$   
 $P(\text{doc} | \text{Fantas}) = \text{Horor}$

$P(\text{doc} | \text{Romantis}) <$   
 $P(\text{doc} | \text{Horor}) <$   
 $P(\text{doc} | \text{Misteri}) >$

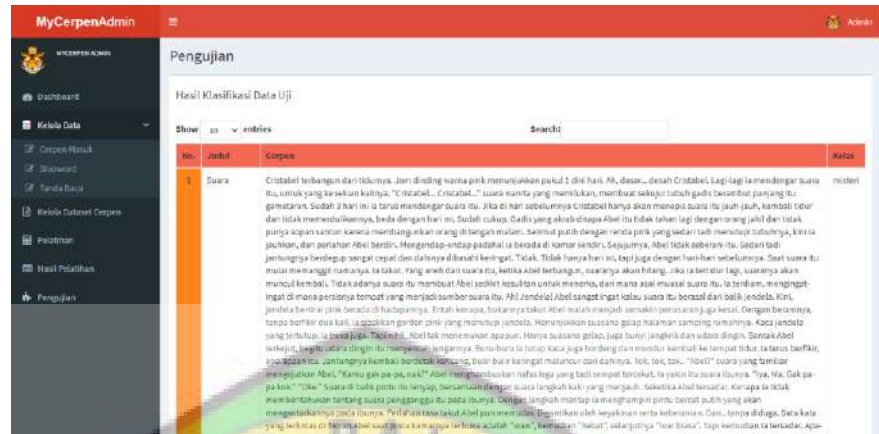
Showing 1 to 4 of 4 entries Previous Next

© 2020 Petra Wika Nandayani R. DBC 116 038

Gambar 4.15 Hasil Klasifikasi *Naïve Bayes Classifier*

Halaman ini adalah halaman untuk menguji hasil pengujian, yaitu menguji apakah sistem dapat mengklasifikasi data cerpen baru. Pada halaman ini, admin dapat menginputkan data uji. Halaman pengujian akan menampilkan hasil dari *case folding*, *tokenizing*, *filtering*, dan hasil klasifikasi *Naïve Bayes Classifier*.

## 11. Implementasi Hasil Klasifikasi Data Uji

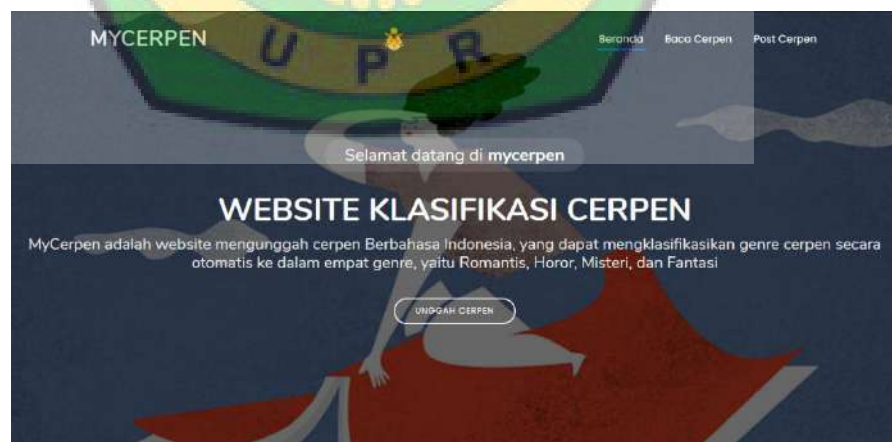


**Gambar 4.16 Hasil Klasifikasi Data Uji**

Halaman hasil klasifikasi data uji menampilkan hasil klasifikasi dari data yang diuji, yaitu menampilkan data cerpen uji dan genrenya.

### 4.1.1.2 Implementasi Halaman Penulis

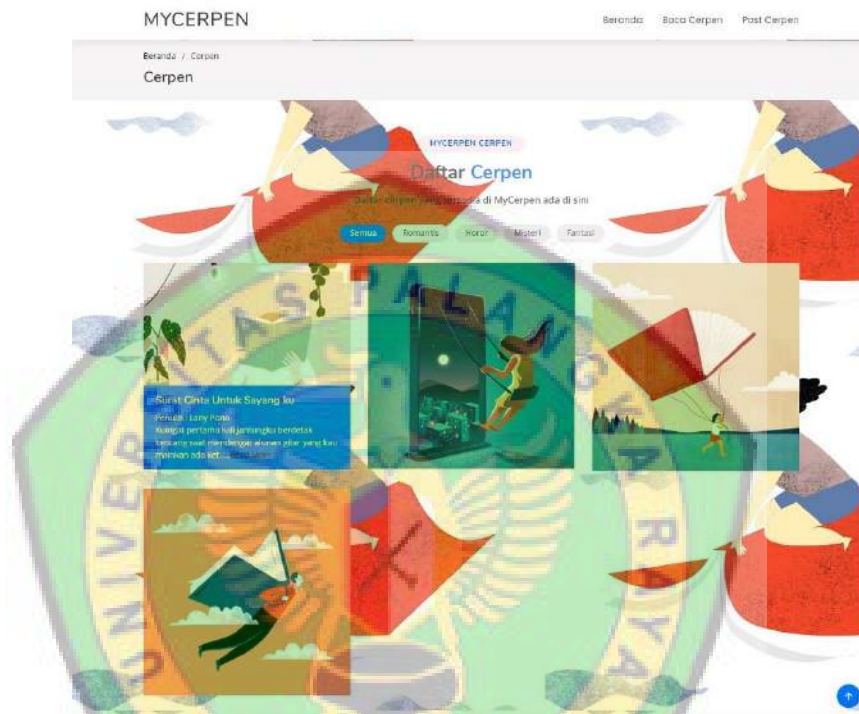
#### 1. Implementasi Halaman Beranda



**Gambar 4.17 Halaman Beranda**

Halaman beranda adalah halaman yang akan tampil saat penulis pertama kali mengakses *website*. Halaman beranda terlihat seperti gambar di atas.

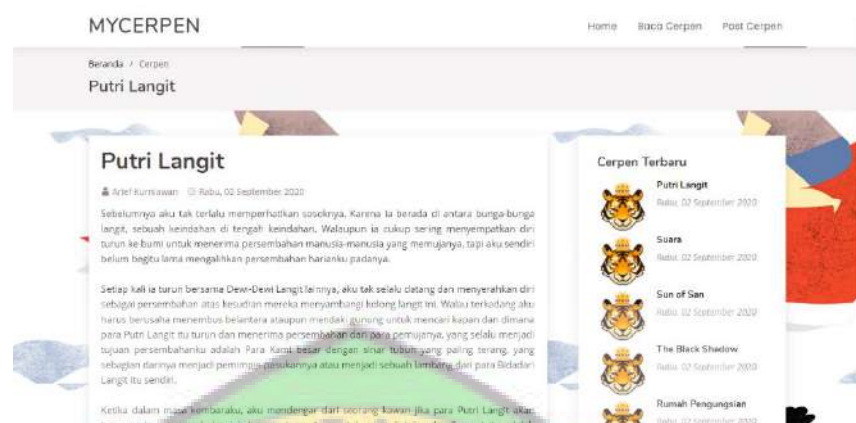
## 2. Implementasi Halaman Baca Cerpen



**Gambar 4.18 Halaman Baca Cerpen**

Halaman ini halaman yang menampilkan daftar cerpen yang tersedia di *website*. Penulis dapat memilih untuk melihat daftar cerpen sesuai genrenya.

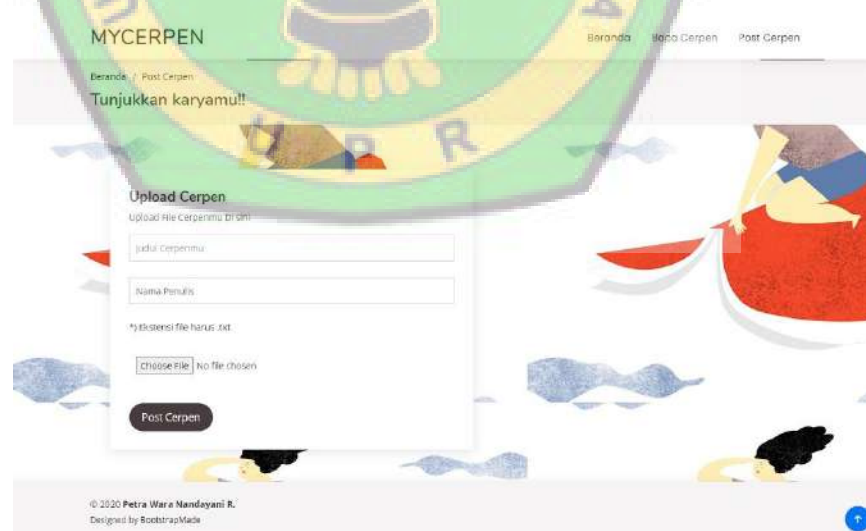
### 3. Implementasi Halaman Isi Cerpen



**Gambar 4.19 Halaman Isi Cerpen**

Halaman ini adalah halaman yang akan tampil saat penulis memilih sebuah cerpen. Pada halaman ini akan ditampilkan isi cerpen yang dipilih.

### 4. Implementasi Halaman *Post Cerpen*



**Gambar 4.20 Halaman *Post Cerpen***

Pada halaman ini, penulis dapat mengunggah cerpen baru dan sistem akan mengklasifikasikan cerpen tersebut ke dalam empat kelas genre yang ada. Pada halaman ini terdapat keterangan format file yang dapat diunggah penulis, yaitu dokumen format file teks dengan ekstensi *.txt*.



**Gambar 4.21 Pemberitahuan Genre Cerpen**

Setelah penulis mengunggah cerpennya, sistem akan memproses cerpen tersebut. Setelah sistem selesai memproses dan mengklasifikasikan cerpen baru tersebut, akan muncul pemberitahuan cerpen yang baru diunggah penulis masuk ke dalam genre apa. Penulis dapat melihat cerpen yang baru saja diunggah pada daftar cerpen di halaman baca cerpen. Pada penelitian ini *website dihosting* dengan nama domain *mycerpen.com*.

## 4.2 Pengujian

Pada penelitian ini, pengujian yang dilakukan yaitu pengujian *performa* klasifikasi sistem menggunakan metode akurasi dan pengujian fungsionalitas sistem menggunakan metode *black box testing*.

### 4.2.1 Black Box Testing

Berikut ini adalah pengujian *black box* terhadap sistem yang telah dibuat.

#### 1. Black Box Testing Admin

Pada bagian admin, pengujian dilakukan terhadap halaman *login*, halaman kelola cerpen masuk, halama kelola *stopword*, halaman kelola tanda baca, halaman kelola *dataset*, halaman pelatihan, dan halaman pengujian.

##### i. Pengujian *Form Login* Admin

**Tabel 4.1 Pengujian *Form Login* Admin**

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Melakukan <i>login</i> (memasukan <i>username</i> dan <i>password</i> salah)	Gagal masuk ke halaman utama dan kembali ke halaman <i>login</i> .	Gagal masuk ke halaman utama dan kembali ke halaman <i>login</i> .	OK

**Tabel 4.1 Pengujian Form Login Admin (lanjutan)**

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
2.	Melakukan <i>login</i> (memasukan <i>username</i> dan <i>password</i> benar)	Berhasil masuk ke halaman utama.	Berhasil masuk ke halaman utama.	OK

## ii. Pengujian Halaman Kelola Cerpen Masuk

**Tabel 4.2 Pengujian Halaman Kelola Cerpen Masuk**

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Klik kelola cerpen masuk	Menampilkan tabel cerpen masuk (judul, penulis, cerpen)	Menampilkan tabel cerpen masuk (judul, penulis, cerpen)	OK

Tabel 4.2 Pengujian Halaman Kelola Cerpen Masuk (lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
2.	Klik aksi hapus	Menampilkan notifikasi yakin hapus data.	Menampilkan notifikasi yakin hapus data.	OK
3.	Klik aksi hapus (pilih Hapus)	Menampilkan pemberitahuan data berhasil dihapus dan kembali ke halaman kelola cerpen masuk.	Menampilkan pemberitahuan data berhasil dihapus dan kembali ke halaman kelola cerpen masuk.	OK

iii. Pengujian Halaman Kelola *Stopword*Tabel 4.3 Pengujian Halaman Kelola *Stopword*

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1.	Klik tambah kata	Menampilkan <i>form</i> tambah kata (masukan	Menampilkan <i>form</i> tambah kata (masukan	OK

Tabel 4.3 Pengujian Halaman Kelola *Stopword* (lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
		kata).	kata).	
2.	Klik simpan pada <i>form</i> tambah cerpen.	Menampilkan pemberitahuan data berhasil disimpan dan kembali ke halaman kelola <i>stopword</i> .	Menampilkan pemberitahuan data berhasil disimpan dan kembali ke halaman kelola <i>stopword</i> .	OK
3.	Klik aksi edit	Menampilkan <i>form</i> ubah data (kata yang akan diubah).	Menampilkan <i>form</i> ubah data (kata yang akan diubah).	OK
4.	Klik simpan pada <i>form</i>	Menampilkan pemberitahuan data berhasil diubah, kembali ke halaman <i>stopword</i>	Menampilkan pemberitahuan data berhasil diubah, kembali ke halaman <i>stopword</i>	

Tabel 4.3 Pengujian Halaman Kelola *Stopword* (lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
5	Klik aksi hapus	Menampilkan notifikasi yakin hapus data.	Menampilkan notifikasi yakin hapus data.	OK
6	Klik aksi hapus (pilih Hapus)	Menampilkan pemberitahuan data berhasil dihapus dan kembali ke halaman <i>stopword</i> .	Menampilkan pemberitahuan data berhasil dihapus dan kembali ke halaman <i>stopword</i> .	OK

## iv. Pengujian Halaman Kelola Tanda Baca

Tabel 4.4 Pengujian Halaman Kelola Tanda Baca

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1	Klik tambah tanda baca	Menampilkan <i>form</i> tambah tanda baca	Menampilkan <i>form</i> tambah tanda baca	OK

Tabel 4.4 Pengujian Halaman Kelola Tanda Baca (lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
		(masukan simbol tanda baca dan nama tanda baca).	(masukan simbol tanda baca dan nama tanda baca).	
2	Klik simpan pada <i>form</i> tambah cerpen.	Menampilkan pemberitahuan data berhasil disimpan dan kembali ke halaman kelola tanda baca.	Menampilkan pemberitahuan data berhasil disimpan dan kembali ke halaman kelola tanda baca.	OK
3	Klik aksi edit	Menampilkan <i>form</i> ubah data (simbol tanda baca dan nama tanda baca yang akan diubah).	Menampilkan <i>form</i> ubah data (simbol tanda baca dan nama tanda baca yang akan diubah).	OK

Tabel 4.4 Pengujian Halaman Kelola Tanda Baca (lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
4	Klik simpan pada <i>form</i>	Menampilkan pemberitahuan data berhasil diubah dan kembali ke halaman tanda baca.	Menampilkan pemberitahuan data berhasil diubah dan kembali ke halaman tanda baca.	OK
5	Klik aksi hapus	Menampilkan notifikasi yakin hapus data.	Menampilkan notifikasi yakin hapus data.	OK
6	Klik aksi hapus (pilih Hapus)	Menampilkan pemberitahuan data berhasil dihapus dan kembali ke halaman tanda baca.	Menampilkan pemberitahuan data berhasil dihapus dan kembali ke halaman tanda baca.	OK

v. Pengujian Halaman Kelola *Dataset*Tabel 4.5 Pengujian Halaman Kelola *Dataset*

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1	Klik aksi pilih kelas genre.	Menampilkan <i>form</i> pilih kelas.	Menampilkan <i>form</i> pilih kelas.	OK
2	Klik pilih kelas genre.	Menampilkan pilihan kelas genre (romantis, horor, misteri, dan fantasi) dan memilih salah satu kelas.	Menampilkan pilihan kelas genre (romantis, horor, misteri, dan fantasi) dan memilih salah satu kelas.	OK
3	Klik simpan pada <i>form</i> .	Menampilkan pemberitahuan data berhasil diubah dan kembali ke <i>dataset</i> .	Menampilkan pemberitahuan data berhasil diubah dan kembali ke <i>dataset</i> .	OK

## vi. Pengujian Halaman Pelatihan

Tabel 4.6 Pengujian Halaman Pelatihan

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1	Pilih pelatihan	Menampilkan proses pelatihan dari <i>dataset</i> cerpen dengan status latih 'belum', menampilkan proses <i>case folding</i> , <i>tokenizing</i>	Menampilkan proses pelatihan dari <i>dataset</i> cerpen dengan status latih 'belum', menampilkan proses <i>case folding</i> , <i>tokenizing</i>	OK

## vii. Pengujian Halaman Hasil Pelatihan

Tabel 4.7 Pengujian Halaman Hasil Pelatihan

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1	Klik Hasil Pelatihan	Menampilkan halaman yang	Menampilkan halaman yang	OK

**Tabel 4.7 Pengujian Halaman Hasil Pelatihan (lanjutan)**

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
		berisi hasil pelatihan, yaitu menampilkan probabilitas prior kelas dan <i>dataset</i> hasil pelatihan	berisi hasil pelatihan, yaitu menampilkan probabilitas prior kelas dan <i>dataset</i> hasil pelatihan	

## viii. Pengujian Halaman Pengujian

**Tabel 4.8 Pengujian Halaman Pengujian**

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1	Pilih Pengujian	Menampilkan halaman pengujian	Menampilkan halaman pengujian	OK
2	Klik <i>upload</i> data uji	Memilih <i>file</i> yang akan diuji	Memilih <i>file</i> yang akan diuji	OK

Tabel 4.8 Pengujian Halaman Pengujian (lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
3	Klik Uji	Menampilkan proses pengujian, menampilkan proses <i>case folding</i> , <i>tokenizing</i> , dan perhitungan <i>Naïve Bayes</i>	Menampilkan proses pengujian, menampilkan proses <i>case folding</i> , <i>tokenizing</i> , dan perhitungan <i>Naïve Bayes</i>	OK

## ix. Pengujian Halaman Hasil Klasifikasi Data Uji

Tabel 4.9 Pengujian Halaman Hasil Klasifikasi Data Uji

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1	Klik Hasil Klasifikasi Data Uji	Menampilkan halaman yang berisi hasil	Menampilkan halaman yang berisi hasil	OK

**Tabel 4.9 Pengujian Halaman Hasil Klasifikasi Data Uji (lanjutan)**

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
		klasifikasi data uji	klasifikasi data uji	

## 2. *Black Box Testing* Penulis

### i. Pengujian Halaman Baca Cerpen

**Tabel 4.10 Pengujian Halaman Baca Cerpen**

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1	Penulis memilih menu baca cerpen	Menampilkan halaman daftar cerpen	Menampilkan halaman daftar cerpen	OK
2	Penulis memilih salah satu kelas genre romantis, horor, misteri, dan fantasi	Menampilkan halaman daftar cerpen berdasarkan genre	Menampilkan halaman daftar cerpen berdasarkan genre	OK

**Tabel 4.10 Pengujian Halaman Baca Cerpen (lanjutan)**

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
3	Penulis memilih salah satu cerpen	Menampilkan halaman isi cerpen	Menampilkan halaman isi cerpen	OK

ii. Pengujian Halaman *Post* Cerpen**Tabel 4.11 Pengujian Halaman *Post* Cerpen**

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
1	Penulis memilih menu <i>post</i> cerpen	Menampilkan halaman <i>post</i> cerpen, berisi <i>form post</i> cerpen	Menampilkan halaman <i>post</i> cerpen, berisi <i>form post</i> cerpen	OK
2	Penulis mengunggah cerpen	Cerpen berhasil terunggah dan sistem otomatis mengklasifikasi genre cerpen	Cerpen berhasil terunggah dan sistem otomatis mengklasifikasi genre cerpen	OK

Tabel 4.11 Pengujian Halaman *Post Cerpen* (lanjutan)

No.	Kondisi Awal	Hasil Yang Diharapkan	Hasil Keluaran	Hasil Pengujian
3	Penulis mengunggah file cerpen selain format <i>.txt</i>	Cerpen tidak berhasil terunggah, muncul peringatan format file yang dapat diunggah	Cerpen tidak berhasil terunggah, muncul peringatan format file yang dapat diunggah	OK

#### 4.2.2 Akurasi

Perangkat keras yang (komputer) yang digunakan dalam pengujian ini adalah laptop *Asus VivoBook A442U* dengan spesifikasi sebagai berikut :

1. *Processor Inter Core i5-8250U*
2. *RAM 8 GB*
3. *HDD 1 TB*
4. Sistem operasi *Windows 10*

Pengujian menggunakan 200 data cerpen dengan kelas genre romantis, horor, misteri, dan fantasi, berupa file teks dengan ekstensi *.txt*. Komposisi data yang digunakan dalam lima kali pelatihan dan pengujian untuk mendapatkan pola data yang optimal, seperti pada tabel 4.1.

**Tabel 4.12 Pembagian Data**

Komposisi Data		Jumlah Data	
Data Latih	Data Uji	Data Latih	Data Uji
50%	50%	100	100
60%	40%	120	80
70%	30%	140	60
80%	20%	160	40
90%	10%	180	20

Hasil pengujian sistem klasifikasi genre cerpen Berbahasa Indonesia dengan menggunakan algoritma *Naïve Bayes Classifier*, dilakukan dengan menggunakan skenario pengujian pada sub bab 3.3.4.2.

Pengujian pertama menggunakan model hasil pelatihan dengan 349 *stopword*. Akurasi didapat dari persentase jumlah data yang diklasifikasikan secara benar kedalam kelasnya terhadap keseluruhan data uji. Hasil perhitungan klasifikasi dapat dilihat pada tabel 4.13, 4.14, 4.15, 4.16, dan 4.17, sebagai berikut.

**Tabel 4.13 Matriks Confusion Pengujian 50%-50%**

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Romantis	10	8	6	0
	Horor	7	8	9	1

Tabel 4.13 Matriks Confusion Pengujian 50%-50% (lanjutan)

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Misteri	6	12	3	5
	Fantasi	10	8	5	2

Akurasi =

$$\begin{aligned}
 & \frac{10 + 8 + 3 + 2}{10 + 8 + 3 + 6 + 0 + 7 + 8 + 9 + 1 + 6 + 12 + 3 + 5 + 10 + 8 + 5 + 2} \times 100\% \\
 &= \frac{23}{100} \times 100\% \\
 &= 23\%
 \end{aligned}$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk mengklasifikasikan data teks dengan data latih sebanyak 100 data dan data uji sebanyak 100 data sebesar 23%.

Tabel 4.14 Matriks Confusion Pengujian 60%-40%

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Romantis	8	7	5	0
	Horor	1	12	6	1
	Misteri	5	9	4	2

Tabel 4.14 Matriks Confusion Pengujian 60%-40% (lanjutan)

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Fantasi	8	6	4	2

Akurasi =

$$\begin{aligned}
 & \frac{8 + 12 + 4 + 2}{8 + 7 + 5 + 0 + 1 + 12 + 6 + 1 + 5 + 9 + 4 + 2 + 8 + 6 + 4 + 2} \times 100\% \\
 &= \frac{26}{80} \times 100\% \\
 &= 32,5\%
 \end{aligned}$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk mengklasifikasikan data teks dengan data latih sebanyak 120 data dan data uji sebanyak 80 data sebesar 32,5%.

Tabel 4.15 Matriks Confusion Pengujian 70%-30%

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Romantis	6	9	0	0
	Horor	2	9	3	1
	Misteri	4	5	3	2
	Fantasi	8	3	2	2

*Akurasi =*

$$\frac{6 + 9 + 3 + 2}{6 + 9 + 0 + 0 + 2 + 9 + 3 + 1 + 4 + 5 + 3 + 2 + 8 + 3 + 2 + 2} \times 100\%$$

$$= \frac{20}{60} \times 100\%$$

$$= 33,3\%$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk mengklasifikasikan data teks dengan data latih sebanyak 140 data dan data uji sebanyak 60 data sebesar 33,3%.

**Tabel 4.16 Matriks Confusion Pengujian 80%-20%**

<b>Kelas</b>		<b>Kelas Hasil Klasifikasi</b>			
		<b>Romantis</b>	<b>Horor</b>	<b>Misteri</b>	<b>Fantasi</b>
<b>Kelas Asli</b>	<b>Romantis</b>	7	2	1	0
	<b>Horor</b>	2	7	1	0
<b>Kelas Asli</b>	<b>Misteri</b>	4	2	4	0
	<b>Fantasi</b>	3	3	2	2

*Akurasi =*

$$\frac{7 + 7 + 4 + 2}{7 + 2 + 1 + 0 + 2 + 7 + 1 + 0 + 4 + 2 + 4 + 0 + 3 + 3 + 2 + 2} \times 100\%$$

$$= \frac{20}{40} \times 100\%$$

$$= 50\%$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk mengklasifikasikan data teks dengan data latih sebanyak 160 data dan data uji sebanyak 40 data sebesar 50%.

**Tabel 4.17 Matriks Confusion Pengujian 90%-10%**

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Romantis	5	0	0	0
	Horor	0	5	0	0
	Misteri	1	0	4	0
	Fantasi	2	1	1	1

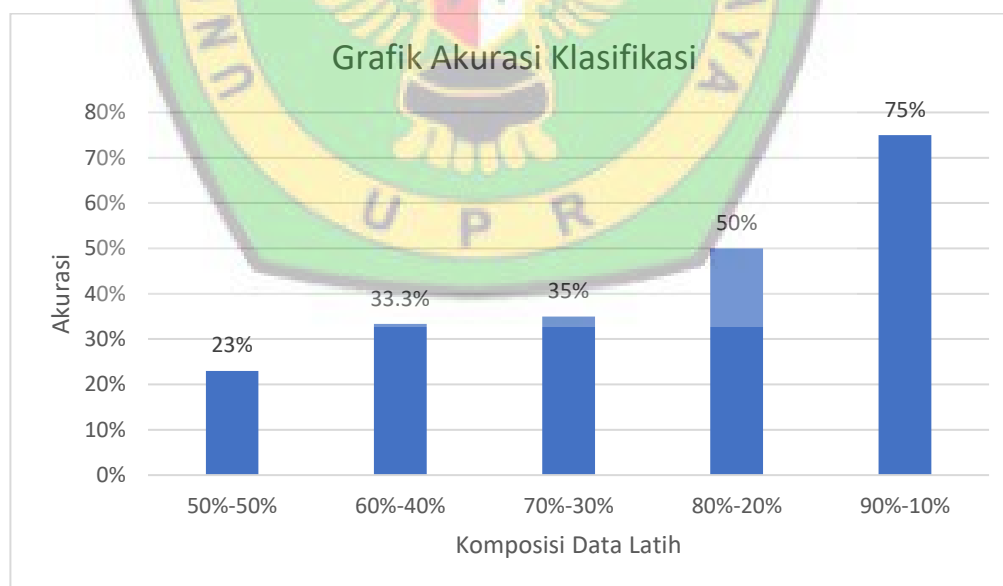
*Akurasi* =

$$\begin{aligned}
 & \frac{5 + 5 + 4 + 1}{5 + 0 + 0 + 0 + 0 + 5 + 0 + 0 + 1 + 0 + 4 + 0 + 2 + 1 + 1 + 1} \times 100\% \\
 & = \frac{15}{20} \times 100\% \\
 & = 75\%
 \end{aligned}$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk mengklasifikasikan data teks dengan data latih sebanyak 180 data dan data uji sebanyak 20 data sebesar 75%.

Tabel 4.18 Hasil Pengujian

Komposisi Data		Jumlah Data		Waktu Pelatihan	Waktu Pengujian (per satu data)	Akurasi
Data Latih	Data Uji	Data Latih	Data Uji			
50%	50%	100	100	5 jam	3-5 menit	23%
60%	40%	120	80	6 jam	3-5 menit	33,33%
70%	30%	140	60	7 jam	3-5 menit	35%
80%	20%	160	40	8 jam	3-5 menit	50%
90%	10%	180	20	9 jam	3-5 menit	75%



Gambar 4.22 Grafik Akurasi Klasifikasi Sistem

Dari lima pengujian yang dilakukan, didapat hasil dengan akurasi tertinggi sebesar 75% dengan komposisi 90% data latih dan 10% data uji.

Setelah dilakukan analisis data pada *stopword* awal, tidak terdapat kata yang dihapuskan dari daftar *stopword*. Karena kata-kata awal yang sudah ada dalam daftar *stopword* yang terdapat dalam sumber awal merupakan kata-kata umum yang memang harus dimasukkan dalam daftar *stopword*. Sehingga tidak ada kata yang dihapuskan dari daftar *stopword*. Setelah dilakukan evaluasi terhadap *stopword* yang digunakan terdapat perubahan performa sistem dari skenario pengujian sebelumnya.

**Tabel 4.19 Matriks Confusion Pengujian 50%-50%**

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Romantis	9	1	5	7
	Horor	6	13	13	8
	Misteri	6	8	3	6
	Fantasi	4	3	4	4

*Akurasi* =

$$\begin{aligned}
 & \frac{4 + 13 + 3 + 9}{9 + 6 + 6 + 4 + 1 + 13 + 8 + 3 + 5 + 13 + 3 + 4 + 7 + 8 + 6 + 4} \times 100\% \\
 & = \frac{29}{100} \times 100\% \\
 & = 29\%
 \end{aligned}$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk mengklasifikasikan data teks dengan data latih sebanyak 100 data dan data uji sebanyak 100 data sebesar 29%.

**Tabel 4.20 Matriks Confusion Pengujian 60%-40%**

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Romantis	6	2	5	6
	Horor	8	11	9	5
	Misteri	3	5	3	5
	Fantasi	3	2	3	4

*Akurasi* =

$$\frac{6 + 11 + 3 + 4}{6 + 8 + 3 + 3 + 2 + 11 + 5 + 2 + 5 + 9 + 3 + 3 + 6 + 5 + 5 + 4} \times 100\%$$

$$= \frac{24}{80} \times 100\%$$

$$= 30\%$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk mengklasifikasikan data teks dengan data latih sebanyak 120 data dan data uji sebanyak 80 data sebesar 30%.

Tabel 4.21 Matriks Confusion Pengujian 70%-30%

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Romantis	7	0	1	2
	Horor	4	10	8	5
	Misteri	3	3	4	3
	Fantasi	1	2	2	5

*Akurasi* =

$$\begin{aligned}
 & \frac{7 + 10 + 4 + 5}{7 + 4 + 3 + 1 + 0 + 10 + 3 + 2 + 1 + 8 + 4 + 2 + 2 + 5 + 3 + 5} \times 100\% \\
 &= \frac{26}{60} \times 100\% \\
 &= 43,3\%
 \end{aligned}$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk mengklasifikasikan data teks dengan data latih sebanyak 140 data dan data uji sebanyak 60 data sebesar 43,3%.

Tabel 4.22 Matriks Confusion Pengujian 80%-20%

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Romantis	6	0	1	0

Tabel 4.22 Matriks Confusion Pengujian 80%-20% (lanjutan)

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Horor	2	8	4	2
	Misteri	0	2	4	2
	Fantasi	2	0	1	6

Akurasi =

$$\begin{aligned}
 & \frac{6 + 8 + 4 + 6}{6 + 2 + 0 + 2 + 0 + 8 + 2 + 0 + 1 + 4 + 4 + 1 + 0 + 2 + 2 + 6} \times 100\% \\
 &= \frac{24}{40} \times 100\% \\
 &= 60\%
 \end{aligned}$$

Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk mengklasifikasikan data teks dengan data latih sebanyak 160 data dan data uji sebanyak 40 data sebesar 60%.

Tabel 4.23 Matriks Confusion Pengujian 90%-10%

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Romantis	5	0	0	0
	Horor	0	5	0	0

Tabel 4.23 Matriks Confusion Pengujian 90%-10% (lanjutan)

Kelas		Kelas Hasil Klasifikasi			
		Romantis	Horor	Misteri	Fantasi
Kelas Asli	Misteri	0	3	2	0
	Fantasi	0	0	0	5

Akurasi =

$$\frac{5 + 5 + 2 + 5}{5 + 0 + 0 + 0 + 0 + 0 + 5 + 0 + 0 + 0 + 3 + 2 + 0 + 0 + 0 + 0 + 5} \times 100\%$$

$$= \frac{17}{20} \times 100\%$$

$$= 85\%$$

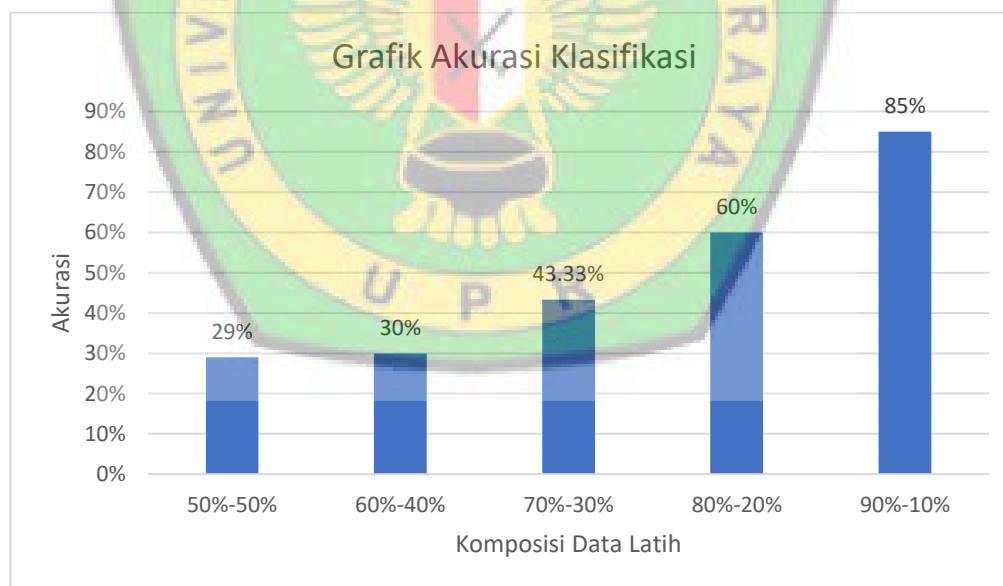
Dari hasil pengujian diatas dapat diketahui kemampuan sistem untuk mengklasifikasikan data teks dengan data latih sebanyak 180 data dan data uji sebanyak 20 data sebesar 85%.

Tabel 4.24 Hasil Pengujian

Komposisi Data		Jumlah Data		Waktu Pelatihan	Waktu Pengujian (per satu data)	Akurasi
Data Latih	Data Uji	Data Latih	Data Uji			
50%	50%	100	100	1 jam 9 menit	1-2 menit	29%

Tabel 4.24 Hasil Pengujian (lanjutan)

Komposisi Data		Jumlah Data		Waktu Pelatihan	Waktu Pengujian (per satu data)	Akurasi
Data Latih	Data Uji	Data Latih	Data Uji			
60%	40%	120	80	1 jam 16 menit	1-2 menit	30%
70%	30%	140	60	1 jam 46 menit	1-2 menit	43,33%
80%	20%	160	40	2 jam 7 menit	1-2 menit	60%
90%	10%	180	20	2 jam 30 menit	1-2 menit	85%



Gambar 4.23 Grafik Akurasi Klasifikasi Sistem

Berdasarkan lima pengujian kedua, didapat hasil dengan akurasi tertinggi sebesar 85% dengan komposisi 90% data latih dan 10% data uji. Dari hasil pengujian, didapat bahwa

jumlah *dataset* yang digunakan mempengaruhi hasil klasifikasi. Pada Tabel 4.8, Tabel 4.9, Tabel 4.10, Tabel 4.11, dan Tabel 4.12 dapat disimpulkan bahwa semakin banyak data latih yang digunakan maka semakin baik pula performa sistem dalam mengklasifikasi data. Karena semakin banyak data latih, maka semakin banyak pula kata yang disimpan untuk melakukan proses pembelajaran oleh sistem, sehingga sistem memiliki pengetahuan yang lebih banyak untuk melakukan klasifikasi. Seperti pada penelitian yang dilakukan oleh Hilman, Setio Basuki, dan Vinna Rahmayanti Setyaning Nastiti, yaitu bahwa jumlah *dataset* yang digunakan mempengaruhi hasil klasifikasi, semakin banyak jumlah *dataset* yang digunakan dapat meningkatkan akurasi. Sehingga komposisi data latih dan data uji paling tinggi menghasilkan performa sistem yang lebih baik. Jadi, hasil lima pengujian yang dilakukan, dapat diketahui kemampuan sistem untuk mengklasifikasi data teks yang optimal, yaitu dengan pola data kelima dengan komposisi data 90% data latih dan 10% data uji, yang menghasilkan akurasi 85%.

Kesalahan klasifikasi yang terjadi pada sistem disebabkan adanya kemiripan antara kata-kata penyusun suatu kelas dengan kelas lainnya. Jadi, kata yang didapat tidak merepresentasikan suatu kelas dengan baik. Walaupun bobot kata yang muncul telah diberikan nilai yang bersesuaian terhadap jumlah kemunculannya pada sebuah dokumen disetiap kelasnya. Sebagai contoh, kata suara muncul disetiap kelas tetapi kata tersebut memiliki bobot yang lebih tinggi di kelas horor. Maka jika sebuah dokumen uji dengan genre fantasi terdapat banyak kata suara, dokumen tersebut dapat

diklasifikasikan salah ke dalam kelas horor karena bobot di kelas horor lebih tinggi dari kelas fantasi. Dari hasil pengamatan, berikut adalah 50 kata yang sering muncul pada masing-masing kelas genre.

**Tabel 4.25 Kata yang Sering Muncul**

<b>Romantis</b>	<b>Horor</b>	<b>Misteri</b>	<b>Fantasi</b>
arah	aneh	aneh	aneh
bagiku	arah	arah	arah
bahagia	berjalan	berjalan	baiklah
berjalan	berlari	bertemu	bangun
bertemu	berteriak	berubah	berjalan
berusaha	berusaha	berusaha	berlari
bingung	darah	darah	bertemu
cantik	dingin	ditemukan	berteriak
cinta	gelap	hati	berubah
dingin	hantu	hitam	berusaha
hangat	hati	jalan	berwarna
hati	hilang	kamar	bingung
hatiku	hitam	kantor	cahaya
indah	jalan	kejadian	cantik
jalan	kaget	kepala	dunia
jatuh	kamar	korban	hati

Tabel 4.25 Kata yang Sering Muncul (lanjutan)

Romantis	Horor	Misteri	Fantasi
kasih	kejadian	lakukan	hidup
kelas	ketakutan	lantai	hitam
lupa	lihat	mata	hutan
makan	mata	mati	jalan
malu	mati	meja	jatuh
mata	membawa	membawa	kasih
mata	menangis	membunuh	langit
mataku	mencari	menangis	lihat
matanya	mencoba	menatap	manusia
membalas	mendengar	mencari	mata
membawa	menemukan	mendengar	matanya
membuatku	menghilang	menemukan	mati
menatap	meninggal	menunggu	membawa
mencari	meninggalkan	pembunuhan	memiliki
mencoba	menyeramkan	penasaran	mencari
mendengar	penasaran	percaya	mendengar
menikmati	pintu	pikir	menemukan
menunggu	putih	pintu	menghampiri
perasaan	sadar	polisi	merah

Tabel 4.25 Kata yang Sering Muncul (lanjutan)

Romantis	Horor	Misteri	Fantasi
rindu	salah	ruang	namaku
sadar	sekolah	ruangan	penasaran
salah	sosok	salah	pohon
senyum	suara	suara	putih
suara	takut	takut	sang
suka	tangan	tangan	senang
tangan	tanyaku	teman	suara
tangannya	teman	tenang	takut
tatapan	tenang	terdengar	tangan
tau	terdengar	tersenyum	tangannya
teman	tersenyum	tinggal	teman
terdengar	tidur	tubuh	terkejut
tersenyum	tubuh	tubuhnya	tersenyum
tertawa	wajah	wajah	tidur
wajah	wajahnya	wajahnya	tubuh

Selain itu, kesalahan klasifikasi yang terjadi pada genre misteri dimana genre tersebut dapat diklasifikasikan salah ke dalam genre horor. Hal tersebut disebabkan karena kata penyusun data teks genre misteri dan horor memiliki kemiripan atau

memiliki kesamaan atribut. Sebagai contoh, kata takut, suara, aneh, berteriak, kaget, dan darah.

Pada beberapa pengujian, beberapa kali terjadi kesalahan klasifikasi genre lain ke dalam genre fantasi atau sebaliknya. Hal ini karena genre fantasi memiliki kata yang juga sering muncul di genre lainnya. Misalnya kata suara, takut, terkejut, bingung, aneh, adalah kata-kata yang sering muncul pada genre fantasi, tetapi kata-kata tersebut juga muncul pada genre misteri dan horor. Selain itu kata indah, hati, suka, senang, dunia, dan cantik yang muncul pada genre fantasi, juga sering muncul pada genre romantis. Kata yang menggambarkan genre tersebut terlalu umum dan banyak terdapat di genre lain, sehingga tidak dapat merepresentasikan kelasnya dengan baik. Jadi, selain jumlah data latih yang digunakan dalam pelatihan, atribut kata yang pada setiap dokumen dalam setiap kelas juga mempengaruhi akurasi sistem. Kesalahan klasifikasi yang disebabkan oleh hal-hal tersebut menyebabkan rendahnya akurasi, yang berarti rendahnya performa klasifikasi sistem.

Kendala yang dialami dalam melakukan klasifikasi genre cerpen, yaitu kata yang dihasilkan dalam kamus kata tidak memiliki kata unik yang dapat mewakili setiap kelas genre. Dapat dilihat dari kata atau atribut yang dimiliki setiap kelas memiliki kemiripan. Selain itu, kendala yang terjadi, yaitu lamanya waktu pelatihan dan pengujian pada sistem. Hal tersebut terjadi karena panjang data cerpen. Cerita yang masuk kategori cerpen adalah cerita yang langsung pada inti cerita, panjang cerpen maksimal 10.000 kata. Seperti pada gambar dibawah, pengujian dilakukan pada dua

cerpen, yaitu cerpen dengan panjang 1227 kata dan 259 kata. Dari hasil pengujian didapat bahwa waktu eksekusi atau waktu klasifikasi cerpen dengan panjang 1227 kata adalah 238 detik, dan waktu klasifikasi cerpen dengan panjang 259 kata adalah 71 detik.



**Gambar 4.24 Waktu Pengujian Data**

Sehingga salah satu faktor yang mempengaruhi waktu proses, baik pelatihan maupun pengujian, adalah panjang data atau banyak kata dalam satu cerpen. Setelah dilakukan percobaan antara dua cerpen dengan jumlah kata berbeda, dapat disimpulkan bahwa, waktu proses cerpen dengan jumlah kata lebih sedikit relatif lebih singkat dari pada waktu proses cerpen dengan jumlah kata yang lebih banyak.

Selain itu, penyebab lamanya waktu proses pelatihan dan pengujian juga disebabkan oleh akses dan pengecekan terhadap *database* penyimpanan kamus kata hasil pelatihan yang cukup memakan waktu karena jumlah data kata dan ukuran *database* yang besar. Oleh karena itu, dilakukan pengujian dengan membandingkan proses pelatihan dan pengujian dengan dilakukan pengurangan fitur atau seleksi fitur,

untuk mengurangi atau menghilangkan fitur yang dianggap kurang penting dan kemunculannya sedikit dalam data latih. Seleksi fitur yang dilakukan adalah dengan menyeleksi *term* atau kata hasil *filtering* berdasarkan *DF* (*document frequency*), yaitu frekuensi dokumen yang mengandung kata tertentu. Kemudian dilakukan dua pengujian, yaitu dengan menyeleksi kata dengan *DF* kurang dari 5% dan 10% dari data latih yang digunakan. Berikut adalah hasil dari pengujian setelah dilakukan seleksi fitur.

**Tabel 4.26 Hasil Seleksi Fitur**

<b>Persentasi Data Latih</b>	<b>Jumlah Data Latih</b>	<b>Waktu Pelatihan</b>	<b>Jumlah Term</b>	<b>Waktu Pengujian (per satu data)</b>	<b>Akurasi</b>
50%	100	1 jam	13.389	30 detik – 1 menit	22%
60%	120	1 jam 11 menit	16.762	30 detik – 1 menit	28,7%
70%	140	1 jam 24 menit	19.508	30 detik – 1 menit	30%
80%	160	1 jam 27 menit	23.231	30 detik – 1 menit	37,5%
90%	180	1 jam 35 menit	25.259	30 detik – 1 menit	65%

Tabel 4.26 merupakan tabel hasil seleksi fitur dengan *DF* kurang dari 5% dari jumlah data latih. Misalnya, 5% dari 100 data latih, berarti kata yang kemunculannya kurang 5 dokumen akan dihilangkan dari kamus kata. Berikut adalah kurva jumlah *term* dan akurasi dari hasil seleksi fitur *DF* yang kurang dari 5%.



**Gambar 4.25 Kurva Hasil Seleksi Fitur**

**Tabel 4.27 Hasil Seleksi Fitur**

<b>Persentasi Data Latih</b>	<b>Jumlah Data Latih</b>	<b>Waktu Pelatihan</b>	<b>Jumlah <i>Term</i></b>	<b>Waktu Pengujian (per satu data)</b>	<b>Akurasi</b>
50%	100	55 menit	8.572	15 detik – 1 menit	19%
60%	120	1 jam	10.597	15 detik – 1 menit	21,25%

Tabel 4.27 Hasil Seleksi Fitur (lanjutan)

Persentasi Data Latih	Jumlah Data Latih	Waktu Pelatihan	Jumlah Term	Waktu Pengujian (per satu data)	Akurasi
70%	140	1 jam 13 menit	12.698	15 detik – 1 menit	26,67%
80%	160	1 jam 16 menit	13.738	15 detik – 1 menit	37,5%
90%	180	1 jam 22 menit	16.407	15 detik – 1 menit	45%

Tabel 4.27 merupakan tabel hasil seleksi fitur dengan *DF* kurang dari 10% dari jumlah data latih. Misalnya, 10% dari 100 data latih, berarti kata yang kemunculannya kurang 10 dokumen akan dihilangkan dari kamus kata. Berikut adalah kurva jumlah *term* dan akurasi dari hasil seleksi fitur *DF* yang kurang dari 10%.



**Gambar 4.26 Kurva Hasil Seleksi Fitur**

Berdasarkan seleksi fitur yang dilakukan didapatkan hasil seperti kurva pada gambar 4.25 dan 4.26. kurva tersebut menunjukkan peningkatan akurasi dari setiap pengujian. Tetapi, jika dibandingkan dengan hasil skenario pengujian sebelumnya, akurasi yang didapatkan lebih rendah. Hal tersebut disebabkan karena kamus kata yang dimiliki berkurang. Ketika dilakukan seleksi fitur, maka terjadi pengurangan atribut kata dari kamus kata yang dimiliki, kata yang dibuang adalah kata yang dokumen frekuensinya rendah. Karena setiap kelas memiliki atribut kata yang mirip, maka dalam perhitungan klasifikasi atribut kata yang dokumen frekuensinya rendah tersebutlah yang mempengaruhi perhitungan klasifikasi. Karena kata tersebut hanya muncul dikelas tertentu, dan tidak muncul dikelas lainnya. Jadi, ketika dilakukan seleksi fitur, kata yang hanya muncul dikelas tertentu akan hilang. Selain itu, pada setiap genre dan setiap dokumen dalam genre tersebut memiliki objek yang berbeda tergantung penulis

cerpen, sehingga kemungkinan muncul kata tersebut kurang dari batasan seleksi fitur yang dilakukan. Walaupun kemunculan kata dalam dokumen latih rendah, tetapi bobot yang dari kata tersebut tinggi, sehingga mempengaruhi hasil dari klasifikasinya.

Selain itu, jika dibandingkan dengan skenario pengujian sebelumnya, waktu yang digunakan untuk melakukan proses pelatihan dan pengujian lebih cepat, karena fitur yang dimiliki lebih sedikit, sehingga akses *database* dan perhitungan yang dilakukan menjadi lebih cepat. Jadi, dalam penelitian ini, seleksi fitur tidak menaikkan performa klasifikasi sistem, tetapi seleksi fitur dapat mempercepat proses perhitungan.



## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Dari penelitian ini dapat disimpulkan bahwa dalam merancang dan membangun sistem yang dapat mengklasifikasi genre cerpen Berbahasa Indonesia ini menggunakan metodologi pengembangan perangkat lunak *Waterfall* Pressman, dengan empat tahapan, yaitu analisis, desain, pengkodean, dan pengujian. Analisis sistem menggunakan *flowchart*, desain sistem menggunakan *Unified Modelling Language (UML)*. Pengkodean dan implementasi desain sistem menggunakan bahasa pemrograman *PHP* dan *database MySQL*. Pengujian menggunakan metode *black box testing* dan metode akurasi.

Klasifikasi genre cerpen Berbahasa Indonesia ke dalam empat kelas genre, yaitu romantis, horor, misteri, dan fantasi, menggunakan algoritma *Naïve Bayes Classifier*. Tahapan klasifikasi meliputi ekstraksi data, pemrosesan teks, dan pengujian. Tahapan pemrosesan teks terdiri dari empat tahapan, yaitu *text preprocessing*, *text transformation*, *feature selection*, dan *pattern discovery*. Tahap *text preprocessing* terdiri dari tahap *tokenizing* dan *case folding*, tahap *text transformation* terdiri dari tahap *filtering*, tahap *feature selection* yang terdiri dari tahap pembobotan teks menggunakan *TF-IDF (Term Frequency – Inverse Document Frequency)*. Tahap klasifikasi menggunakan algoritma *Naïve Bayes Classifier*.

Hasil pengujian dengan *black box* menunjukkan sistem dapat berfungsi seperti tujuan. Pengujian dengan metode akurasi, didapat hasil akurasi tertinggi 85% pada pengujian dengan data 90% latih dan 10 % data uji dari 200 *dataset*. Dari hasil pengujian, dapat disimpulkan bahwa semakin banyak data latih maka semakin tinggi akurasi klasifikasi sistem, karena semakin banyak kata yang dipelajari sistem, dan panjang data mempengaruhi waktu proses dalam sistem, baik pelatihan maupun pengujian.

## 5.2 Saran

Beberapa saran yang dirasa perlu untuk perbaikan dan peningkatan sistem yang diperlukan disebagai berikut.

1. Mencoba algoritma pengklasifikasian lain untuk meningkatkan performa klasifikasi sistem.
2. Menambah jumlah data agar dapat membantu memperbanyak kamus kata data dalam data latih.
3. Pada penelitian selanjutnya dapat menggunakan proses *stemming* pada tahap pemrosesan teks.

## DAFTAR PUSTAKA

- Akrhamsiagustinah. 2015. *Makalah Pengantar Data Mining Word Cloud Of Corruption Eradication Commission (Studi Kasus : Akun Twitter Komisi Pemberantasan Korupsi @KPK\_RI)*. Makalah. Yogyakarta: Universitas Islam Indonesia
- Apriliana dkk. 2017. Implementasi *Text Mining* Klasifikasi Skripsi Menggunakan Metode *Naïve Bayes Classifier*. Jurnal. *semanTIK*, Vol.3, No.2. Kendari: Universitas Halu Oleo
- Budiman, Rudyanto. 2015. *Klasifikasi Novel Sesuai Dengan Genre Menggunakan TF-IDF*. Skripsi. Medan: Universitas Sumatera Utara
- Febriani, Enggal Suci. 2017. *Klasifikasi Konten Berita Surat Kabar Berdasarkan Judul Dengan Text Mining Menggunakan Metode Naïve Bayes (Studi Kasus : Radar Kediri)*. Kediri: Universitas Nusantara PGRI Kediri
- Gustiawan, Mas Aji. 2016. *Cara Membuat Buku Fiksi : Lima Genre Tulisan Fiksi Yang Wajib Kamu Ketahui !!*. Didapat dari <https://penerbitdeepublish.com/cara-membuat-buku-penerbit-buku-g002/>. Diakses pada Rabu, 19 Agustus 2020
- Haryalesmana, Devid. 2016. Sumber *stopword* <https://github.com/masdevid/ID-Stopwords>
- Hendryadi. 2014. *Metode Pengumpulan Data*. Teorionile Personal Paper No. 01/Jan-2014
- Hariyanto, Budi dkk. 2018. *Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Retensi Arsip*. Jurnal. *Jurnal Informatika Polinema* Volume 4, Edisi 2. Malang: Politeknik Negeri Malang
- Herlambang, Admaja Dwi dan Satrio Hadi Wijoyo. 2019. Algoritma *Naïve Bayes* Untuk Klasifikasi Sumber Belajar Berbasis Teks Pada Mata Pelajaran Produktif Di SMK Rumpun Teknologi Informasi dan Komunikasi. Jurnal. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)* Vol.6, No.4. Malang: Universitas Brawijaya
- Hermawati, Fajar Astuti. 2013. *Data Mining*. Yogyakarta: Andi
- Indriyani, Iin. 2015. Pembelajaran Apresiasi Cerpen Melalui Model *Discovery Learning* Berbasis Nilai-nilai Karakter. Jurnal. *Jurnal STKIP Siliwangi Bandung* Vol 4, No 2 (2015). Bandung: STKIP Siliwangi
- Indriyono, Bonaficius dkk. 2015. Klasifikasi Jenis Buku Berdasarkan Judul dan Sinopsis Menggunakan Metode *Naïve Bayes Classifier* (Studi Kasus: STMIK Kadiri). *Jurnal Sistem Informasi*, Volume 5, Nomor 4 (2015). Yogyakarta: STMIK AMIKOM Yogyakarta
- Jaya, Tri Snadhika. 2018. Pengujian Aplikasi dengan Metode *Black Box Testing Boundary Value Analysis* (Studi Kasus: Kantor Digital Politeknik Negeri Lampung). Jurnal. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, Vol. 03, No. 02. Bandar Lampung: Politeknik Negeri Lampung

- Lestari, Irma Dewi. 2016. *Klasifikasi Online dan Google*. Jurnal. *Jurnal Iqra' Volume 10 No. 02*. Medan: Universitas Islam Negeri Sumatera Utara
- Merawati, Fitri. 2015. *Perkembangan Fiksi Ilmiah Karya Pengarang Indonesia Tahun 1980-an dan 2000-an*. Makalah. Yogyakarta: Universitas Ahmad Dahlan
- Munawir. 2013. *Analisis Gaya Bahasa Pada Novel Anak Pondok Senja Karya Mulasih Tary (Kajian Silistika)*. Skripsi. Purwokerto: Universitas Muhammadiyah Purwokerto
- Mustofa, Roni. 2016. *Resepsi Siswa Terhadap Cerpen Mata yang Enak Dipandang Karya Ahmad Tohari dan Implikasinya pada Pembelajaran Sastra di SMA*. Skripsi. Lampung: Universitas Lampung
- Nastiti, Vinna Rahmayanti Setyaning dkk. 2019. *Klasifikasi Sinopsis Novel Menggunakan Metode Naïve Bayes Classifier*. Jurnal. *Repositor Vol. 1, No. 2*. Malang: Universitas Muhammadiyah Malang
- Nindito, Hendro. 2016. *Teori Text Mining dan Web Mining*. <https://sis.binus.ac.id/2016/12/15/teori-text-mining-dan-web-mining/>. Diakses pada Senin, 24 Februari 2020
- Nugroho, Eko. 2011. *Perancangan Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-KARP*. Skripsi. Malang: Universitas Brawijaya
- Nurjanah, Musfiroh dkk. 2013. *Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) Untuk Text Mining*. Jurnal. *Jurnal Informatika Mulawarman Vol.8, No.3*. Samarinda: Universitas Mulawarman
- Prasetyo, Eko. 2014. *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi
- Prasetyo, Eko. 2013. *Data Mining – Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi
- Rahutomo, Faisal, Ariadi Retno Tri Hayati Ririd. 2019. *Evaluasi Daftar Stopword Bahasa Indonesia*. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK) Vol 6, No. 1*. Malang: Politeknik Negeri Malang
- Ryansyah, Adi dan Sri Andayani. 2017. *Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen*. *Jurnal. Jurnal Sistem & Teknologi Informasi Komunikasi, Volume : 1, No : 1*. Palembang: Universitas Katolik Musi Charitas
- S, Rosa A, dan M. Shalahuddin. 2016. *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Bandung: Informatika
- Saputra, A. dan Agustin F. 2012. *Membangun Sistem Aplikasi E-Commerce dan SMS*. Jakarta: PT Elex Media Komputindo
- Soleka, Amalia. 2019. *Mengenal 17 Genre Paling Populer dalam Fiksi, Sudah Tahu?*. <https://www.idntimes.com/life/education/amelia-solekha/mengenal-17-genre-paling-populer-dalam-fiksi-c1c2/full>. Diakses pada Kamis, 13 Maret 2020
- Setiawan, Ganang Wahyu. 2011. *Pengujian Perangkat Lunak Menggunakan Metode Black Box Studi Kasus Exexsa Universitas Sanata Dharma*. Skripsi. Yogyakarta: Universitas Sanata Dharma

- Somantri, Oman. 2017. *Text Mining Untuk Klasifikasi Kategori Cerita Pendek Menggunakan Naïve Bayes (NB)*. Jurnal. Jurnal Telematika, vol. 12 no. 1. Bandung: Institut Teknologi Harapan Bangsa
- Sulhaerati, dkk. 2017. *Analisis Tanggapan Terhadap Perusahaan Ritel Raksasa Indomart dan Alfamart*. Laporan Penelitian. Yogyakarta: Universitas Islam Indonesia
- Suyanto. 2017. *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika
- Suyanto. 2018. *Machine Learning Tingkat Dasar dan Lanjut*. Bandung: Informatika
- Wahyudi, Rizki dan Aprilita Dwi Aristantia. 2017. Aplikasi Pengolahan Data Pelanggaran Siswa Pada SMK Yayasan Pendidikan Teknologi 1 Purbalingga Terintegrasi Dengan *SMS Gateway*. Jurnal. *Jurnal Telematika Vol. 10 No. 2*. Banyuma: STMIK Amikom Purwokerto
- Wijaya, Akhmad Pandhu dan Heru Agus Santoso. 2016. *Naïve Bayes Classifier pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government*. *Journal of Applied Intelligent System, Vol. 1, No. 1 (2016)*. Semarang: Universitas Dian Nuswantoro
- Wijaya, Herry Derajad dan Saruni Dwiasnati. 2020. Implementasi *Data Mining* dengan Algoritma *Naïve Bayes* pada Penjualan Obat. *Jurnal Informatika, Vol. 7 no. 1 (2020)*. Jakarta: Universitas Mercu Buana Jakarta
- Zaki, Mohammed J., & Wagner Meira Jr. 2013. *Data Mining and Analysis: Fundamental Concepts and Algorithms*.