

**SISTEM *GRABBING* KONTEN ARTIKEL *WEBSITE* KAMPUS
DI RIAU MENGGUNAKAN METODE *NAIVE BAYES*
CLASSIFIER BERBASIS *WEB MOBILE***

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat
Penyusunan Skripsi Pada Fakultas Teknik
Universitas Islam Riau



RIKO SAPUTRA

143510282

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS ISLAM RIAU

PEKANBARU

2021

**Sistem Grabbing Konten Artikel Website Kampus Di Riau Menggunakan
Metode Naive Bayes Classifier Berbasis Web Mobile**

Riko Saputra

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Riau

e-mail : ricosahputra94@student.uir.ac.id

ABSTRAK

Kota Pekanbaru memiliki 57 perguruan tinggi aktif yang terdiri dari perguruan tinggi negeri dan perguruan tinggi swasta. Dari perguruan tinggi aktif yang berada di Pekanbaru, terdapat 16 perguruan tinggi yang memiliki *website* resmi. Dimana setiap *website* perguruan tinggi tersebut memiliki berita-berita mengenai prestasi-prestasi yang diraih kampus, informasi pemberitaan seperti jadwal pendaftaran mahasiswa baru serta mengenai kegiatan-kegiatan yang dilakukan di suatu kampus. Dari penelitian sebelumnya dikalangan mahasiswa dan siswa di Pekanbaru. Didapatkan bahwa mahasiswa/siswa tertarik dengan berita-berita kampus. Berita-berita yang dominan dicari adalah berita seputar *event* yang diselenggarakan kampus atau kampus lain. Akan tetapi ada beberapa dikalangan mahasiswa/siswa yang kesulitan menemukan berita dan informasi dari beberapa perguruan tinggi yang ada. Oleh Karena itu, dibangunlah portal web yang dapat menampilkan berita-berita dari perguruan tinggi yang ada di Pekanbaru. Portal web ini dibangun untuk memberikan kemudahan bagi mahasiswa/siswa untuk menemukan dan memperoleh berita yang diinginkan dari perguruan tinggi yang ada. Portal web ini dibangun menggunakan teknik *grabbing* dengan metode *naive bayes classifier* berbasis web *mobile*. Teknik *grabbing* merupakan teknik untuk mengintegrasikan semua informasi yang terdapat di *website* dan mengambil data dari tempat atau sumber lain. Dengan adanya sistem ini diharapkan dapat mempermudah mahasiswa/siswa untuk mencari dan mendapatkan informasi dari kampus-kampus yang ada di Pekanbaru.

Kata Kunci : *grabbing, naive bayes classifier, universitas, website*

*Campus Website Articles Grabbing Content System In Riau Using The Naive
Bayes Classifier Method Based On Mobile Web*

Riko Saputra

Informatics Engineering, Faculty of Engineering, Islamic University of Riau

e-mail : ricoshahputra94@student.ac.id

ABSTRAK

Pekanbaru City have 57 universities consisting of national universities and private universities. Of the universities in Pekanbaru, there are 16 universities that have official websites. Where each college website has news about the achievements of the campus, news information such as new student registration schedules and about activities carried out on a campus. From previous research among students in Pekanbaru. It was found that students were interested in campus news. The news that is dominantly sought after is news about events organized by each campuses. However, there are some students who have difficulty finding news and information from several existing universities. Therefore, a web portal was built that can display news from universities in Pekanbaru. This web portal was built to make it easy for students to find and obtain the desired news from existing universities. This web portal was built using the grabbing technique with the mobile web-based Naive Bayes classifier method. Grabbing technique to integrate all the website and retrieve data from other places or sources. With this system, it is hoped that it will make it easier for students to find and get information from campuses in Pekanbaru.

Keyword : *grabbing, naive bayes classifier, university, website*

KATA PENGANTAR

Dengan menyebut nama Allah SWT yang Maha Pengasih lagi Maha Penyayang, Penulis ucapkan puji syukur atas kehadiran Allah SWT, yang telah melimpahkan rahmat dan hidayah-nya, sehingga Penulis dapat menyelesaikan proposal skripsi yang berjudul “ Sistem *Grabbing* Konten Artikel Website Kampus Di Riau Menggunakan Metode *Naive Bayes Classifier* Berbasis Web *Mobile* ” ini tepat pada waktunya.

Dalam penyusunan proposal skripsi ini, penulis mendapatkan bantuan dari berbagai pihak sehingga dapat memperlancar penyusunan proposal ini. Untuk itu Penulis menyampaikan banyak terima kasih kepada semua pihak yang telah berkontribusi dalam pembuatan laporan ini.

Terlepas dari semua itu, penulis menyadari bahwa masih ada kekurangan baik dari bentuk penyusunan maupun materinya. Oleh karena itu kritik dan saran dari pembaca sangat penulis harapkan untuk dapat menyempurnakan proposal ini. Akhir kata penulis berharap semoga proposal skripsi ini dapat memberikan manfaat untuk kita semua.

Pekanbaru, Desember 2021

Riko Saputra

DAFTAR ISI

KATA PENGANTAR.....	i
DAFTAR ISI.....	ii
DAFTAR TABEL.....	v
DAFTAR GAMBAR.....	vii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	3
1.3 Batasan Masalah.....	3
1.4 Rumusan Masalah	4
1.5 Tujuan Penelitian.....	4
1.6 Manfaat Penelitian.....	4
BAB II LANDASAN TEORI	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori.....	7
2.2.1 Perguruan Tinggi.....	7
2.2.2 Berita.....	7
2.2.3 Portal Web	8
2.2.4 Teknik <i>Grabbing</i>	8
2.2.5 <i>Text Mining</i>	9
2.2.5.1 <i>Case folding</i>	9
2.2.5.2 <i>Cleaning</i>	10
2.2.5.3 <i>Tokenizing</i>	10

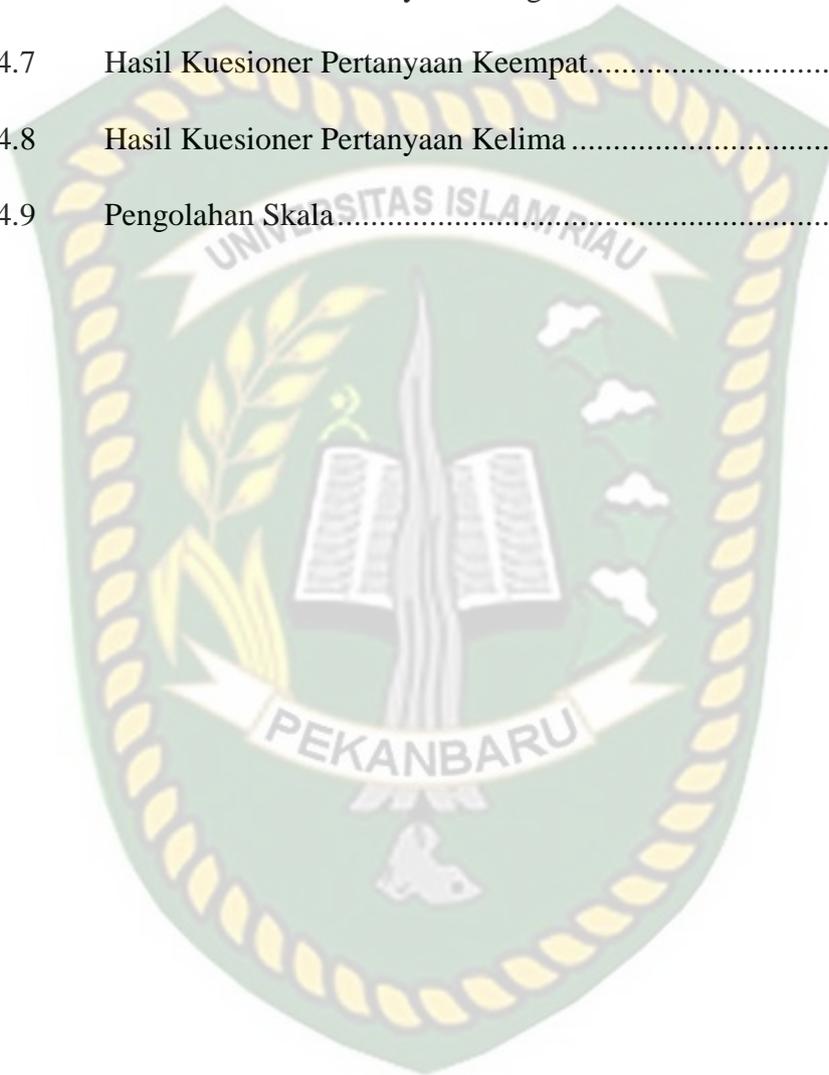
2.2.5.4	Normalisasi.....	10
2.2.5.5	<i>Filtering (Stopword)</i>	10
2.2.5.6	<i>Stemming</i>	10
2.2.5.7	Tahap Pembobotan (TF-IDF)	11
2.2.6	<i>Naive Bayes Classifier</i>	11
2.2.7	<i>Data Flow Diagram</i>	14
2.2.8	<i>Flowchart</i>	15
2.2.9	<i>Hipotesis</i>	16
BAB III METODOLOGI PENELITIAN		17
3.1	Alat Dan Bahan Penelitian Yang Digunakan.....	17
3.1.1	Alat Penelitian	17
3.1.1.1	Spesifikasi Kebutuhan <i>Hardware</i>	17
3.1.1.2	Spesifikasi Kebutuhan <i>Software</i>	18
3.2	Bahan Penelitian.....	18
3.2.1	Jenis Data Penelitian.....	18
3.3	Pengembangan Sistem	19
3.3.1	Penerapan Metode <i>Naive Bayes</i>	20
3.4	Perancangan Sistem.....	33
3.4.1	Diagram Konteks.....	33
3.4.2	<i>Hierarchy Chart</i>	34
3.4.3	<i>Data Flow Diagram (DFD)</i>	35
3.4.3.1	DFD Level 0.....	35

3.4.3.2 DFD Level 1 Proses 1	36
3.4.3.3 DFD Level 1 Proses 3	37
3.4.4 Desain <i>Output</i>	37
3.4.5 Desain <i>Input</i>	38
3.4.6 Perancangan <i>Database</i>	41
3.4.6.1 <i>Entity Relationship Diagram</i> (ERD).....	41
3.4.6.2 Desain <i>Database</i>	42
3.4.7 Desain Logika Program	43
BAB IV HASIL DAN PEMBAHASAN	46
4.1 Hasil Penelitian.....	46
4.2 Pengujian <i>Black Box</i>	46
4.2.1 Pengujian Tampilan Pengunjung	47
4.2.2 Pengujian Tampilan <i>Admin</i>	48
4.3 Implementasi Sistem.....	52
BAB V KESIMPULAN DAN SARAN	59
5.1 Kesimpulan	59
5.2 ... Saran	59
DAFTAR PUSTAKA.....	60

DAFTAR TABEL

Tabel 2.1	Simbol <i>Data Flow Diagram</i>	14
Tabel 2.2	Simbol <i>Flowchart</i>	16
Tabel 3.1	Contoh Data Artikel.....	20
Tabel 3.2	<i>Case folding</i>	22
Tabel 3.3	Hasil <i>Cleaning</i>	22
Tabel 3.4	Hasil <i>Tokenizing</i>	23
Tabel 3.5	Contoh Hasil Normalisasi.....	23
Tabel 3.6	Contoh Hasil Filtering.....	24
Tabel 3.7	Hasil <i>Stemming</i>	24
Tabel 3.8	Data Pembobotan Kata.....	25
Tabel 3.9	Hasil Pembobotan Kata.....	26
Tabel 3.10	Hasil Probabilitas.....	29
Tabel 3.11	Contoh Data Testing (Uji).....	30
Tabel 3.12	Hasil Data Testing (Uji).....	30
Tabel 3.13	Data Latih dan Data Uji.....	31
Tabel 3.14	Tabel User.....	42
Tabel 3.15	Tabel web.....	43
Tabel 3.16	Tabel Berita.....	43
Tabel 4.1	Hasil Pengujian Pengguna.....	53
Tabel 4.2	Skor Maksimum.....	54
Tabel 4.3	Kriteria Skor.....	54

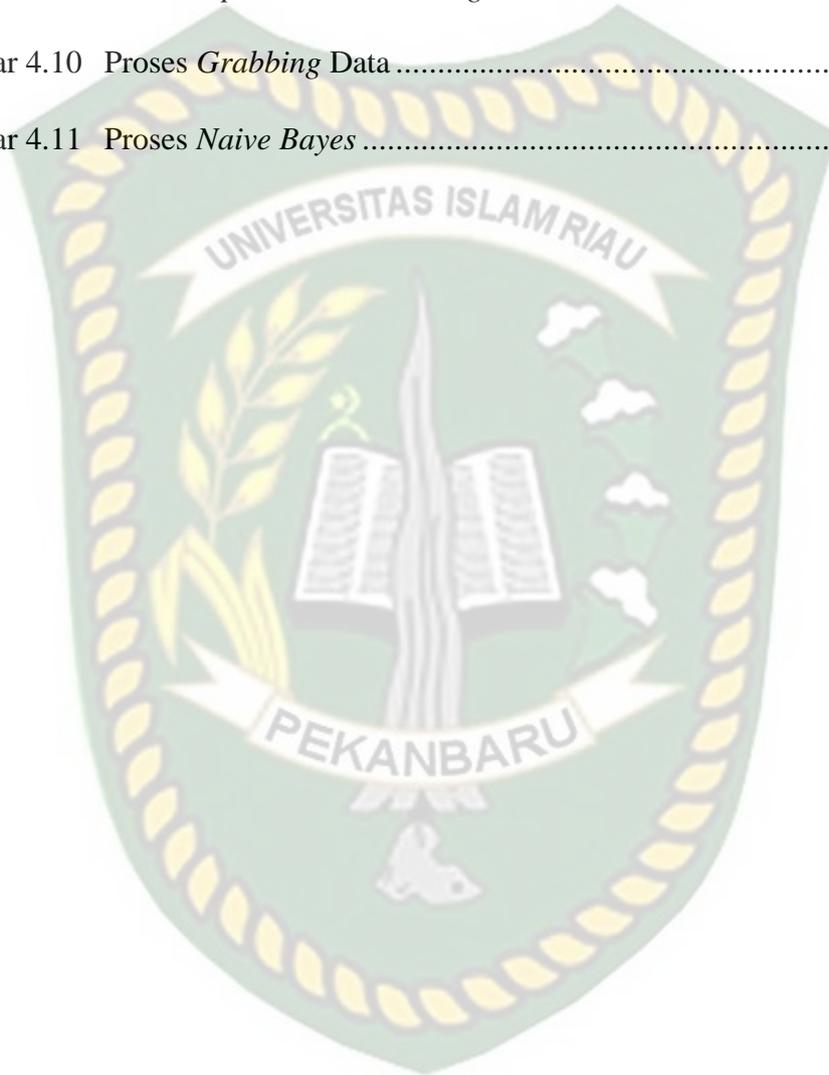
Tabel 4.4	Hasil Kuesioner Pertanyaan Pertama.....	55
Tabel 4.5	Hasil Kuesioner Pertanyaan Kedua	55
Tabel 4.6	Hasil Kuesioner Pertanyaan Ketiga	56
Tabel 4.7	Hasil Kuesioner Pertanyaan Keempat.....	56
Tabel 4.8	Hasil Kuesioner Pertanyaan Kelima	57
Tabel 4.9	Pengolahan Skala.....	57



DAFTAR GAMBAR

Gambar 3.1	<i>Flowchart</i> Sistem.....	19
Gambar 3.2	Alur Tahapan <i>Preprocessing</i>	21
Gambar 3.3	Diagram Konteks	34
Gambar 3.4	<i>Hierarchy Chart</i>	34
Gambar 3.5	DFD Level 0 Sistem Portal <i>Web</i>	36
Gambar 3.6	DFD Level 1 Proses 1 Pengolahan <i>Data Master</i>	36
Gambar 3.7	DFD Level 1 Proses 3 Penilaian Akhir.....	37
Gambar 3.8	Desain <i>Output</i> Portal Berita <i>Admin</i>	38
Gambar 3.9	Rekam <i>Website</i>	39
Gambar 3.10	Rekam <i>Training</i> Berita.....	40
Gambar 3.11	Rekam Data <i>User</i>	41
Gambar 3.12	ERD.....	42
Gambar 3.13	<i>Flowchart</i> Login	44
Gambar 3.14	<i>Flowchart</i> Menu <i>Master</i>	45
Gambar 3.15	<i>Flowchart</i> Proses	45
Gambar 4.1	Pengujian Tampilan <i>Home</i> Pengunjung.....	47
Gambar 4.2	Tampilan <i>Detail</i> Berita.....	47
Gambar 4.3	Pengujian <i>Form</i> Login	48
Gambar 4.4	Tampilan <i>Home</i>	48
Gambar 4.5	Proses <i>Inputan</i> Data <i>Admin</i>	49
Gambar 4.6	Tampilan Data <i>Admin</i>	49

Gambar 4.7	Proses <i>Inputan</i> Data Kampus.....	50
Gambar 4.8	Tampilan Data Kampus.....	50
Gambar 4.9	Proses <i>Inputan</i> Data <i>Training</i>	51
Gambar 4.10	Proses <i>Grabbing</i> Data.....	51
Gambar 4.11	Proses <i>Naive Bayes</i>	52



BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut PDDikti (Pangkalan Data Pendidikan Tinggi) pada tahun ajaran 2018/2019, Provinsi Riau memiliki 100 perguruan tinggi aktif yang tersebar di beberapa daerah, baik perguruan tinggi negeri maupun perguruan tinggi swasta. Dan untuk Kota Pekanbaru itu sendiri terdapat 57 perguruan tinggi, yang terdiri dari perguruan tinggi negeri dan perguruan tinggi swasta. Dari daftar perguruan tinggi aktif yang berada di Pekanbaru, terdapat 39 perguruan tinggi yang memiliki *website* resmi aktif. Dimana setiap *website* perguruan tinggi tersebut memiliki berita-berita mengenai prestasi-prestasi yang diraih kampus, informasi pemberitaan jadwal penting seperti jadwal pendaftaran mahasiswa baru, serta berita mengenai kegiatan-kegiatan yang dilakukan di suatu kampus.

Saat ini hampir seluruh kampus mempunyai *website* yang menampilkan profil dan info dari universitas. Tetapi ternyata, masih banyak calon mahasiswa yang belum mendapatkan info tentang kampus yang nantinya akan menjadi salah satu kandidat kampus pilihan mereka. Calon mahasiswa juga kesulitan, jika harus mendatangi kampus satu-persatu ataupun mengakses *website* resmi dari perguruan tinggi tersebut. Di zaman ini, teknologi sudah sangat berkembang sehingga sangat mudah untuk mendapatkan informasi. Salah satu contohnya ialah dengan memakai sistem portal *website* yang bisa menampilkan artikel/berita yang berasal dari *website* lain dalam satu *website*.

Penelitian terkait penggabungan objek pada *website* telah dilakukan oleh peneliti sebelumnya. Sistem belum memiliki *exception handling* dalam menangani *error*. Masih ada beberapa konten yang ditampilkan di sistem berupa *error* yang disebabkan oleh *website* perguruan tinggi yang menjadi acuan data dan informasi mengalami *error*. Sehingga sistem akan mengalami *error* jika terdapat *error* pada *website* asli (Husin, 2013). Selanjutnya sistem hanya mengambil informasi secara keseluruhan dari *website* asli (Dauni, 2017).

Oleh karena itu, maka dibangunlah sebuah portal web yang dapat menampilkan berita-berita perguruan tinggi yang ada di Provinsi Riau yang khususnya berada di ruang lingkup Kota Pekanbaru. Portal web ini dibangun untuk memberikan kemudahan bagi mahasiswa/siswa untuk menemukan dan memperoleh berita yang diinginkan dari perguruan tinggi di Kota Pekanbaru. Portal web ini akan mengambil berita-berita dari *website* resmi setiap perguruan tinggi dan menampilkannya pada portal web yang akan dibangun. Berita-berita ini akan diperbaharui secara real time. Berita-berita pada portal web akan diperbaharui jika *website* resmi melakukan *update* berita.

Portal web ini akan dibangun menggunakan teknik *grabbing*, yaitu teknik untuk mengintegrasikan semua informasi yang terdapat di *website* untuk mengambil data dari tempat atau dari sumber lain. Dengan metode *grabbing*, informasi yang ada dalam beberapa halaman *website*, dapat diakses melalui satu halaman *website* saja sehingga akan memberikan kemudahan bagi *user* untuk memperoleh informasi dengan cepat dan akurat (Dauni, 2017).

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah yang sudah diuraikan sebelumnya maka dapat diidentifikasi masalah sebagai berikut:

1. Banyak informasi kampus tidak diketahui oleh masyarakat terutama siswa SMA atau Sederajat.
2. Informasi diperlukan seperti informasi beasiswa, informasi penerimaan mahasiswa baru dan informasi lainnya yang dapat mendukung kecepatan penyebaran informasi dikalangan masyarakat terutama siswa SMA sederajat.
3. Bagaimana sebuah portal web dapat menampilkan berita-berita atau artikel dari beberapa *website* universitas yang ada di Pekanbaru dalam satu *website* saja

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini dibatasi dalam beberapa hal, berikut ini adalah penjabaran batasan masalah dari penelitian ini :

1. Sistem informasi dibuat dengan berbasis web. Dimana informasi yang ditampilkan berupa berita-berita dari setiap *website* resmi perguruan tinggi.
2. Sistem ini akan mengambil berita secara *real time* dari berita yang ditambahkan pada *website* resmi setiap perguruan tinggi.
3. Sistem ini hanya menampilkan artikel berita dari website universitas yang berada di Pekanbaru.

4. Sistem ini mempunyai 6 kategori berita yang terdiri dari Kesehatan, IT, Sosial, Budaya, Prestasi dan Olahraga.
5. Sistem dibangun dengan menggunakan *framework* CI dan bahasa pemrograman PHP khususnya library cURL

1.4 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, didapatkan rumusan masalah sebagai berikut :

1. Berita-berita yang ada saat ini masih terpisah pada masing-masing *website* perguruan tinggi.
2. Sistem pada penelitian sebelumnya belum memiliki *error handling*.
3. Membangun portal web berita yang berkategori.

1.5 Tujuan Penelitian

Tujuan dari penelitian ialah dari penelitian ini ialah membuat portal *website* yang dapat menampung berita-berita dari setiap perguruan tinggi di Kota Pekanbaru.

1.6 Manfaat Penelitian

Adapun manfaat yang ingin dicapai dari penelitian ini adalah:

1. Manfaat yang diharapkan dari penelitian ini adalah membantu mahasiswa/siswa dalam menemukan informasi dari *website* resmi setiap perguruan tinggi di Kota Pekanbaru tanpa harus membuka *website* satu per satu.

BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian terdahulu yang dilakukan (Husin, 2013) membangun sistem portal web Perguruan Tinggi Agama Islam Negeri (PTAIN) menggunakan metode *grabbing*. Sistem ini mengambil informasi dari 13 *website* resmi Universitas Islam Negeri (UIN) dan Institut Agama Islam Negeri (IAIN) di Indonesia. Sistem belum memiliki *exception handling* dalam menangani *error*. Sehingga ada beberapa konten yang ditampilkan berupa *error* yang disebabkan oleh *website* perguruan tinggi yang menjadi acuan data dan informasi mengalami *error*. Sistem akan mengalami *error* jika terdapat *error* pada *website* asli. Hasil *output* dari sistem ini adalah menampilkan konten-konten terbaru dari *website* resmi Perguruan Tinggi Agama Islam Negeri.

Penelitian yang dilakukan oleh (Dauni, 2017) membangun sistem penyedia informasi beasiswa menggunakan metode *grabbing*. Sistem ini mengambil informasi dari 3 *website* penyedia beasiswa yaitu *beasiswaparscasarja.com*, *scholars4dev.com* dan *indbeasiswa.com*. Sistem ini belum mengkategorikan berita-berita yang diambil dari *website* asli. Hasil *output* dari sistem ini adalah menampilkan informasi-informasi dari 3 *website* penyedia beasiswa.

Penelitian ketiga yang dilakukan oleh (Efendi, 2019) membangun sistem sebuah aplikasi toko *online* yang bisa secara bersamaan melakukan *multivariable comparison* (Perbandingan Multivariabel). Pada sebuah produk dari beberapa

aplikasi toko *online* yang ada. Multivariable comparison yang digunakan yaitu harga, diskon, rating dan ulasan dari konsumen. Dengan menggunakan teknik *grabbing* langsung ke aplikasi atau web toko *onlinenya*. Dimana konsumen hanya perlu memasukan kata kunci produk yang ingin dicari, lalu pada saat itu juga aplikasi akan menampilkan produk yang relevan. Setelah konsumen memilih produk yang sesuai maka sistem akan otomatis menampilkan harga produk, ulasan produk, dan diskon produk dari beberapa situs toko *online* serta mengurutkan dari harga terendah hingga harga tertinggi untuk produk yang dipilih, ketika konsumen memilih salah satu dari daftar perbandingan tersebut maka pengguna akan langsung diarahkan ke toko *online* yang menjual produk tersebut. Tujuannya mempermudah dan mempersingkat waktu para konsumen untuk membandingkan produk yang diinginkan dengan hanya membuka satu aplikasi. Aplikasi ini bisa digunakan *smartphone* yang berbasis *android*.

Penelitian keempat yang dilakukan oleh (Pratama, 2018) membangun sistem plagiarisme dengan metode *cosine similarity*, metode *cosine similarity* merupakan salah satu metode yang dapat digunakan untuk menghitung nilai kemiripan antar dokumen. Tahapan yang dilakukan sistem untuk menghasilkan nilai kemiripan antar dokumen yaitu dengan membandingkan dokumen jurnal yang diupload dengan dokumen *repository* yang diperoleh dari hasil *grabbing* data DOAJ dan tersimpan di *database*. Perhitungan metode yang dilakukan akan diperoleh presentasi nilai kemiripan antar dokumen.

2.2 Landasan Teori

2.2.1 Perguruan Tinggi

Perguruan tinggi adalah satuan pendidikan penyelenggara pendidikan tinggi. Peserta didiknya disebut mahasiswa, sedangkan tenaga pendidiknya disebut dosen. Dibentuknya perguruan tinggi bukan tanpa tujuan, tujuan perguruan tinggi ialah untuk menyelenggarakan pendidikan tinggi yang mencakup program diploma, program sarjana, program magister, program doctor, dan lain sebagainya. Perguruan tinggi di Indonesia terbagi menjadi dua yaitu terdiri dari perguruan tinggi negeri dan perguruan tinggi swasta. Perguruan tinggi negeri penyelenggaraannya dilakukan oleh pemerintah, sedang perguruan tinggi swasta penyelenggaraannya dilakukan oleh pihak swasta. Di Indonesia, perguruan tinggi dapat berbentuk universitas, institut, sekolah tinggi, politeknik, dan akademi.

2.2.2 Berita

Berita adalah laporan mengenai suatu peristiwa atau kejadian yang terbaru, laporan mengenai fakta-fakta yang aktual, menarik perhatian, dinilai penting atau luar biasa (Mahmudy dan Widodo, 2014). Menurut (Dyoti, 2003), berita adalah semua peristiwa yang sudah atau akan terjadi yang perlu diketahui oleh manusia sebagai pembaca atau pemirsa. Secara umum, penulisan berita menggunakan model deduktif, artinya pembeberan fakta dimulai dari hal-hal yang bersifat umum hingga hal yang khusus. Dari yang paling penting ke hal yang kurang penting. Struktur tulisannya dikenal dengan bentuk piramida terbalik.

2.2.3 Portal Web

Menurut (Hidayat, 2010), *website* atau situs dapat diartikan sebagai kumpulan halaman-halaman yang digunakan untuk menampilkan informasi teks, gambar, animasi, suara dan/atau gabungan dari semuanya, baik yang bersifat statis maupun dinamis. Sehingga membentuk satu rangkaian bangunan yang saling terkait, yang masing-masing dihubungkan dengan jaringan-jaringan halaman. Sedangkan menurut (Simarmata, 2010), web dapat diartikan sebagai alat untuk menciptakan sistem informasi global yang mudah berdasarkan pada hiperteks.

Menurut (Zabir, 2008), portal web merupakan situs yang memberikan pengguna beranda pribadi dengan akses penuh ke semua informasi dari seluruh web, serta *dashboard* yang mengirimkan agregasi konten penting bagi pengguna. Portal web dirancang khusus untuk menyatukan informasi dari berbagai sumber dengan cara (*format* atau *layout*) yang seragam. Setiap sumber informasi mendapat area khusus pada halaman *website* portal dalam menampilkan informasinya (Bumi, 2003).

2.2.4 Teknik *Grabbing*

Teknik *grabbing* adalah suatu teknik untuk mengambil text atau secara umum obyek yang ada pada situs lain yang nantinya dipergunakan untuk mendapatkan data dari tampilan sebuah website. Untuk melakukan *grabbing*, bias menggunakan CURL. Dengan CURL ini, akan membaca semua kode HTML dari suatu situs lalu mengambil informasi yang di inginkan dan kemudian informasi tersebut ditampilkan dalam satu website.

2.2.5 Text Mining

Text mining dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang *user* berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan komponen-komponen dalam data *mining* yang salah satunya adalah kategorisasi. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan pada *text mining* adalah kumpulan teks yang memiliki *format* yang tidak terstruktur atau minimal semi terstruktur. Adapun tugas khusus dari *text mining* antara lain yaitu mengkategorikan teks (*text categorization*) dan pengelompokan teks (*text clustering*).

Text mining merupakan bagian dari data *mining* yang berfungsi untuk menemukan pola yang menarik dari kumpulan data yang berjumlah besar (Kurniawan, 2012). *Text mining* juga merupakan sebuah proses pengetahuan intensif dimana pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis (Feldman dan Sanger, 2007). *Text Mining* memiliki 6 tahapan proses pokok, yaitu pemrosesan awal *text* (*text preprocessing*), transformasi teks (*text transformation*) atau *feature generation* dan pemilihan fitur (*feature selection*).

Penjelasan proses tersebut sebagai berikut :

2.2.5.1 Case folding

Case folding merupakan proses pengubahan semua huruf kapital dalam dokumen menjadi huruf kecil, hanya huruf a-z yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.

2.2.5.2 Cleaning

Cleaning merupakan proses membersihkan dokumen dari komponen-komponen yang tidak mempunyai hubungan dengan informasi yang ada pada dokumen, seperti karakter atau simbol, angka, emoji ataupun *link* URL.

2.2.5.3 Tokenizing

Tokenizing/Tokenisasi adalah proses memecah teks menjadi kata, frasa, sesuai dengan kata penyusunnya. Tujuan tokenizing adalah eksplorasi kata-kata dalam kalimat (Gurusamy & Kannan, 2014).

2.2.5.4 Normalisasi

Normalisasi adalah proses untuk mengubah kata-kata yang tidak baku menjadi kata baku sesuai dengan Kamus Besar Bahasa Indonesia (KBBI). Misalnya kata singkatan seperti ‘dgan’ menjadi kata ‘dengan’, ‘kpd’ menjadi kata ‘kepada’ dan lain sebagainya.

2.2.5.5 Filtering (Stopword)

Pada langkah ini dilakukan penghapusan kata-kata yang tidak memiliki makna yang cukup berarti. Salah satu contoh *stopword* seperti: dari, dengan, maka, yang, dll (Gurusamy dan Kannan, 2014).

2.2.5.6 Stemming

Stemming proses yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke bentuk dasarnya. Pada tahap ini dilakukan proses menghilangkan imbuhan seperti ‘ke’, ‘ber’, ‘me’, dll. Menjadi bentuk kata dasarnya (Gurusamy dan Kannan, 2014). Misalnya terdapat kata ‘mengambil’ maka akan diubah menjadi kata ‘ambil’.

2.2.5.7 Tahap Pembobotan (TF-IDF)

Pembobotan TF-IDF merupakan metode untuk memberikan bobot hubungan suatu kata (*term*) terhadap suatu dokumen berdasarkan dua konsep, yaitu frekuensi kemunculan kata di dalam sebuah dokumen dan inverse dokumen yang mengandung kata tersebut.

2.2.6 *Naive bayes Classifier*

Pengklasifikasian *Naive Bayes* didasari oleh teorema *bayes* yang ditemukan oleh Thomas *Bayes* pada abad ke-18. Dalam studi perbandingan algoritma klasifikasi telah ditemukan simple *bayes-ian* atau yang biasa dikenal dengan *Naive Bayes Classifier* (NBC). Algoritma *Naive Bayes Classifier* telah banyak digunakan dalam penelitian mengenai *text mining* karena memiliki kelebihan yaitu algoritma sederhana tapi memiliki akurasi yang tinggi (Rish, 2001). Metode *Naive Bayes Classifier* untuk klasifikasi atau kategorisasi teks menggunakan atribut kata yang muncul dalam suatu dokumen sebagai dasar klasifikasinya (Hamzah, 2012).

Metode *Naive Bayes Classifier* menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Pada tahap pelatihan dilakukan proses analisis terhadap sampel dokumen berupa pemilihan *vocabulary*, yaitu kata yang mungkin muncul dalam koleksi dokumen sampel yang sebisa mungkin dapat menjadi representasi dokumen. Selanjutnya adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel dokumen. Pada tahap klasifikasi ditentukan nilai kategori dari suatu dokumen berdasarkan *term* yang muncul dalam dokumen yang diklasifikasi.

Keterangan :

$D = \{d_i, |i = 1, 2, \dots, |D|\} = \{d_1, d_2, \dots, d_{|D|}\}$: koleksi dokumen

$C = \{c_j | j = 1, 2, \dots, |C|\} = \{c_1, c_2, \dots, c_{|C|}\}$: koleksi kategori

$P = P(d_i | c_j)$: peluang dokumen i pada kategori j

Klasifikasi *Naive Bayes Classifier* dilakukan dengan cara mencari probabilitas, yaitu probabilitas kategori v_j jika diketahui dokumen d_i . Dokumen d_i dianggap sebagai *tuple* dari kata-kata dalam dokumen, yaitu a_1, a_2, \dots, a_n , yang frekuensi kemunculannya diasumsikan sebagai variabel acak dengan distribusi probabilitas (McCallum & Nigam, 1998). Selanjutnya klasifikasi dokumen adalah mencari nilai maksimum dari :

$$V_{\text{MAP}} = \operatorname{argmax}_{c_j \in C} (c_j | a_1, a_2, \dots, a_n) \quad (1)$$

Teorema *Bayes* menyatakan tentang probabilitas bersyarat menyatakan :

$$P(B | A) = \frac{P(A | B) P(B)}{P(A)} \quad (2)$$

Dengan menerapkan teorema *Bayes*, persamaan (1) dapat ditulis :

$$V_{\text{MAP}} = \operatorname{argmax}_{c_j \in C} \frac{P(a_1, a_2, \dots, a_n | c_j) P(c_j)}{P(a_1, a_2, \dots, a_n)} \quad (3)$$

Karena nilai $P(a_1, a_2, \dots, a_n)$ untuk semua c_j besarnya sama maka nilainya dapat diabaikan, sehingga persamaan (3) menjadi :

$$V_{\text{MAP}} = \operatorname{argmax}_{v_j \in V} P(a_1, a_2, \dots, a_n | c_j) P(c_j) \quad (4)$$

Dengan mengasumsikan bahwa setiap kata dalam $\langle a_1, a_2, \dots, a_n \rangle$ adalah independent, maka $P(a_1, a_2, \dots, a_n | c_j)$ dalam persamaan (4) dapat ditulis sebagai:

$$P(a_1, a_2, \dots, a_n | c_j) = \prod_i P(a_i | c_j) \quad (5)$$

Sehingga persamaan (4) dapat ditulis:

$$V_{\text{MAP}} = \operatorname{argmax}_{v_j \in V} P(c_j) \prod_i P(a_i | c_j) \quad (6)$$

Nilai $P(v_j)$ ditentukan pada saat pelatihan, yang nilainya didekati dengan :

$$P(c_j) = \frac{|doc_j|}{|Contoh|} \quad (7)$$

Dimana $|doc_j|$ adalah banyaknya dokumen *testing* yang memiliki kategori j dalam pelatihan, sedangkan $|Contoh|$ banyaknya dokumen *training* dalam contoh yang digunakan untuk pelatihan. Untuk nilai $P(c_j)$, yaitu probabilitas kata w_k dalam kategori j ditentukan dengan :

$$P(w_k | c_j) = \frac{n_k + 1}{n + |vocabulary|} \quad (8)$$

Dimana n_k adalah frekuensi munculnya kata w_k dalam dokumen yang berkategori c_j , sedangkan nilai n adalah banyaknya seluruh kata dalam dokumen berkategori c_j , dan $|vocabulary|$ adalah banyaknya kata dalam contoh pelatihan. Tahapan klasifikasi dilakukan dengan proses *preprocessing* dokumen, proses *training* dan proses klasifikasi. *Preprocessing* dilakukan dengan mengidentifikasi kata unik dalam seluruh dokumen yang dipilih sebagai contoh, membuang *stop word* seperti kata 'ini', 'itu', 'yang', dan lain-lain. Dan *feature selection* kata kunci sebagai *vocabulary* berdasarkan frekuensi kemunculan kata dalam dokumen. Selanjutnya disusun matriks kata-dokumen yang mencatat frekuensi kemunculan kata dalam tiap dokumen.

2.2.7 Data Flow Diagram

Data Flow Diagram adalah suatu diagram yang menggambarkan aliran data dari sebuah proses yang sering disebut dengan sistem informasi. Di dalam *data flow diagram* juga menyediakan informasi mengenai *input* dan *output* dari tiap entitas dan proses itu sendiri. DFD adalah representasi grafik dari sebuah sistem. DFD menggambarkan komponen-komponen sebuah sistem, aliran-aliran data diantara komponen-komponen tersebut, asal, tujuan dan penyimpanan dari data tersebut.

DFD merupakan model proses data yang dibuat ataupun dirancang untuk menggambarkan aliran data, mulai darimana data tersebut masuk dan kemana tujuan dari data itu sendiri. Secara umum, DFD adalah model data atau data yang memberikan gambaran asal data dan kemana tujuan data (*input* dan *output*) di dalam sebuah sistem. Berikut ini ialah symbol dari DFD.

Tabel 2.1 Simbol Data Flow Diagram

Simbol	Keterangan	Fungsi
	Entitas Luar	Menunjukkan tempat asal data atau <i>sumber</i> data.
	Proses	Proses menunjukkan transformasi dari input menjadi output
	Penyimpanan data	Penyimpanan data yang dapat berupa file atau database
	Arus data	Menunjukkan aliran data pada sistem

2.2.8 *Flowchart*

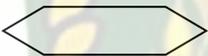
Flowchart adalah diagram yang menampilkan langkah-langkah dan keputusan untuk melakukan sebuah proses dari suatu program. Setiap langkah digambarkan dalam bentuk simbol dan dihubungkan dengan garis atau panah. Dimana masing masing simbol merepresentasikan kegiatan tertentu. *Flowchart* membantu analis dan programmer untuk memecahkan masalah kedalam segmen-segmen yang lebih kecil dan membantu dalam menganalisis alternatif-alternatif dalam pengoperasian.

Flowchart diawali dengan penerimaan *input* dan diakhiri dengan penampilan *output*. *Flowchart* adalah suatu gambaran yang menjelaskan urutan:

1. Pembacaan data.
2. Pemrosesan data.
3. Pengambilan keputusan terhadap data.
4. Penyajian hasil pemrosesan data.

Simbol-simbol *flowchart* yang bisa dipakai adalah simbol-simbol *flowchart standart* yang dikeluarkan oleh ANSI dan ISO. Berikut ini akan dibahas tentang simbol-simbol yang digunakan untuk menyusun *flowchart* adalah:

Tabel 2.3 Simbol *Flowchart*

No.	Simbol	Fungsi
1		Terminal, simbol yang menyatakan awal atau akhir suatu program.
2		Proses, simbol yang menyatakan suatu proses yang dilakukan <i>computer</i> .
3		<i>Input-output</i> , simbol yang menyatakan proses <i>input / output</i> .
4		<i>Decision</i> , simbol yang menunjukkan kondisi tertentu yang menghasilkan dua kemungkinan jawaban, yaitu ya dan tidak.
5		<i>Predefined</i> proses, suatu simbol untuk menyediakan tempat-tempat pengolahan data dalam <i>storage</i> .
6		<i>Connector</i> , <i>Penghubung alur dalam halaman yang sama</i> .
7		<i>Offline Connector</i> , menghubungkan proses dalam lembar kerja yang berbeda.
8		Arus / flow, simbol yang digunakan untuk menggabungkan antara simbol yang satu dengan simbol yang lain
9		Menyatakan sekumpulan langkah proses yang ditulis sebagai prosedur.

2.2.9 Hipotesis

Dengan adanya portal web mobile informasi kampus dapat menjadi sarana untuk menyebarkan informasi kampus-kampus di Pekanbaru menjadi lebih cepat.

BAB III

METODOLOGI PENELITIAN

3.1 Bahan Dan Alat Penelitian Yang Digunakan

Bahan dan alat yang digunakan dalam penelitian ini adalah sebagai berikut:

3.1.1 Alat Penelitian

Dalam penelitian ini memerlukan bahan dan alat sebagai pendukung perancangan sistem portal web dengan teknik *grabbing* dan menggunakan metode *naive bayes*, yang memiliki tahapan seperti yang disebutkan pada tinjauan pustaka yaitu tahap eksplorasi, perencanaan dan iterasi pengembangan sistem. Spesifikasi perangkat keras yang dipakai untuk perancangan di penelitian ini adalah :

3.1.1.1 Spesifikasi Perangkat Keras (*Hardware*)

Dalam menjalankan aplikasi ini dibutuhkan perangkat keras (*hardware*) yang mumpuni, agar aplikasi dapat berjalan dengan baik. Untuk itu spesifikasi perangkat keras haruslah memenuhi standar kebutuhan aplikasi, adapun minimum aplikasi terhadap struktur komputer adalah:

1. *Processor* : *Intel Celeron 1007U*
2. *RAM* : *4,00 GB*
3. *Hardisk* : *500 GB*
4. *System Type* : *64 bit Operating System*

3.1.1.2 Spesifikasi Perangkat Lunak (*Software*)

Software yang dipergunakan dalam pembuatan sistem portal web dengan teknik *grabbing* dan menggunakan metode *naive bayes* untuk konten *website* portal kampus adalah:

1. Sistem Operasi : *Microsoft Windows 10*
2. Bahasa Pemrograman : *HTML, PHP*
3. *Database Management System* : *MySQL*
4. *Browser* : *Google Chrome*
5. Desain Logika Program : *Microsoft Visio 2016*

3.2 Bahan Penelitian

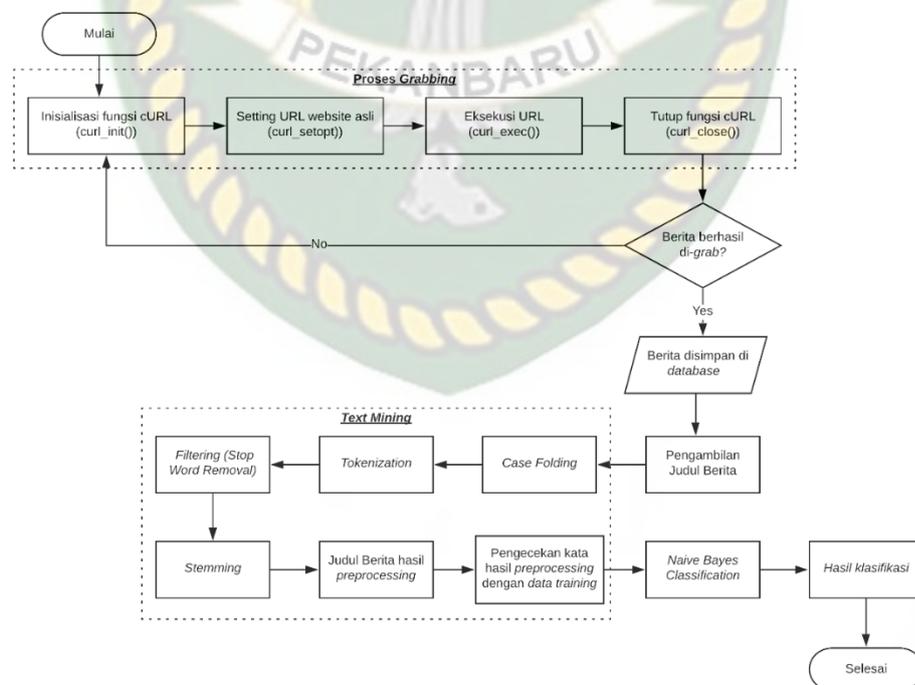
3.2.1 Jenis Data Penelitian

Adapun jenis data yang digunakan dalam penelitian ini adalah data primer yang dikumpulkan melalui observasi terhadap *website-website* kampus yang ada di Kota Pekanbaru, sehingga didapat data-data sebagai berikut:

1. Data mengenai daftar *website* di ambil dari Pangkalan Data Pendidikan Tinggi (PDDikti) mengenai kampus yang terdaftar di Kota Pekanbaru.
2. Bagaimana cara pengambilan data *grabbing* dari masing-masing *website* kampus dengan metode cURL.

3.3. Pengembangan Sistem

Pada penelitian yang dilakukan akan membangun sistem *web mobile* untuk melakukan *grabbing* konten *website* dan melakukan proses klasifikasi kategori menggunakan metode *naive bayes*. Pada pengembangan sistem dijelaskan alur kerja dari sistem portal *website* yang akan dibangun. Alur kerja dimulai dari pengguna yang sudah terkoneksi internet melakukan *request* ke *web server*. *Web server* menampilkan berita-berita yang diambil dari setiap web perguruan tinggi dengan teknik *grabbing*. Berita-berita yang diambil akan disimpan di *database*. Kemudian *web server* akan memproses permintaan pengguna dan menampilkan *response* kepada pengguna sesuai berita yang telah di *request*. Berikut adalah *flowchart* sistem yang dibangun.



Gambar 3.1 Flowchart Sistem

3.3.1 Penerapan Metode *Naive Bayes*

Pada contoh penerapan metode akan menggunakan artikel yang telah dilakukan *grabbing* menggunakan CURL. Berikut ini adalah proses pengambilan artikel menggunakan CURL.

```
$data = curl_init();
curl_setopt($data, CURLOPT_RETURNTRANSFER, 1);
curl_setopt($data, CURLOPT_URL, 'https://sar.ac.id/');
$hasil = curl_exec($data);
curl_close($data);
```

Hasil dari pengambilan data seperti contoh artikel berikut ini:

Tabel 3.1 Contoh Data Artikel

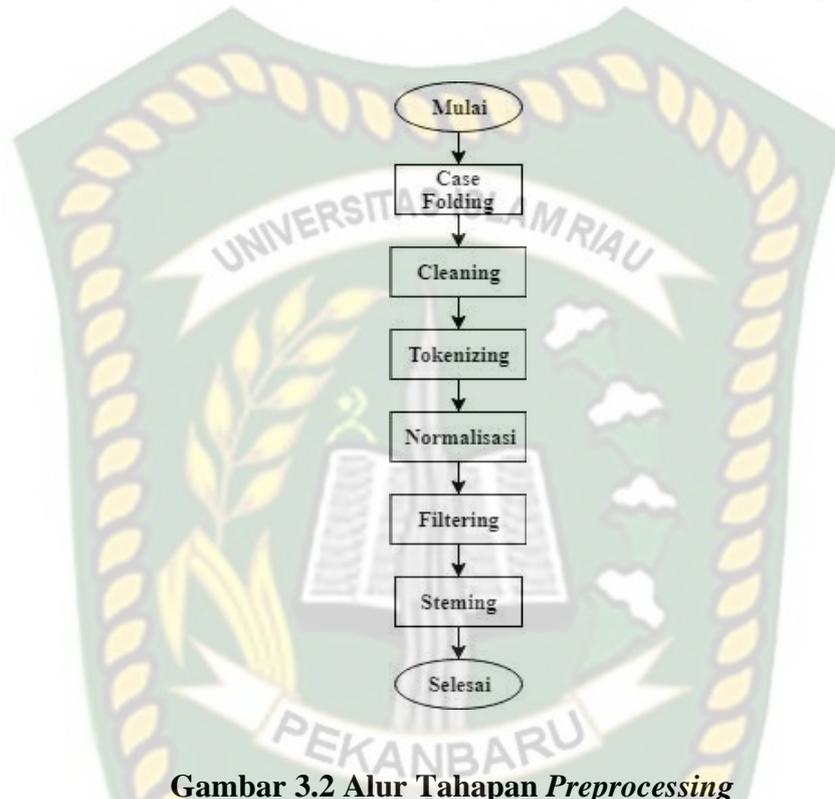
No	Judul	Kategori
1	Mahasiswa UIR terpilih mengikuti program bangkit kampus merdeka	Prestasi
2	Mahasiswa UIR Juara 1 nasional piala Panglima TNI	Olahraga
3	Sidang Terbuka Senat UIR Wisuda Daring 2020	Sosial

Setelah proses pengambilan data pelabelan pada setiap judul artikel, selanjutnya dilakukan proses *Preprocessing*.

1. Tahap *Preprocessing*

Preprocessing merupakan salah satu tahapan menghilangkan permasalahan-permasalahan yang dapat mengganggu hasil daripada proses data.

Dalam kasus klasifikasi dokumen yang menggunakan data bertipe teks, terdapat beberapa macam proses yang dilakukan diantaranya *case folding*, *stemming* dan sebagainya. Berikut ini merupakan tahapan pada *preprocessing* pada gambar 3.2.



Gambar 3.2 Alur Tahapan *Preprocessing*

2. *Case folding*

Case Folding adalah tahapan proses mengubah semua huruf dalam teks dokumen menjadi huruf kecil, serta menghilangkan karakter selain a-z. berikut ini adalah contoh hasil data artikel yang telah di proses.

Tabel 3.2 Hasil *Case folding*

No	Judul	Hasil <i>Case Folding</i>
1	Mahasiswa UIR Terpilih Mengikuti Program Bangkit Kampus Merdeka	mahasiswa uir terpilih mengikuti program bangkit kampus merdeka
2	Mahasiswa UIR Juara 1 Nasional Piala Panglima TNI	mahasiswa uir juara 1 nasional piala panglima tni
3	Sidang Terbuka Senat UIR Wisuda Daring 2020	sidang terbuka senat uir wisuda daring 2020

3. *Cleaning*

Cleaning merupakan tahapan membersihkan dokumen dari komponen-komponen yang tidak mempunyai hubungan dengan informasi yang ada pada dokumen, seperti karakter atau simbol, *emoji* dan *link* URL.

Tabel 3.3 Hasil *Cleaning*

No	Judul	Hasil <i>Cleaning</i>
1	mahasiswa uir terpilih mengikuti, program bangkit kampus merdeka	mahasiswa uir terpilih mengikuti program bangkit kampus merdeka
2	mahasiswa uir juara 1 nasional piala panglima tni!	mahasiswa uir juara 1 nasional piala panglima tni
3	sidang terbuka senat uir wisuda daring 2020?	sidang terbuka senat uir wisuda daring 2020

4. *Tokenizing*

Tokenizing/Tokenisasi adalah proses memecah teks menjadi kata, frasa, sesuai dengan kata penyusunnya. Tujuan tokenizing adalah eksplorasi kata-kata dalam kalimat (Gurusamy & Kannan, 2014). Berikut ini adalah contoh hasil data artikel yang telah di tokenisasi.

Tabel 3.4 Hasil *Tokenizing*

D1	D2	D3
mahasiswa	mahasiswa	sidang
uir	uir	terbuka
terpilih	juara	senat
mengikuti	1	uir
program	nasional	wisuda
bangkit	piala	daring
kampus	panglima	2020
merdeka	tni	

5. Normalisasi

Proses normalisasi merupakan proses untuk mengubah kata-kata singkatan dan kata yang tidak baku atau tidak sesuai ejaan/menjadi kata yang baku. Contoh kata 'dgn' menjadi 'dengan', 'analisa' menjadi 'analisis', dan lain sebagainya.

Tabel 3.5 Contoh Hasil Normalisasi

No	Kata Tidak Baku	Hasil Normalisasi
1	aktifitas	aktivitas
2	diagnosa	diagnosis
3	kpd	kepada

6. *Filtering (Stopword)*

Pada langkah ini dilakukan penghapusan kata-kata yang tidak memiliki makna yang cukup berarti. Salah satu contoh *stopword* seperti: dari, dengan, maka, yang, dll (Gurusamy dan Kannan, 2014).

Tabel 3.6 Contoh Hasil Filtering

No	Contoh Kata	Hasil <i>Filtering</i>
1	manajemen	manajemen
2	pengetahuan	pengetahuan
3	adalah	konsep
4	sebuah	baru
5	konsep	dunia
6	baru	bisnis
7	di	
8	dunia	
9	bisnis	

7. *Stemming*

Stemming proses yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke bentuk dasarnya. Pada tahap ini dilakukan proses menghilangkan imbuhan seperti ‘ke’, ‘ber’, ‘me’, dll, menjadi bentuk kata dasarnya (Gurusamy & Kannan, 2014). Misalnya terdapat kata ‘mengambil’ maka akan menjadi kata ‘ambil’. Berikut ini adalah contoh hasil data artikel yang telah ditransformasi dari kata sebelumnya.

Tabel 3.7 Hasil *Stemming*

D1	D2	D3
mahasiswa	mahasiswa	sidang
uir	uir	buka
pilih	juara	senat
ikuti	1	uir
program	nasional	wisuda
bangkit	piala	daring
kampus	Panglima	2020
merdeka	tni	

8. Tahap Pembobotan (TF-IDF)

Pembobotan adalah proses mengubah kata menjadi bentuk angka atau *vector*. Sedangkan TF-IDF merupakan untuk menentukan nilai frekuensi sebuah kata di dalam sebuah dokumen dan pemberian bobot pada setiap kata pada setiap dokumen untuk mencari dan menghitung berapa kali kata itu muncul. Berikut ini adalah hasil dari stemming sebelumnya, yang nantinya akan dilakukan tahap pembobotan pada hasil stemming tersebut :

Tabel 3.8 Data Pembobotan Kata

D1	D2	D3
mahasiswa	mahasiswa	sidang
uir	uir	buka
pilih	juara	senat
ikuti	1	uir
program	Nasional	wisuda
bangkit	piala	daring
kampus	panglima	2020
merdeka	tni	

Dari tabel 3.8 akan dilakukan tahapan pembobotan dengan menghitung TF-IDF dari data tersebut.

Menghitung *invers document frequency* dengan rumus :

$$idf = \log\left(\frac{tf}{df}\right)$$

Dimana, TF = *term frequency* (kemunculan kata didalam sebuah dokumen)

DF = *document frequency* (frekuensi dokumen yang mengandung kata tersebut)

Setelah menghitung nilai idf, selanjutnya tentukan nilai dari TF-IDF dengan mengkalikan $tf \times idf$

Berikut ini adalah hasil dari pembobotan data dari Tabel 3.8 yang dapat dilihat pada tabel 3.9

Tabel 3.9 Hasil Pembobotan kata

Term (t)	TF			DF	IDF	TF.IDF		
	D1	D2	D3			D1	D2	D3
mahasiswa	1	1	0	2	$\log(3/2)=0,176$	0,176	0,176	0
uir	1	1	1	3	$\log(3/3)=0$	0	0	0
pilih	1	0	0	1	$\log(3/1)=0,477$	0,477	0	0
ikut	1	0	0	1	$\log(3/1)=0,477$	0,477	0	0
program	1	0	0	1	$\log(3/1)=0,477$	0,477	0	0
bangkit	1	0	0	1	$\log(3/1)=0,477$	0,477	0	0
kampus	1	0	0	1	$\log(3/1)=0,477$	0	0,477	0
merdeka	1	0	0	1	$\log(3/1)=0,477$	0	0,477	0
juara	0	1	0	1	$\log(3/1)=0,176$	0	0,477	0
1	0	1	0	1	$\log(3/1)=0,477$	0	0,477	0
nasional	0	1	0	1	$\log(3/1)=0,477$	0	0,477	0
piala	0	1	0	1	$\log(3/1)=0,477$	0	0,477	0
panglima	0	1	0	1	$\log(3/1)=0,477$	0	0,477	0
tni	0	1	0	1	$\log(3/1)=0,477$	0	0,477	0
sidang	0	0	1	1	$\log(3/1)=0,477$	0	0	0,477
buka	0	0	1	1	$\log(3/1)=0,477$	0	0	0,477
senat	0	0	1	1	$\log(3/1)=0,477$	0	0	0,477
wisuda	0	0	1	1	$\log(3/1)=0,477$	0	0	0,477
daring	0	0	1	1	$\log(3/1)=0,477$	0	0	0,477
2020	0	0	1	1	$\log(3/1)=0,477$	0	0	0,477

9. Tahap Klasifikasi *Naive Bayes*

Berdasarkan pada Analisa kebutuhan data dalam melakukan proses *text mining* pada sebelumnya, pada proses ini akan menjelaskan proses-proses yang akan dilakukan dalam klasifikasi data.

Dalam proses klasifikasi *Naive Bayes* ini, *dataset* akan dibagi menjadi proses data *training* (latih) dan proses data *testing* (uji). Berikut merupakan penjelasan langkah-langkah dan contoh perhitungan.

1. Data *training* (latih)

Pada tahap ini, nilai pada data yang telah diketahui bobotnya akan dijadikan data latih sebagai acuan dalam membuat model klasifikasi. Kemudian akan dicari nilai probabilitas kategori dan probabilitas masing-masing kata pada setiap *term* untuk setiap kelas dari data latih. Berikut adalah perhitungannya:

- a. Pertama hitung probabilitas setiap kategori (*prior*), pada penelitian ini yang menjadi kategori ada yaitu kategori Kesehatan, IT, Sosial, Budaya, Prestasi dan Olahraga.

$$P(\text{kategori}) = \frac{d(\text{kategori})}{|c|}$$

$$P(\text{Kesehatan}) = \frac{d(\text{Kesehatan})}{|c|} = \frac{1}{6}$$

$$P(\text{IT}) = \frac{d(\text{IT})}{|c|} = \frac{1}{6}$$

$$P(\text{Sosial}) = \frac{d(\text{Sosial})}{|c|} = \frac{1}{6}$$

$$P(\text{Budaya}) = \frac{d(\text{Beasiswa})}{|c|} = \frac{1}{6}$$

$$P(\text{Prestasi}) = \frac{d(\text{Prestasi})}{|c|} = \frac{1}{6}$$

$$P(\text{Olahraga}) = \frac{d(\text{Olahraga})}{|c|} = \frac{1}{6}$$

- b. Kemudian dihitung probabilitas pada setiap *term* dari semua dokumen. jumlah keseluruhan *term* yang digunakan pada perhitungan ini adalah 21 (tabel 3.8), 10 *term* kategori prestasi, 10 *term* kategori olahraga dan 9 *term* dari kategori sosial. Jumlah *term* tergantung dari hasil *preprocessing* data. Berikut adalah merupakan perhitungan probabilitas pada setiap *term*:

$$P(W_k | \text{kategori}) = \frac{(n_k, \text{Kategori}) + 1}{(n, \text{Kategori}) + |\text{kosakata}|}$$

Diketahui:

$$|\text{kosakata}| = 22$$

$$\text{Term Prestasi} = 10$$

$$\text{Term Olahraga} = 10$$

$$\text{Term Sosial} = 9$$

- i. Probabilitas kata “ada”

$$\begin{aligned} P(\text{"mahasiswa"} | \text{"Prestasi"}) &= \frac{(\text{"mahasiswa"} | \text{"Prestasi"}) + 1}{(\text{"Prestasi"}) + |\text{kosakata}|} = \frac{0.176 + 1}{10 + 22} \\ &= 0.037 \end{aligned}$$

$$P(\text{"mahasiswa"}|\text{"Olahraga"}) = \frac{(\text{"mahasiswa"}|\text{"Olahraga"}) + 1}{(\text{"Olahraga"}) + |\text{kosakata}|} = \frac{0.176 + 1}{10 + 22} = 0.037$$

$$P(\text{"mahasiswa"}|\text{"Sosial"}) = \frac{(\text{"mahasiswa"}|\text{"Sosial"}) + 1}{(\text{"Sosial"}) + |\text{kosakata}|} = \frac{0 + 1}{9 + 22} = 0.032$$

Pada proses ini secara keseluruhan dapat dilihat pada tabel berikut ini :

Tabel 3.10 Hasil Probablitas

Term (t)	Probabilitas		
	Prestasi	Olahraga	Sosial
mahasiswa	0,037	0,037	0,032
uir	0,03125	0,03125	0,032258
pilih	0,046156	0,03125	0,032258
ikut	0,046156	0,03125	0,032258
program	0,046156	0,03125	0,032258
bangkit	0,046156	0,03125	0,032258
kampus	0,03125	0,046156	0,032258
merdeka	0,03125	0,046156	0,032258
juara	0,03125	0,046156	0,032258
1	0,03125	0,046156	0,032258
piala	0,03125	0,046156	0,032258
panglima	0,03125	0,046156	0,032258
tni	0,03125	0,046156	0,032258
sidang	0,03125	0,046156	0,032258
buka	0,03125	0,03125	0,047645
senat	0,03125	0,03125	0,047645
wisuda	0,03125	0,03125	0,047645
daring	0,03125	0,03125	0,047645
2020	0,03125	0,03125	0,047645
nasional	0,03125	0,03125	0,047645

2. Data *testing* (uji)

Di proses ini akan dilakukan pengujian pada data uji menggunakan data uji pada data proses *training* sebelumnya. Pada Tabel 3.10 merupakan contoh data yang menjadi data *testing* (uji).

Tabel 3.11 Contoh Data *Testing* (Uji)

Judul	Kategori
UIR Meraih Juara di Lomba Catur Nasional	?

Pada pengujian data diatas, selanjutnya akan dihitung probabilitas berdasarkan nilai probabilitas pada setiap *term*. Menghitung nilai probabilitas menggunakan metode *Naive bayes* pada data *testing* ini ialah dengan mengkalikan nilai probabilitas semua kategori dengan probabilitas pada setiap *term* yang diambil dari semua data *training*.

Sebelum melakukan proses perhitungan, data uji terlebih dahulu akan melakukan proses *text mining*. Tabel 3.11 merupakan hasil data *testing* yang telah melalui proses *preprocessing*.

Tabel 3.12 Hasil Data *Testing* (Uji)

Judul	Kategori
uir raih juara lomba catur nasional	?

Setelah itu akan dilakukan penghitungan probabilitas setiap kata di kategori pada data *testing* berdasarkan hasil dari probabilitas *term* yang didapat pada data *training*. Tabel 3.12 merupakan data *training* dan data *testing* setelah melalui tahap *preprocessing*.

Tabel 3.13 Data *Training* dan Data *Testing*

Data Latih (n)	Judul	Kategori
1	mahasiswa uir terpilih mengikuti program bangkit kampus merdeka	Prestasi
2	mahasiswa uir juara 1 nasional piala panglima tni	Olahraga
3	sidang terbuka senat uir wisuda daring 2020	Sosial
Data Uji		
4	uir meraih juara di lomba catur nasional	?

Selanjutnya akan dilakukan perhitungan probabilitas pada masing-masing kata kategori pada data uji dengan menggunakan rumus persamaan pada 2.2.6.

Diketahui:

$$|\text{kosakata}| = 22$$

$$\text{Term Prestasi} = 10$$

$$\text{Term Olahraga} = 10$$

$$\text{Term Sosial} = 9$$

$$P(\text{uir} | \text{Prestasi}) = (0,031+1) / (10+22) = 0.032$$

$$P(\text{raih} | \text{Prestasi}) = (0+1) / (10+22) = 0.031$$

$$P(\text{juara} | \text{Prestasi}) = (0,031+1) / (10+22) = 0.032$$

$$P(\text{lomba} | \text{Prestasi}) = (0+1) / (10+22) = 0.031$$

$$P(\text{catur} | \text{Prestasi}) = (0+1) / (10+22) = 0.031$$

$$P(\text{nasional} | \text{Prestasi}) = (0+1) / (10+22) = 0.031$$

$$P(\text{uir} | \text{Olahraga}) = (0,031+1) / (10+22) = 0.032$$

$$P(\text{raih} | \text{Olahraga}) = (0+1) / (10+22) = 0.031$$

$$P(\text{juara} | \text{Olahraga}) = (0,046+1) / (10+22) = 0.032$$

$$\begin{aligned}
 P(\text{lomba} \mid \text{Olahraga}) &= (0+1) / (10+22) = 0.031 \\
 P(\text{catur} \mid \text{Olahraga}) &= (0+1) / (10+22) = 0.031 \\
 P(\text{nasional} \mid \text{Olahraga}) &= (0,046+1) / (10+22) = 0.032 \\
 \\
 P(\text{uir} \mid \text{Sosial}) &= (0,032+1) / (9+22) = 0.033 \\
 P(\text{raih} \mid \text{Sosial}) &= (0+1) / (9+22) = 0.032 \\
 P(\text{juara} \mid \text{Sosial}) &= (0,032+1) / (9+22) = 0.033 \\
 P(\text{lomba} \mid \text{Sosial}) &= (0+1) / (9+22) = 0.032 \\
 P(\text{catur} \mid \text{Sosial}) &= (0+1) / (9+22) = 0.032 \\
 P(\text{nasional} \mid \text{Sosial}) &= (0+1) / (9+22) = 0.032
 \end{aligned}$$

Selanjutnya ialah melakukan penghitungan probabilitas kategori menggunakan rumus persamaan pada 2.5 :

$$P(\text{Kategori} \mid d) = P(\text{Kategori}) * \prod_i P(a_i \mid \text{Kategori})$$

$$\begin{aligned}
 P(\text{uji} \mid \text{prestasi}) &= P(\text{prestasi}) \times P(\text{uir} \mid \text{prestasi}) \times P(\text{raih} \mid \text{prestasi}) \times P(\text{juara} \mid \\
 &\text{prestasi}) \times P(\text{lomba} \mid \text{prestasi}) \times P(\text{catur} \mid \text{prestasi}) \times P(\text{nasional} \mid \\
 &\text{prestasi})
 \end{aligned}$$

$$\begin{aligned}
 P(\text{uji} \mid \text{prestasi}) &= 0.2 \times 0.032 \times 0.032 \times 0.032 \times 0.032 \times 0.032 \times 0.032 \\
 &= 0.0000000000205
 \end{aligned}$$

$$\begin{aligned}
 P(\text{uji} \mid \text{olahraga}) &= P(\text{olahraga}) \times P(\text{uir} \mid \text{olahraga}) \times P(\text{raih} \mid \text{olahraga}) \times P(\text{juara} \mid \\
 &\text{olahraga}) \times P(\text{lomba} \mid \text{olahraga}) \times P(\text{catur} \mid \text{olahraga}) \times P(\text{nasional} \\
 &\mid \text{olahraga})
 \end{aligned}$$

$$P(\text{uji} \mid \text{olahraga}) = 0.2 \times 0.031 \times 0.031 \times 0.032 \times 0.031 \times 0.031 \times 0.032$$

$$= 0.00000000000218$$

$$P(\text{uji}|\text{sosial}) = P(\text{sosial}) \times P(\text{uir} | \text{sosial}) \times P(\text{raih} | \text{sosial}) \times P(\text{juara} | \text{sosial}) \times \\ P(\text{lomba} | \text{sosial}) \times P(\text{catur} | \text{sosial}) \times P(\text{nasional} | \text{sosial})$$

$$P(\text{uji} | \text{sosial}) = 0.2 \times 0.033 \times 0.033 \times 0.033 \times 0.033 \times 0.033 \times 0.033 \\ = 0.00000000000265$$

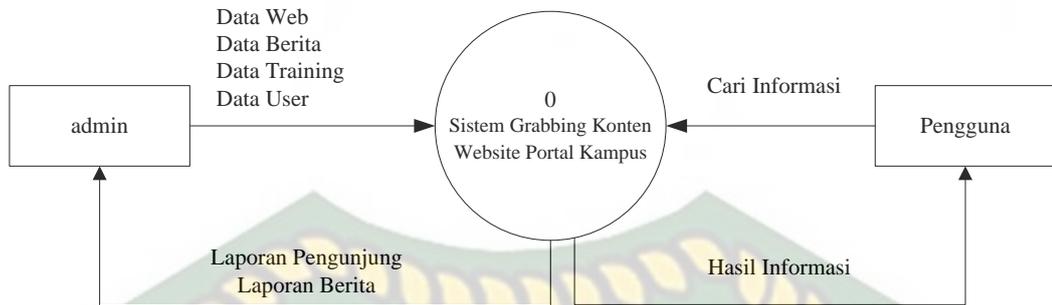
Hasil dari perhitungan nilai probabilitas di atas didapat nilai tertinggi pada $P(\text{uji}|\text{Sosial})$ yaitu sebesar 0.00000000000265, sehingga hasil dari artikel pada data uji tersebut diklasifikasikan dalam kategori Sosial.

3.4 Perancangan Sistem

Pada tahapan ini akan dijelaskan hal yang berhubungan dengan perancangan sistem yang akan dibuat:

3.4.1 Diagram Konteks

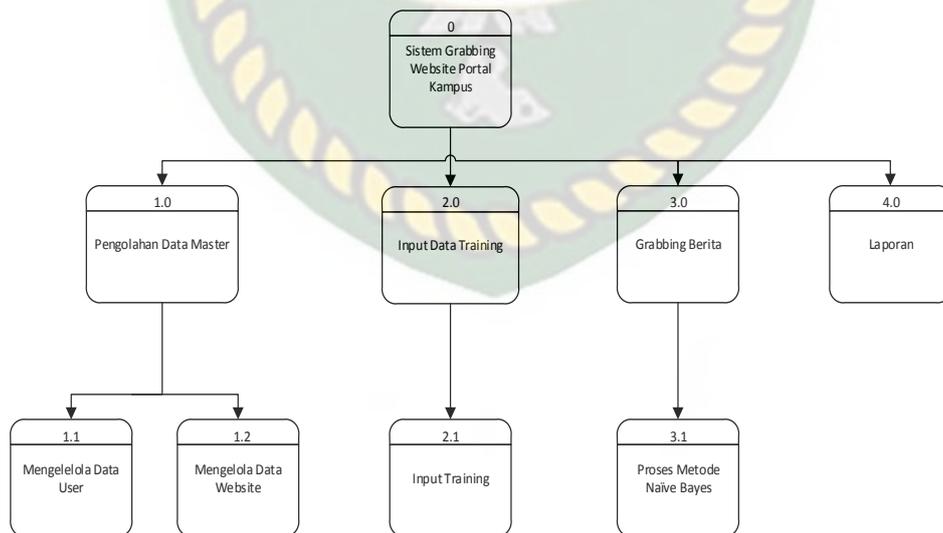
Diagram Konteks adalah sebuah diagram sederhana yang menggambarkan hubungan antara *entity* luar, masukan dan keluaran dari sistem. Diagram konteks direpresentasikan dengan lingkaran tunggal yang mewakili keseluruhan sistem. Salah satu fungsi dari diagram konteks adalah melakukan penyimpanan data eksternal lain.



Gambar 3.3 Diagram Konteks

3.4.2 Hierarchy Chart

Hierarchy chart merupakan gambaran subsistem yang menjelaskan proses-proses yang terdapat dalam sistem utama dimana semua subsistem yang berada dalam ruang lingkup sistem utama saling berhubungan satu dan lainnya yang membedakan adalah pada level prosesnya. Berikut ini adalah gambaran hierarchy chart pada sistem ini dapat dilihat pada gambar dibawah ini.



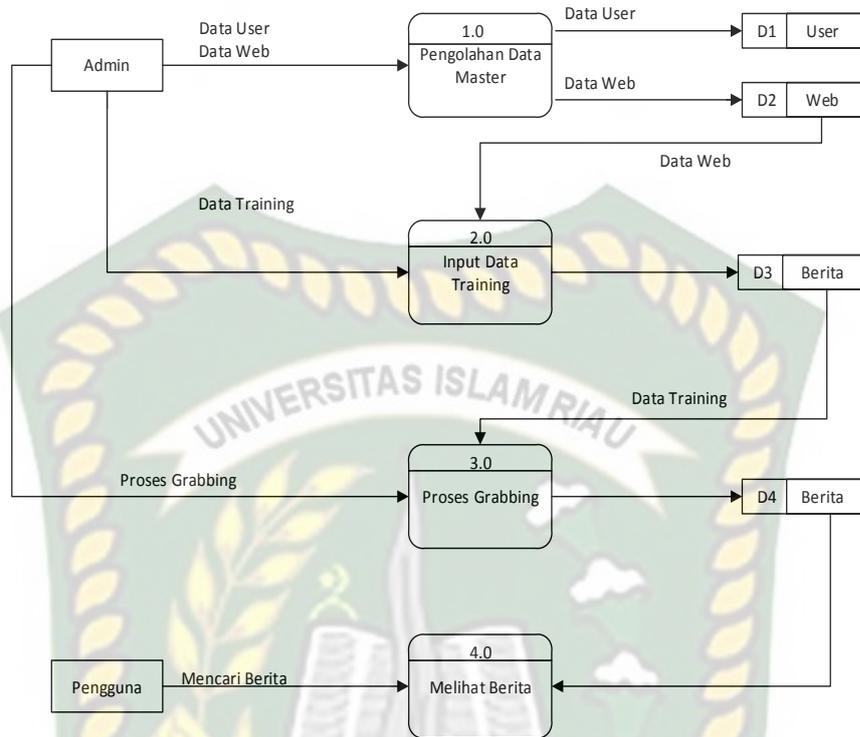
Gambar 3.4 Hierarchy Chart

3.4.3 Data Flow Diagram (DFD)

Data Flow Diagram adalah suatu diagram yang menggambarkan aliran data dari sebuah proses yang disebut dengan sistem informasi. Di *dalam data flow diagram* juga menyediakan informasi mengenai *input* dan *output* dari tiap entitas dan proses itu sendiri. DFD adalah representasi grafik dari sebuah sistem yang menggambarkan alur sistem. Pada sistem *grabbing* konten *website* ini terdapat beberapa level proses yaitu:

3.4.3.1 DFD Level 0

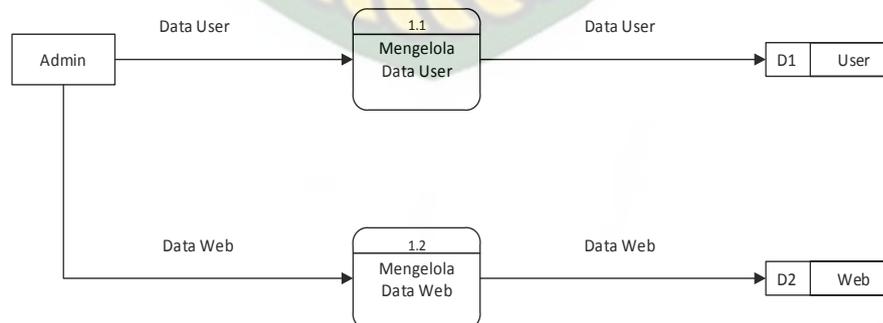
DFD level 0 bisa dilihat pada gambar 3.5 proses pengolahan data master bertugas mengelola data *user* dan data web. Data *user* dan data web yang di inputkan oleh admin kemudian disimpan pada data *store*. Sedangkan data *training* di inputkan oleh admin. Selanjutnya dari data *store* ke data *training* dan berita tersebut digunakan untuk proses metode *naive bayes*. Hasil proses tersebut merupakan sistem portal web dengan teknik *grabbing* dan menggunakan metode *naive bayes* untuk konten *website* portal kampus.



Gambar 3.5 DFD Level 0 Sistem portal web

3.4.3.2 DFD Level 1 Proses 1

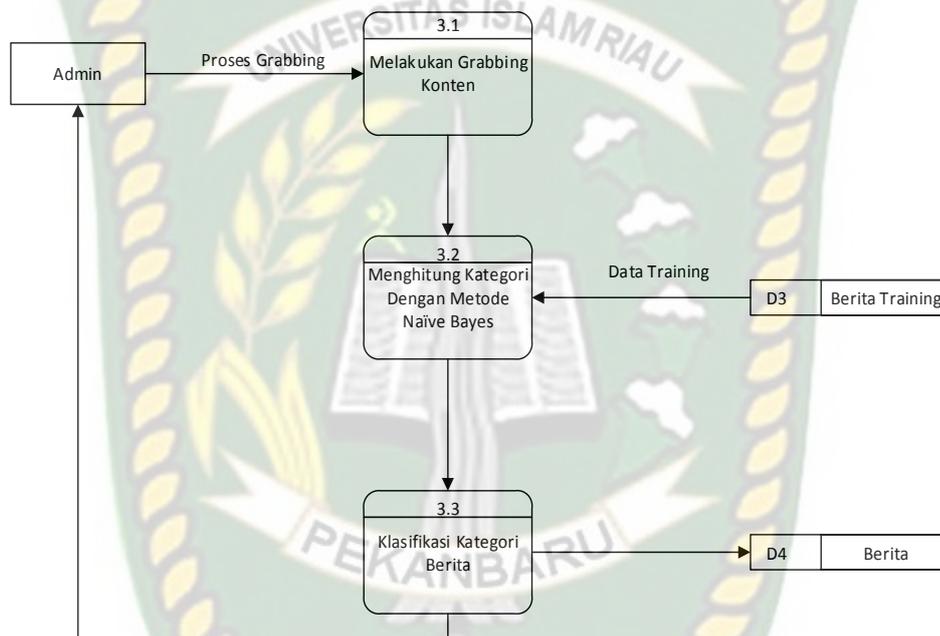
Pada proses pengelolaan data dibagi dalam 2 proses yaitu mengelola data *user*, dan data web yang akan dikelola oleh admin, dapat dilihat pada gambar 3.6.



Gambar 3.6 DFD Level 1 Proses 1 Pengolahan Data Master

3.4.3.3 DFD Level 1 Proses 3

Pada gambar dibawah ini dijelaskan bahwa tim penguji menginputkan data *training* berita. Pada proses *naive bayes* mengambil data *training* dari konten berita sebelumnya yang *diinputkan* oleh admin, dan kemudian dilakukan pengkategorian pada saat melakukan *grabbing* data *website* kampus.



Gambar 3.7 DFD Level 1 Proses 3 Proses penilaian akhir

3.4.4 Desain Output

Desain *output* dari sistem portal web dengan teknik *grabbing* dan menggunakan metode *naive bayes* untuk konten *website* portal kampus dapat dilihat pada gambar 3.8.

BERITA DAN INFORMASI

No	JUDUL	KAMPUS	TANGGAL	KATEGORI
9(2)	X(100)	X(100)	99-99-9999	X(10)
9(2)	X(100)	X(100)	99-99-9999	X(10)

Gambar 3.8 Desain Output Portal Berita Admin

Pada gambar 3.8 hasil dari proses *grabbing website* dan penggunaan *naive bayes* untuk proses klasifikasi kategori. Hasil akhir menampilkan judul, kampus, tanggal, kategori. Hasil akhir akan menjadi pilihan alternatif terbaik untuk admin dalam penilaian portal berita dengan klasifikasi *naive bayes*. Pada keterangan akan menampilkan data mengenai keterangan dari hasil yang didapatkan dari proses *grabbing*.

3.4.5 Desain Input

Desain *input* pada sistem portal web dengan teknik *grabbing* dan menggunakan metode *naive bayes* untuk konten *website* portal kampus ini terdiri dari:

1. Rekam Web

Fungsi : Mengelola Web

Nama tabel : web

Tombol : Menyimpan data web yang di *input*

Tombol Reset : Mengatur ulang *form* yang di *input* seperti semula

Tombol Edit : Mengedit data website yang sudah di *input*

Tombol Hapus : Menghapus data website

DATA WEBSITE

Form Web	List Web									
<p>Kampus</p> <input style="width: 100%;" type="text" value="X(50)"/> <p>URL</p> <input style="width: 100%;" type="text" value="X(100)"/> <p style="text-align: center;"> <input type="button" value="Save"/> <input type="button" value="Reset"/> </p>	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 33%;">Kampus</th> <th style="width: 33%;">URL</th> <th style="width: 34%;">Aksi</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">X(100)</td> <td style="text-align: center;">X(100)</td> <td style="text-align: center;">Edit Hapus</td> </tr> <tr> <td style="text-align: center;">X(100)</td> <td style="text-align: center;">X(100)</td> <td style="text-align: center;">Edit Hapus</td> </tr> </tbody> </table>	Kampus	URL	Aksi	X(100)	X(100)	Edit Hapus	X(100)	X(100)	Edit Hapus
Kampus	URL	Aksi								
X(100)	X(100)	Edit Hapus								
X(100)	X(100)	Edit Hapus								

Gambar 3.9 Rekam Website

2. Rekam *Training* Berita

Fungsi : Mengelola *Training* Berita

Nama tabel : Berita

Tombol Simpan : Menyimpan data berita yang di *input*

Tombol Reset : Mereset *form* yang sudah di *input* seperti semula

Tombol Edit : Mengedit data *training* yang sudah di *input*

Tombol Hapus : Menghapus data *Training* Berita

DATA TRAINING

Form Training

Judul

Tanggal Kategori Kampus

Isi

Link Gambar

Gambar 3.10 Rekam *Training* Berita

3. Rekam Data *User*

Fungsi : Mengelola data *User*

Nama tabel : *User*

Tombol Simpan : Menyimpan data pengguna yang di *input*

Tombol Reset : Mengatur ulang *form* yang di *input* seperti semula

Tombol Edit : Mengedit data pengguna yang sudah di *input*

Tombol Hapus : Menghapus data pengguna

DATA USER

Form User

Nama

Username

Password

List User

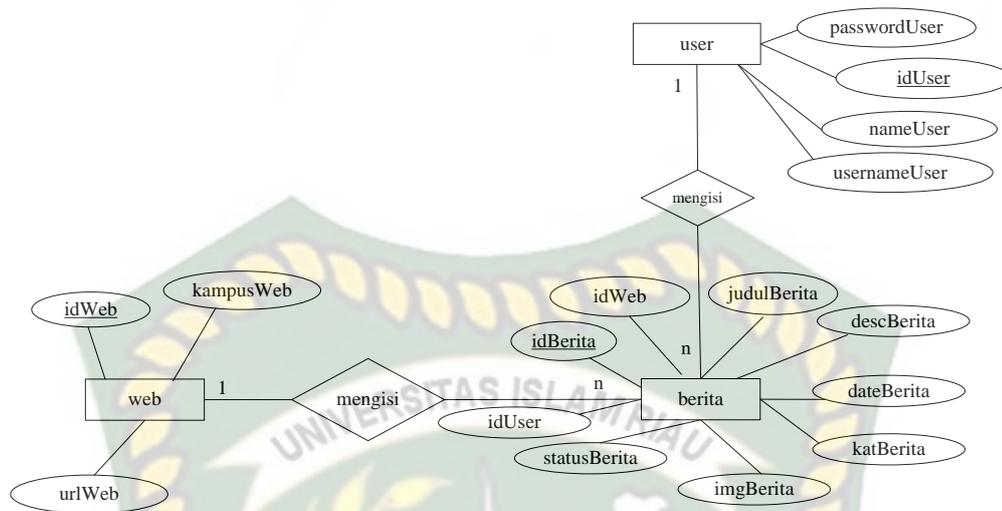
No	Nama	Username	Aksi	
999	X(50)	X(20)	Edit	Hapus
999	X(50)	X(20)	Edit	Hapus

Gambar 3.11 Rekam Data *User*

3.4.6 Perancangan *Database*

3.4.6.1 *Entity Relationship Diagram (ERD)*

ERD adalah diagram yang digunakan untuk perancangan suatu *database* dan menunjukkan relasi antar objek beserta atribut-atributnya pada *database*.



Gambar 3.12 ERD

3.4.6.2 Desain Database

Dalam perancangan basis data sistem portal web, penulis membuat beberapa buah tabel yang saling berelasi. Tabel-tabel tersebut terdiri dari tabel data *user*, tabel data berita, dan tabel lainnya.

1. Tabel *user*

Nama Database : *db_infokampus*

Nama Tabel : *user*

Tabel 3.14 Tabel *user*

No	Field	Data Type	Size	Ket
1	<i>idUser</i>	Int	10	<i>Primary Key</i>
2	<i>nameUser</i>	Varchar	50	
3	<i>usernameUser</i>	Varchar	20	
4	<i>passwordUser</i>	Varchar	50	

2. Tabel Web

Nama Database : *db_infokampus*

Nama Tabel : web

Tabel 3.15 Tabel web

No	Field	Data Type	Size	Ket
1	idWeb	Int	10	Primary Key
2	kampusWeb	Varchar	50	
3	urlWeb	Varchar	100	

3. Tabel Berita

Nama Database : *db_infokampus*

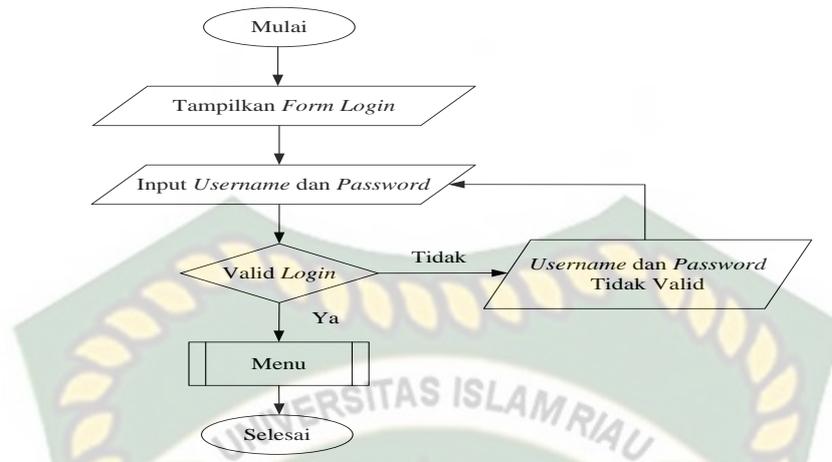
Nama Tabel : berita

Tabel 3.16 Tabel berita

No	Field	Data Type	Size	Ket
1	idBerita	Int	10	Primary Key
2	idWeb	Int	10	Foreign Key
3	idUser	Int	10	Foreign Key
4	judulBerita	Text		
5	descBerita	Text		
6	dateBerita	Date		
7	katBerita	Varchar	20	
8	imgBerita	Text		
9	statusBerita	Varchar	10	

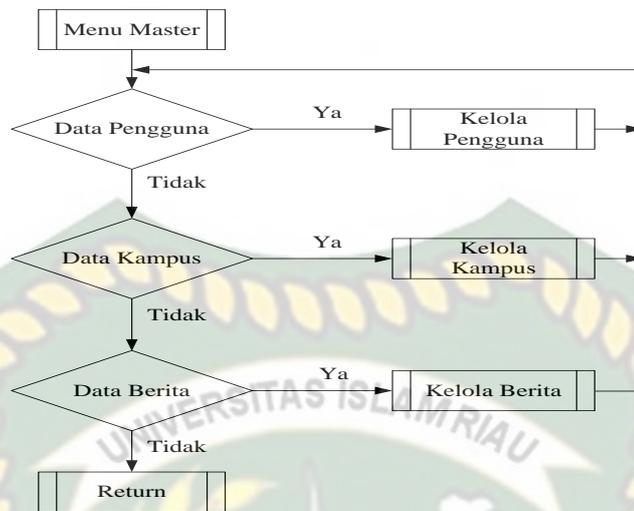
3.4.7 Desain Logika Program

Dalam membuat suatu sistem, pengembangan alur adalah hal yang sangat penting dalam memahami proses dari sebuah sistem. Pada tahapan ini akan dijabarkan alur proses pada sistem *grabbing* konten web dengan metode *naive bayes* dalam bentuk *Flowchart*. Berikut ini gambar 3.13 *Flowchart Login*.



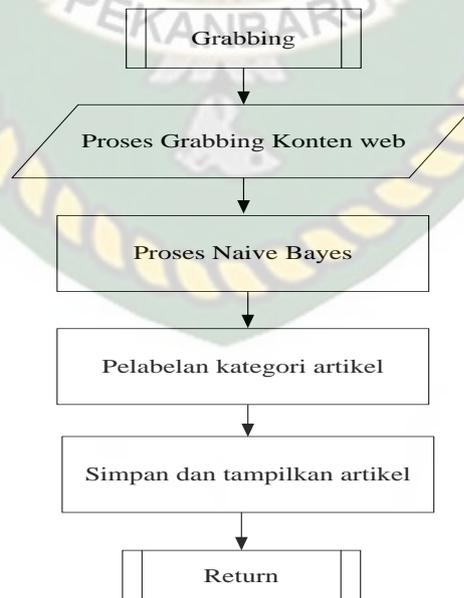
Gambar 3.13 Flowchart Login

Pada halaman utama akan muncul menu login, pengguna diharuskan untuk login terlebih dahulu dan menginputkan *username* dan *password*. Jika *password* benar maka pengguna diizinkan masuk ke sistem, dan jika pengguna salah menginputkan *username* dan *password*. Maka sistem akan memberi peringatan bahwasanya *username* dan *password* yang dimasukkan salah atau tidak valid. Ketika *login* berhasil maka akan diarahkan ke menu master untuk mengelola kampus, mengelola pengguna, dan mengelola data berita bisa dilihat pada gambar 3.14.



Gambar 3.14 Flowchart Menu Master

Flowchart proses sistem dapat dilihat pada gambar 3.15. Pada sistem admin melakukan *login*, admin dapat langsung menggunakan sistem dengan cara memilih menu yang tersedia. Dalam proses *grabbing* konten *website* kampus, proses *flowchart* akan dilakukan seperti gambar dibawah ini.



Gambar 3.15 Flowchart Proses

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Hasil penelitian akan membahas *Interface* dari seluruh aplikasi Sistem *Grabbing* Konten Artikel *Website* Kampus Di Riau Menggunakan Metode Naive Bayes Classifier Berbasis Web Mobile. Setelah melakukan penelitian dan membangun sebuah sistem, maka dilakukan pengujian terlebih dahulu. Pengujian ini berfungsi untuk mengetahui hasil yang diberikan Sistem *Grabbing* Konten Artikel *Website* Kampus Di Pekanbaru Menggunakan Metode *Naive Bayes Classifier*. Hal ini dilakukan agar sewaktu sistem sudah benar-benar dipublikasikan tidak terjadi kesalahan.

4.2 Pengujian *Black Box*

Black box testing merupakan suatu metode pengujian terhadap fungsionalitas atau kegunaan sebuah aplikasi. *Black box testing* adalah pengujian yang sepenuhnya dengan hanya menilai kebutuhan dan spesifikasi software. *Black box testing* cukup meninjau *input* dan *output* sistem *software* tersebut tanpa pengetahuan tentang internal programnya. Pengujian ini juga berguna untuk menentukan apakah aplikasi tersebut sudah berjalan dengan baik tanpa ada masalah. Jika ada masalah, maka perlu dilakukan pengecekan lebih lanjut dan perbaikan pada sistem. Berikut ini pengujian *black box* terhadap sistem ini.

4.2.1 Pengujian Tampilan Pengunjung

Tampilan data pengunjung dapat diakses melalui web dan portal *website* akan tampil dengan data berita dari proses *grabbing* dari *website* kampus.



Gambar 4.1 Pengujian Tampilan *Home* Pengunjung

Pada gambar 4.1 dijelaskan bahwa merupakan tampilan berita dari beberapa kampus di Pekanbaru. Jika salah satu berita di klik maka akan menampilkan *detail* berita dapat dilihat pada gambar 4.2.



Gambar 4.2 Tampilan *Detail* Berita

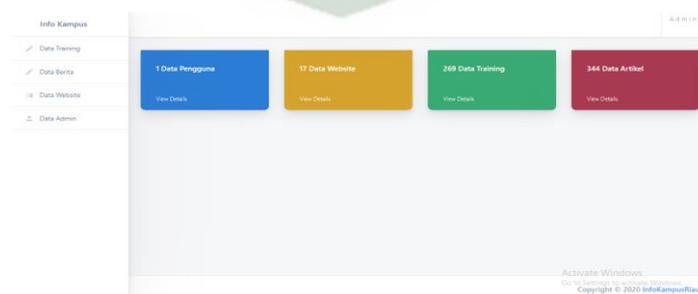
Pada gambar 4.2 menyatakan bahwa akan di alihkan ke *website* utama dari kampus sehingga tidak dikategorikan sebagai plagiarisme.

4.2.2 Pengujian Tampilan Admin

Admin dapat mengelola data *training*, data kampus, data berita dan data admin dengan masuk kedalam sistem dan sebelum masuk admin memasukan data *login* dengan memasukan data *username* dan *password*. berikut tampilan *form login*.

Gambar 4.3 Pengujian *Form login*

Setelah memasukan data, halaman utama akan ditampilkan. Tampilannya dapat dilihat pada gambar 4.4.



Gambar 4.4 Tampilan *Home*

Pada gambar 4.4 menyatakan bahwa admin dapat mengelola data *training*, data kampus, data berita dan data admin. Berikut ini adalah menu-menu secara *detail* pada halaman admin.

1. Menu *User*

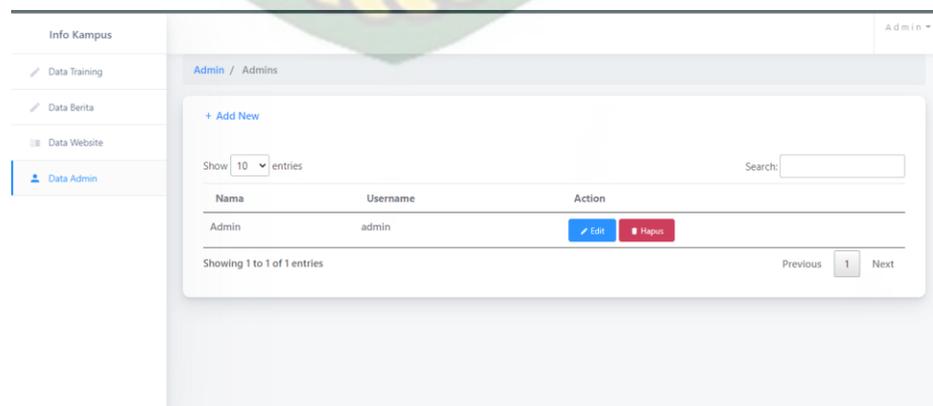
Pada menu *user* admin dapat mengelola data admin yang dapat masuk kedalam halaman admin, seperti menambah, menghapus, dan *update* data. berikut ini adalah *form* data *user* admin.



The screenshot shows a web application interface for an administrator. On the left is a sidebar menu with the following items: 'Info Kampus', 'Data Training', 'Data Berita', 'Data Website', and 'Data Admin' (which is currently selected). The main content area displays a form for adding or editing an admin user. The form has a 'Back' link at the top left. It contains three input fields: 'Nama*' with the value 'Admin', 'Username' with the value 'admin', and 'Password' which is masked with dots. Below the password field is a green 'Save' button. At the bottom left of the form, there is a small asterisk and the text '*Harus diisi'. In the top right corner of the main area, there is a dropdown menu labeled 'Admin'. At the bottom right of the page, there is a watermark that says 'Activate Windows'.

Gambar 4.5 Proses *Inputan* Data Admin

Jika data sudah terisi maka sistem akan menampilkan tabel data *user* seperti gambar 4.6 berikut ini.



The screenshot shows the same web application interface as in Gambar 4.5, but now displaying a table of admin users. The sidebar menu is identical. The main content area shows a breadcrumb 'Admin / Admins' and a '+ Add New' link. Below that is a search bar and a dropdown menu set to 'Show 10 entries'. The table has three columns: 'Nama', 'Username', and 'Action'. There is one row with 'Admin' in the 'Nama' column and 'admin' in the 'Username' column. The 'Action' column for this row contains two buttons: a blue 'Edit' button and a red 'Hapus' button. At the bottom left of the table, it says 'Showing 1 to 1 of 1 entries'. At the bottom right, there are 'Previous' and 'Next' navigation buttons, with '1' in a box between them. The 'Activate Windows' watermark is still present at the bottom right.

Nama	Username	Action
Admin	admin	Edit Hapus

Gambar 4.6 Tampilan Data Admin

2. Menu *Website*

Pada menu *user* admin dapat mengelola data kampus dan *websitenya* yang masuk kedalam portal *website* kampus, seperti menambah, menghapus, dan *update*. berikut ini adalah *form* data kampus.

Gambar 4.7 Proses *Inputan* Data Kampus

Jika data sudah terisi maka sistem akan menampilkan tabel data kampus seperti gambar 4.8 berikut ini.

Kampus	Url	Action
STMIK Hangtuah	http://www.stmikhtp.ac.id/search/label/Kegiatan%20Kampus	Edit Hapus
UIN SUSKA RIAU	https://uin-suska.ac.id	Edit Hapus
UIR	https://uir.ac.id	Edit Hapus
UMRI	https://umri.ac.id	Edit Hapus

Gambar 4.8 Tampilan Data Kampus

3. Menu *Training*

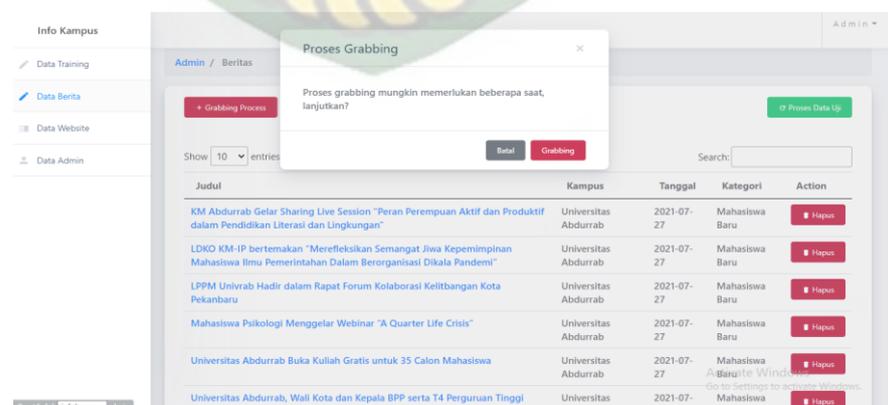
Menu data *training* merupakan menu khusus untuk memasukan data sample berita yang diambil dari portal berita yang bertujuan untuk proses *naive bayes*. Data sample dapat dilihat pada gambar berikut ini.



Gambar 4.9 Proses *Inputan Data Training*

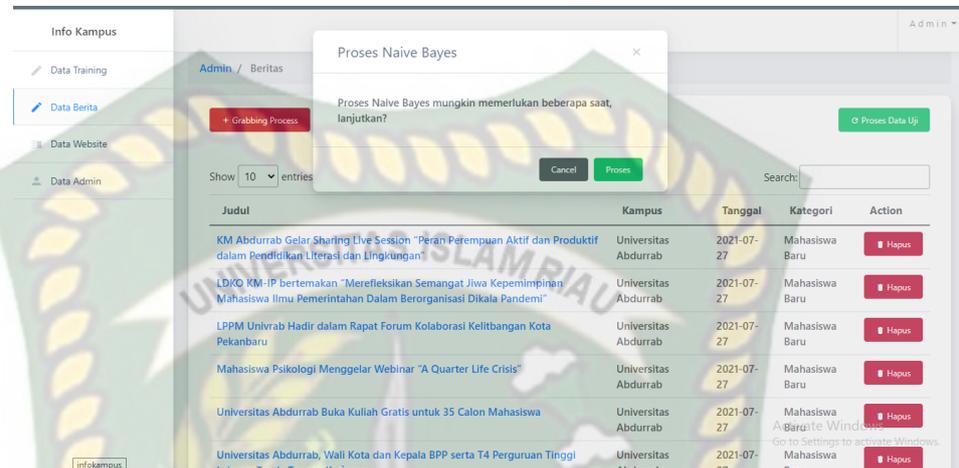
4. Menu Berita

Menu data berita merupakan menu khusus untuk memasukan data *grabbing* berita dari kampus dan pemrosesan *naive bayes* untuk mendapatkan kategori yang cocok terhadap berita. Proses kategori menggunakan metode *naive bayes*.



Gambar 4.10 Proses *Grabbing Data*

Setelah proses *grabbing* langkah selanjut proses pengkategorian menggunakan *naive bayes classifier* yang didasarkan pada data *training*.



Gambar 4.11 Proses *Naive bayes*

Dari proses pengujian diatas dapat disimpulkan bahwa setiap data yang diinputkan kedalam sistem berfungsi dengan baik, dan sistem akan memberi peringatan jika ada *field* yang kosong ataupun terdapat kesalahan dalam penginputan data yang tidak sesuai dengan format sistem yang dibuat. Maka akan muncul kolom berwarna merah sebagai peringatan bahwa *field* kosong atau terjadi kesalahan saat menginputkan data.

4.3 Implementasi Sistem

Implementasi sistem yang digunakan adalah dengan membuat kuesioner dengan 5 pertanyaan dan 10 responden umum yang terdiri dari pengguna dan masyarakat sebagai pengguna sistem. Setelah dilakukan pengujian terhadap aplikasi, maka didapatkan beberapa kritik dan saran. Data dari hasil pengujian pengguna dapat dilihat pada Tabel 4.1

Tabel 4.1 Hasil Pengujian Pengguna

Skenario	Penguji	Nilai	Saran	Kritik
Interface	Sri Ramahdani	A	UI nya bisa dikembangkan lagi	-
	Nur Fadhilah	A	Aplikasi dapat dikembangkan lagi	-
	Imam Syahrianto	A	Tampilan website dibuat lebih menarik lagi	-
	Siti Annisa	A	Aplikasi dikembangkan lagi agar cakupannya tidak sebatas pekanbaru saja	-
	Irwandi	A	Tampilan dibuat lebih menarik lagi	-
	Elsa Lutfi	A	Tampilannya mungkin bisa dibuat lebih bagus lagi	-
	Tama Dwi Anggara	A	Kategori beritanya kalau bisa ditambah lagi	-
	Abdussalam	B	Semoga dapat dikembangkan lagi	-
	Annisa	A	Lebih ditingkatkan lagi	-
	Ardhit Pratama	A	Menu dan tampilan dibuat lebih menarik	-

Implementasi sistem dilakukan dengan tujuan untuk mengetahui tanggapan dari pengguna tentang Sistem *Grabbing* Konten Artikel *Website* Kampus Di Riau Menggunakan Metode Naive Bayes Classifier Berbasis Web *Mobile*. Hasil implementasi dengan memberikan kuesioner kepada 10 orang skala *likert* adalah metode perhitungan yang digunakan untuk keperluan riset atas jawaban setuju atau tidaknya seorang responden terhadap suatu pernyataan terhadap suatu pernyataan. Untuk menghitung skor maksimum tiap jawaban

dengan mengalikan skor dengan jumlah responden, yaitu skor dikali 10 responden. Nilai skor maksimum dapat dilihat dari table 4.2

Tabel 4.2 Skor Maksimum

Jawaban	Skor	Skor Maksimum (Skor * Jumlah Responden)
Sangat Baik	4	75
Baik	3	45
Cukup	2	15
Kurang Baik	1	15

Setelah itu, dapat dicari persentase masing-masing jawaban dengan menggunakan rumus:

$$Y = \frac{TS}{\text{Skor Ideal}} \times 100\%$$

Dimana:

Y = Nilai persentase

TS = Total skor responden = \sum skor x responden

Skor ideal = skor x jumlah responden = 4 x 10 = 40

Kriteria skor untuk persentase dapat dilihat pada Tabel 4.3

Tabel 4.3 Kriteria Skor

Kategori	Keterangan
76%-100%	Sangat baik
51%-75%	Baik
26%-50%	Cukup
0%-25%	Tidak baik

Berikut ini adalah hasil persentase masing-masing jawaban yang sudah dihitung nilainya. Kuesioner ini telah diujikan kepada 10 orang responden.

1. Pertanyaan pertama

Apakah fitur dan informasi yang ditampilkan mudah dipahami?

Hasil kuesioner pertanyaan pertama dapat dilihat pada Tabel 4.4

Tabel 4.4 Hasil Kuesioner Pertanyaan Pertama

Pertanyaan	Jawaban	Skor	Responden	Jumlah Skor	Nilai Presentase (%)
1	Sangat baik	4	6	24	$(36:40) \times 100 = 90\%$
	Baik	3	4	12	
	Cukup	2	0	0	
	Tidak baik	1	0	0	
	Jumlah		10	36	

Berdasarkan nilai persentase dari pertanyaan pertama, dapat disimpulkan sebanyak % responden menyatakan bahwa informasi yang diberikan jelas.

2. Pertanyaan Kedua

Apakah aplikasi mudah digunakan ?

Hasil kuesioner pertanyaan kedua dapat dilihat pada Tabel 4.5

Tabel 4.5 Hasil Kuesioner Pertanyaan Kedua

Pertanyaan	Jawaban	Skor	Responden	Jumlah Skor	Nilai Presentase (%)
2	Sangat baik	4	7	28	$(37:40) \times 100 = 92,5\%$
	Baik	3	3	9	
	Cukup	2	0	0	
	Tidak baik	1	0	0	
	Jumlah		10	37	

3. Pertanyaan Ketiga

Bagaimana pendapat anda tentang tampilan aplikasi Sistem *Grabbing* Konten Artikel *Website* Di Riau Menggunakan Metode *Naive Bayes Classifier* Berbasis Web Mobile ?

Hasil kuesioner pertanyaan ketiga dapat dilihat pada Tabel 4.6

Tabel 4.6 Hasil Kuesioner Pertanyaan Ketiga

Pertanyaan	Jawaban	Skor	Responden	Jumlah Skor	Nilai Presentase (%)
3	Sangat baik	4	5	20	$(35:40) \times 100 = 87,5\%$
	Baik	3	5	15	
	Cukup	2	0	0	
	Tidak baik	1	0	0	
	Jumlah		10	35	

4. Pertanyaan Keempat

Apakah tampilan dan fungsi aplikasi mudah dipahami ?

Hasil kuesioner pertanyaan keempat dapat dilihat pada Tabel 4.7

Tabel 4.7 Hasil Kuesioner Pertanyaan Keempat

Pertanyaan	Jawaban	Skor	Responden	Jumlah Skor	Nilai Presentase (%)
4	Sangat baik	4	8	32	$(38:40) \times 100 = 95\%$
	Baik	3	2	6	
	Cukup	2	0	0	
	Tidak baik	1	0	0	
	Jumlah		10	38	

5. Pertanyaan Kelima

Apakah aplikasi ini bermanfaat bagi pengguna nantinya ?

Hasil kuesioner pertanyaan kelima dapat dilihat pada Tabel 4.8

Tabel 4.8 Hasil Kuesioner Pertanyaan Kelima

Pertanyaan	Jawaban	Skor	Responden	Jumlah Skor	Nilai Presentase (%)
5	Sangat baik	4	6	24	$(36:40) \times 100 = 90\%$
	Baik	3	4	12	
	Cukup	2	0	0	
	Tidak baik	1	0	0	
	Jumlah		10	36	

Berdasarkan persentase dari pertanyaan kelima diatas, maka dapat disimpulkan sebanyak 90% responden menyatakan bahwa aplikasi ini bermanfaat bagi pengguna.

Hasil dari setiap pertanyaan dilakukan perhitungan rata-rata secara keseluruhan. Kemudian akan dibandingkan dengan Tabel 4.8 Untuk diambil kesimpulan. Perhitungan secara keseluruhan pengolahan kuesioner dapat dilihat pada Tabel 4.9

Tabel 4.9 Pengolahan Skala

No Pertanyaan	Nilai Persentase	Keterangan
1	90%	Sangat baik
2	92%	Sangat baik
3	87,5%	Sangat baik
4	95%	Sangat baik
5	90%	Sangat baik
Total Persentase	454,5%	Sangat baik
Rata-rata	90.9%	

Dilihat dari tabel 4.9 adalah hasil penghitungan secara keseluruhan pengolahan kuesioner dengan jumlah pertanyaan adalah 5, hasil total presentase yang didapat adalah 454,5% dan rata-rata nilai persentase adalah 90,9%. Dapat disimpulkan bahwa hasil dari pengolahan skala untuk kuesioner adalah Sangat Baik.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisa data dan perancangan sistem portal *website* kampus di Pekanbaru ini dapat disimpulkan yaitu:

1. Sistem ini dapat menampilkan hasil *grabbing* konten *website* kampus dengan baik.
2. Sistem ini dapat mempermudah pengguna karena dilengkapi dengan informasi dari kampus-kampus di Pekanbaru.
3. Proses *naive bayes classifier* dapat digunakan untuk proses pelabelan kategori artikel.

5.2 Saran

Adapun saran yang dapat diberikan untuk penelitian berikutnya yang berkaitan dengan penelitian ini adalah :

1. Diharapkan untuk penelitian berikutnya yang berkaitan dengan penelitian ini menggunakan metode dan teknik yang berbeda agar sistem ini bisa dibandingkan dan dapat melihat perbedaan hasil dari setiap metode yang digunakan.
2. Semoga sistem ini dapat dikembangkan menjadi lebih baik lagi agar dapat digunakan oleh berbagai *device* atau *multiplatform*.

DAFTAR PUSTAKA

- Arikunto, S. (2009). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta : Rineka Cipta.
- Beck, K., & Andres, C. (2004). *Extreme Programming Explained: Embrace Change (2nd Edition)*. Addison-Wesley Professional.
- Bumi, S. (2003). *Definisi dari Portal WEB*. Retrieved from <http://www.websiteflashmedia.com/2013/01/definisi-dari-website-portal.html>
- Dauni, P. (2017). *Implementasi Metode Grabbing Pada Web Penyedia Informasi Beasiswa*.
- Firman, A. (2016). *Sistem Informasi Perpustakaan Online Berbasis Web*.
- Griffiths, A. (2010). *CodeIgniter 1.7 Professional Development*. Birmingham: Packt Publishing.
- Hamzah, A. (2012). *Klasifikasi Teks Dengan Naive Bayes Classifier(Nbc) Untuk Pengelompokan Teks Berita Dan Abstract Akademis*.
- Hidayat, R. (2010). *Cara Praktis Membangun Website Gratis*. Jakarta : PT Elex Media Komputindo Kompas: Gramedia.
- Kurniawan, B., Effendi, S., & Sitompu, O. S. (2012). *Klasifikasi Konten Berita dengan Metode Text Mining*. *Jurnal Dunia Teknologi Informasi 1*, 14-19.
- Kustiyahningsih. (2011). *Pemrograman Basis Data Berbasis Web Menggunakan PHP dan MySQL*. Yogyakarta: Graha Ilmu.
- Liawati, S. (2015). *Minat Memilih Program Studi Pendidikan Akuntansi Ditinjau Dari Minat Menjadi Guru Dan Persepsi Peluang Kerja*.
- Mahmudy, W., & Widodo, A. (2014). *Klasifikasi Artikel Berita Secara Otomatis Menggunakan Metode Naive Bayes Classifier yang Dimodifikasi*.

- McCallum, A., & Nigam, K. (1998). A comparison of event models for Naive Bayes text classification.
- Muslim, M., & Retno, N. (2014). Implementasi Cloud Computing Menggunakan Metode Pengembangan Sistem Agile.
- Prabowo, A., Sholiq, S., & Muqtadiroh, F. (2013). Fakultas Teknologi Informasi Institut Teknologi Sepuluh Nopember. *Rancang Bangun Aplikasi Web Informasi Eksekutif Pada Pemerintah Kabupaten XYZ*.
- Purwoko, H. (2016). Pemanfaatan cURL pada PHP Guna Mendapatkan Informasi Malware dengan Memanfaatkan AVG Thread Labs.
- Rahmi, R., Sari, R. P., & Suhatman, R. (2016, November). *Jurnal Komputer Terapan Vol.2, No.2*. Retrieved from Jurnal Politeknik Caltex Riau: <http://jurnal.pcr.ac.id>
- Rish, I. (2001). An Empirical Study of the *Naive Bayes* Classifier.
- Simarmata, J. (2010). *Rekayasa Perangkat Lunak*. Yogyakarta: Andi Offset.
- Syihan Husin, A. (2013). Sistem Terintegrasi Portal Web Perguruan Tinggi Agama Islam Negeri Menggunakan Metode Grabbing.
- Zabir, O. (2008). *Building a Web 2.0 Portal with ASP.NET 3.5*. O'REILLY.