**ANALISIS SENTIMEN KOMENTAR MASYARAKAT PADA APLIKASI PRO DENPASAR MENGGUNAKAN METODE *BOOLEAN MULTINOMIAL NAÏVE BAYES***

**SKRIPSI**



**COKORDA GEDE AGUNG YUDI DHARMA PUTRA**

**NIM. 1408605015**

**JURUSAN ILMU KOMPUTER**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS UDAYANA**

**BUKIT JIMBARAN**

**2018**

# **SURAT PERNYATAAN KEASLIAN KARYA ILMIAH**

Yang bertanda tangan dibawah ini menyatakan bahwa naskah Skripsi dengan judul:

Nama : Cokorda Gede Agung Yudi Dharma Putra

NIM : 1408605015

Program Studi : Teknik Informatika

E-mail : [cokagungyudi@gmail.com](mailto:cokagungyudi@gmail.com)

Nomor Telp. : 081238738893

Alamat : Banjar Kawan Mas Ubud, Gianyar

Belum pernah dipublikasikan dalam dokumen skripsi, jurnal nasional maupun internasional atau dalam prosiding manapun, dan tidak sedang atau akan diajukan untuk publikasi di jurnal atau prosiding manapun. Apabila di kemudian hari terbukti terdapat pelanggaran kaidah-kaidah akademik pada karya ilmiah saya, maka saya bersedia menanggung sanksi-sanksi yang dijatuhkan karena kesalahan tersebut, sebagaimana diatur oleh Peraturan Menteri Pendidikan Nasional Nomor 17 Tahun 2010 tentang Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi.

Demikian Surat Pernyataan ini saya buat dengan sesungguhnya untuk dapat dipergunakan bilamana diperlukan.

Bukit Jimbaran, Mei 2018

Yang membuat pernyataan

(Cokorda Gede Agung Yudi Dharma Putra)

NIM. 1408605015

|  |  |  |
| --- | --- | --- |
| Judul | : | Analisis Sentimen Komentar Masyarakat Pada Aplikasi Pro Denpasar Menggunakan Metode *Boolean Multinomial Naïve Bayes* |
| Kompetensi | : | Komputasi |
| Nama | : | Cokorda Gede Agung Yudi Dharma Putra |
| NIM | : | 1408605015 |
| Tanggal Seminar | : |  |

# **LEMBAR PENGESAHAN TUGAS AKHIR**

|  |  |
| --- | --- |
| Pembimbing I | Penguji I |
|  |  |
| **(I Putu Gede Hendra Suputra, S.Kom., M.Kom)** |  |
| NIP. 198812282014041001 | NIP. |

|  |  |
| --- | --- |
| Pembimbing II | Penguji II |
|  |  |
| **(Ida Bagus Gede Dwidasmara, S.Kom., M.Cs)** |  |
| NIP. 198503152010121007 | NIP. |

|  |
| --- |
| Penguji III |
|  |
|  |
| NIP. |

|  |
| --- |
| Mengetahui,  Ketua Jurusan Ilmu Komputer  FMIPA Universitas Udayana |
|  |
| **(Dr. I Ketut Gede Suhartana, S.Kom., M.Kom)** |
| NIP. 197201102008121001 |

|  |  |  |
| --- | --- | --- |
| Judul | : | Analisis Sentimen Komentar Masyarakat Pada Aplikasi Pro Denpasar Menggunakan Metode *Boolean Multinomial Naïve Bayes* |
| Nama | : | Cokorda Gede Agung Yudi Dharma Putra (NIM: 1408605015) |
| Pembimbing | : | 1. I Putu Gede Hendra Suputra, S.Kom., M.Kom 2. Ida Bagus Gede Dwidasmara, S.Kom., M.Cs |

# **ABSTRAK**

|  |  |  |
| --- | --- | --- |
| Title | : | *Sentiment Analysis Of Community Commentary On Pro Denpasar Applications Using Boolean Multinomial Naïve Bayes Methods* |
| Name | : | Cokorda Gede Agung Yudi Dharma Putra  (Student Number: 1408605015) |
| Supervisor | : | I Putu Gede Hendra Suputra, S.Kom., M.Kom |
| Co-Supervisor | : | Ida Bagus Gede Dwidasmara, S.Kom., M.Cs |

# **ABSTRACT**

# **KATA PENGANTAR**

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas rahmat dan tuntunanNya penulis dapat menyelesaikan laporan proposal penelitian yang berjudul "Analisis Sentimen Kometar Masyarakat pada Aplikasi PRO Denpasar Menggunakan Metode *Boolean Multinomial Naïve Bayes*” secara tepat waktu. Proposal ini disusun dengan harapan dapat menjadi pedoman dan arahan dalam melaksanakan penelitian.

Selama menyelesaikan proposal ini, penulis mengucapkan banyak terima kasih kepada berbagai pihak telah banyak memberi bimbingan, pengarahan, petunjuk dan saran, serta fasilitas yang membantu hingga akhir penulisan proposal ini.

Disadari pula bahwa sudah tentu proposal ini masih mengandung kelemahan dan kekurangan. Masukan dan saran penyempurnaan sangan diharapkan.

Bukit Jimbaran, Desember 2017

Penyusun

# **DAFTAR ISI**

[**SURAT PERNYATAAN KEASLIAN KARYA ILMIAH** ii](#_Toc507617286)

[**LEMBAR PENGESAHAN TUGAS AKHIR** iii](#_Toc507617287)

[**ABSTRAK** iv](#_Toc507617288)

[**ABSTRACT** ii](#_Toc507617289)

[**KATA PENGANTAR** iii](#_Toc507617290)

[**DAFTAR ISI** iv](#_Toc507617291)

[**DAFTAR TABEL** vi](#_Toc507617292)

[**DAFTAR GAMBAR** vii](#_Toc507617293)

[**BAB I** 1](#_Toc507617294)

[**PENDAHUULUAN** 1](#_Toc507617295)

[**1.1** **Latar Belakang** 1](#_Toc507617296)

[**1.2** **Rumusan Masalah** 3](#_Toc507617297)

[**1.3** **Tujuan Penelitian** 3](#_Toc507617298)

[**1.4** **Batasan Masalah** 3](#_Toc507617299)

[**1.5** **Manfaat Penelitian** 4](#_Toc507617300)

[**BAB II** 5](#_Toc507617301)

[**TINJAUAN PUSTAKA** 5](#_Toc507617302)

[**2.1** **Tinjauan Studi** 5](#_Toc507617303)

[**2.2** **Aplikasi PRO Denpasar** 6](#_Toc507617304)

[**2.3** **Analisis Sentimen** 6](#_Toc507617305)

[**2.4** ***Text Preprocessing*** 7](#_Toc507617306)

[**2.5** **Seleksi Fitur *Chi-square*** 7](#_Toc507617307)

[**2.6** ***Boolean Multinomial Naïve Bayes*** 8](#_Toc507617308)

[**BAB III** 11](#_Toc507617309)

[**ANALISIS DAN PERANCANGAN** 11](#_Toc507617310)

[**3.1** ***Fishbone* Penelitian** 11](#_Toc507617311)

[**3.2** **Alur Pengolahan Data** 11](#_Toc507617312)

[**3.3** **Pengumpulan Data** 12](#_Toc507617313)

[**3.4** ***Preproccessing*** 13](#_Toc507617314)

[**3.4.1** ***ToLowerCase*** 13](#_Toc507617315)

[**3.4.2** **Pengembalian kata** 13](#_Toc507617316)

[**3.4.3** ***Tokenizing*** 13](#_Toc507617317)

[**3.4.4** ***Stopword*** 14](#_Toc507617318)

[**3.4.5** ***Stemming*** 15](#_Toc507617319)

[**3.5** **Metode yang digunakan** 15](#_Toc507617320)

[**3.5.1** **Seleksi Fitur *Chi-square*** 16](#_Toc507617321)

[**3.5.2** ***Boolean Multinomial Naïve Bayes*** 17](#_Toc507617322)

[**3.6** **Implementasi** 19](#_Toc507617323)

[**3.7** **Pengujian dan Evaluasi** 19](#_Toc507617324)

[**3.7.1** **Skenario Pengujian** 19](#_Toc507617325)

[**3.7.2** **Evaluasi Sistem** 19](#_Toc507617326)

[**DAFTAR PUSTAKA** 21](#_Toc507617327)

# **DAFTAR TABEL**

[Tabel 1 Contoh Komentar 12](#_Toc503129804)

[Tabel 2 Contoh ToLowerCase 13](#_Toc503129805)

[Tabel 3 Contoh Pengembalian Kata 13](#_Toc503129806)

[Tabel 4 Contoh Tokenizing 14](#_Toc503129807)

[Tabel 5 Contoh proses stopword 14](#_Toc503129808)

[Tabel 6 Contoh Stemming 15](#_Toc503129809)

[Tabel 7 Contoh Tabel Pengetahuan 17](#_Toc503129810)

[Tabel 8 Contoh tabel evaluasi Chi-square 20](#_Toc503129811)

[Tabel 9 Jadwal Pelaksanaan Penelitian 20](#_Toc503129812)

# **DAFTAR GAMBAR**

[Gambar 1 Fishbone Penelitian 11](#_Toc503129790)

[Gambar 2 Alur Metodologi Penelitian 12](#_Toc503129791)

[Gambar 3 Flowchart langkah Chi-square 16](#_Toc503129792)

[Gambar 4 Flowchart langkah metode Boolean Multinomial Naïve Bayes 18](#_Toc503129793)

# **BAB I**

# **PENDAHUULUAN**

## **Latar Belakang**

Pengaduan Rakyat Online Kota Denpasar (PRO Denpasar) merupakan aplikasi berbasis website dan mobile yang menjembatani partisipasi masyarakat dalam melakukan monitoring dan evaluasi pembangunan daerah yang dilakukan oleh Pemerintah Kota Denpasar dalam upaya menjadi Kota Denpasar menjadi *smart city*. Masyarakat umum dapat melakukan pengaduan dan pelaporan terkait pembangunan dan akan ditinjau oleh tim verifikasi PRO Denpasar sebelum dilanjutkan ke instansi terkait untuk ditindak lanjuti. Menurut Dinas Komunikasi dan Informatika Kota Denpasar (2013), aplikasi PRO Denpasar telah terhubung dengan 52 instansi yang terdiri dari Satuan Kerja Perangkat Daerah (SKPD) Pemerintah Kota Denpasar, Perusahaan Daerah dan Kelurahan – kelurahan di Kota Denpasar. Sejak diluncurkan pada tahun 2013, tentunya sudah banyak masyarakat yang sudah menggunakan aplikasi PRO Denpasar untuk melakukan pengaduan, selain itu juga mengomentari kinerja Pemerintah dalam menanggapi pengaduan yang diterima. Komentar yang diterima dapat berupa komentar yang bersifat positif, negatif, dan netral, dimana komentar tersebut dapat dijadikan evaluasi oleh Pemerintah dalam memperbaiki kinerja dalam menanggapi pengaduan. Maka dari itu dibutuhkan suatu sistem yang dapat menganalisis dan melaporkan jenis komentar-komentar yang diterima Pemerintah dalam menanggapi pengaduan masyarakat, apakah komentar tersebut positif, negatif, atau netral.

Analisis sentimen merupakan studi komputasi yang menganalisa opini-opini, sentimen, emosi yang diekspresikan dalam teks (Liu, 2012). Analisis sentimen memiliki tugas untuk mengklasifikasi suatu komentar ke dalam kelas positif, negatif, maupun komentar yang bersifat netral.

Sebelumnya penelitian analisis sentimen sudah banyak dilakukan dengan berbagai macam metode. Metode yang banyak digunakan dalam penelitian analisis sentimen seperti SVM (*Support Vector Machine*), *Maximum Entropy,* dan *Naïve Bayes.* Untuk meningkatkan akurasi ketepatan dalam mengklasifikasi biasanya metode-metode tersebut dikombinasikan dengan metode untuk menyeleksi fitur contohnya seperti metode *Mutual Information, frequency-based, dan Chi-square.* Pada penelitian yang dilakukan Anmol Nayak dan Dr. S Natarajan (2016), Mereka melakukan analisis sentimen terhadap media jejaring sosial Twitter dengan menggunakan beberapa teknik klasifikasi. Metode yang digunakan adalah *Naïve Bayes*, *Support Vector Machine,* dan *Random Forest Classifier*. Hasil yang diperoleh dari penelitian mereka dapat ditarik kesimpulan bahwa *Naïve Bayes* memperlihatkan performa yang lebih baik dari ketiga teknik yang digunakan. Sehingga pada penelitian ini digunakan metode *Naïve Bayes* untuk mengklasifikasikan komentar. Seleksi fitur yang digunakan pada penelitian ini adalah seleksi fitur *Chi-square* karena pada penelitian yang dilakukan Juen Ling et al. (2014) yang meneliti analisis sentimen dengan membandingkan seleksi fitur *Chi-square* dengan *frequence-based*, dimana pada kesimpulan disebutkan seleksi *Chi-square* lebihbaik digunakan daripada *frequence-based.*

Metode *Naïve Bayes* memiliki beberapa varian tergantung kasus yang akan diselesaikan. Untuk kasus klasifikasi teks terdapat dua model dari *Naïve Bayes* yaitu Model *Multinomial Naïve Bayes* dan *Boolean Multinomial Naïve Bayes.* Menurut Dan Jurafsky (2017), model *Multinomial Naïve Bayes* lebih cocok digunakan untuk pencarian dokumen karena pada model *Multinomial Naïve Bayes* memperhatikan *frequence*-*term* dari setiap kata yang ada pada dokumen, sedangkan untuk model *Boolean Multinomial Naïve Bayes* lebih cocok untuk menangani kasus analisis sentimen karena pada model *Boolean Multinomial Naïve Bayes, frequence-term* kata hanya akan dihitung satu disetiap dokumen. Maka dari itu pada penelitian ini digunakan Model *Boolean Multinomial Naïve Bayes.* Selain itu Model *Boolean Multinomial Naïve Bayes* ini juga belum pernah digunakan untuk menganalisis sentimen dari komentar masyarakat terhadap kinerja pemerintah.

Berdasarkan permasalahan yang ada dan jurnal-jurnal yang menjadi dasar penelitian, maka dirancang sistem yang dapat melakukan analisis sentimen pada komentar masyarakat terhadap kinerja pemerintah menggunakan metode *Boolean Multinomial Naïve Bayes* dan diharapkan sistem dapat mengklasifikasi terhadap komentar positif, negatif, dan netral dengan akurasi ketepatan yang tinggi.

## **Rumusan Masalah**

Rumusan masalah dari latar belakang yang telah disampaikan adalah sebagai berikut:

1. Bagaimana merancang dan mengimplementasikan sistem analisis sentimen komentar dengan metode *Boolean Multinomial Naïve Bayes*.
2. Berapa tingkat akurasi dari metode *Boolean Multinomial Naïve Bayes* dalam menganalisis sentimen komentar.
3. Berapa tingkat akurasi sistem dengan menambahkan seleksi fitur *Chi-square* dalam analisis sentimen komentar.

## **Tujuan Penelitian**

Beberapa tujuan dari dilakukannya penelitian ini adalah sebagai berikut:

1. Untuk merancang dan membangun sistem analisis sentimen komentar dengan metode *Boolean Multinomial Naïve Bayes.*
2. Untuk mengetahui tingkat akurasi yang diperoleh dari metode *Boolean Multinomial Naïve Bayes* dalam menganalisis sentimen komentar.
3. Untuk mengetahui tingkat akurasi yang diperoleh sistem dengan menambahkan seleksi fitur *Chi-square* dalam analisis sentimen komentar.
4. Untuk membantu Dinas Kominfo Kota Denpasar dalam mengembangkan aplikasi Pro Denpasar dan melakukan evaluasi dan monitoring saat menanggapi pengaduan masyarakat.

## **Batasan Masalah**

Beberapa batasan masalah dari penelitian ini adalah sebagai berikut:

1. Pada penelitian ini dataset yang digunakan adalah data komentar dari tanggapan pengaduan pada aplikasi PRO Denpasar.
2. Kelas yang dihasilkan oleh analisis sentimen adalah komentar yang positif, negatif, dan netral.
3. Sistem tidak dapat mengenali komentar yang bersifat sarkasme.
4. *Stopword* hanya berlaku pada kata-kata yang berbahasa indonesia saja.
5. Bahasa Bali yang digunakan dalam kamus adalah Bahasa Bali sehari-hari.

## **Manfaat Penelitian**

Beberapa manfaat dari penelitian ini adalah sebagai berikut:

1. Diharapkan dengan adanya penelitian ini hasilnya dapat digunakan untuk dijadikan pedoman dalam melakukan atau mengembangkan penelitian lebih lanjut mengenai analisis sentimen.
2. Hasil penelitian ini dapat dijadikan acuan untuk monitoring dan evaluasi terhadap kinerja pemerintah dalam menanggapi pengaduan yang diterima dari masyarakat.

# **BAB II**

# **TINJAUAN PUSTAKA**

## **Tinjauan Studi**

Penelitian ini menggunakan beberapa jurnal-jurnal dari penelitian sebelumnya mengenai analisis sentimen, metode *Naïve Bayes,* metode *Multinomial Naïve Bayes,* dan seleksi fitur *Chi-square.* Berikut ini merupakan penelitian tersebut.

1. ***Comparative study of Naïve Bayes, Support Vector Machine, and Random Forest Classifier in Sentiment Analysis of Twitter feeds*** (Anmol Nayak & Natarajan, 2016)

Pada penelitian ini, dilakukan analisis sentimen dengan membandingkan 3 metode yaitu metode *Naïve Bayes, Support Vector Machine, dan Random Forest Classifier.* Dataset yang digunakan adalah 2000 data tweet review film pada twitter. Penelitian ini menyimpulkan bahwa dalam menganalisis sentimen review film metode *Naïve Bayes* memiliki keakuratan paling tinggi yaitu 89% diikuti oleh SVM (88%) dan *Random Forest* (85%).

1. ***Online News Classification Using Multinomial Naïve Bayes* (**Amelia Rahman et al, 2017)

Penelitian ini membahas tentang pengklasifikasian berita online yang mengambil data dari 5 situs berita online yang besar, data yang digunakan berjumlah 1011 berita. Metode yang digunakan adalah *Multinomial Naïve Bayes* dengan seleksi fitur menggunakan TF-IDF. Pada penelitian ini menghasilkan akurasi sebesar 94,29%, dan penulis menyarankan pada penelitian berikutnya mencoba menggunakan seleksi fitur yang lain untuk melakukan uji coba.

1. **Analisis Sentimen menggunakan Metode *Naïve Bayes Classifier* dengan Seleksi Fitur *Chi-square*** (Juen Ling et al, 2014)

Pada penelitian ini analisis sentimen dilakukan dengan menggunakan kombinasi metode *Naïve Bayes* dengan seleksi fitur *Chi-square.* Pada penelitian ini dijelaskan bahwa dalam seleksi fitur *Chi-square* frekuensi pada fitur menjadi kurang penting bila fitur tersebut juga muncul pada kategori yang tidak diharapkan, maka dari itu seleksi fitur *Chi-square* lebih baik daripada *frequence-based* yang hanya menggunakan *frequence* *term* dari fitur ada kategori yang diharapkan saja. Akurasi yang dihasilkan oleh penambahan fitur *Chi-square* pada metode *Naïve Bayes* adalah sebesar 90,172 %.

## **Aplikasi PRO Denpasar**

Pengaduan Rakyat Online Kota Denpasar (PRO Denpasar) adalah aplikasi berbasis website yang melibatkan partisipasi publik dan bersifat dua arah, yang digunakan sebagai alat bantu untuk melakukan monitoring dan verifikasi capaian program pembangunan maupun pengaduan masyarakat terkait pelaksanaan program pembangunan di Kota Denpasar. PRO Denpasar dikembangkan oleh Dinas Komunikasi dan Informatika Kota Denpasar.

Aplikasi PRO Denpasar berupaya untuk menjembatani partisipasi publik dalam pembangunan daerah antara masyarakat umum dengan Pemerintah Kota Denpasar. Masyarakat umum dapat memberikan pelaporan tentang pembangunan yang akan ditinjau dan didisposisikan oleh tim verifikasi PRO Denpasar kepada Instansi yang terkait untuk ditindaklanjuti.

PRO Denpasar telah terhubung dengan 52 instansi yang terdiri dari Satuan Kerja Perangkat Daerah (SKPD) Pemerintah Kota Denpasar, Perusahaan Daerah dan Kelurahan – kelurahan di Kota Denpasar.

## **Analisis Sentimen**

Analisis sentimen atau *opinion mining* merupakan studi komputasi yang menganalisis opini-opini, sentimen, serta emosi yang diekspresikan dalam teks. Tugas dari analisis sentimen adalah mengklasifikasi jenis teks yang ada pada dokumen, kalimat, atau pendapat. Teks yang ada pada dokumen, kalimat, atau pendapat memiliki aspek positif atau aspek negatif (Juen Ling, 2014).

Proses utama dalam analisis sentimen yaitu mengelompokkan teks yang terdapat dalam sebuah kalimat atau dokumen kemudian menentukan pendapat yang dikemukakan tersebut apakah bersifat positif, negatif, atau netral. Analisis sentimen dapat digunakan untuk mencari pendapat tentang produk, merek atau tokoh publik dan menentukan apakah mereka dilihat positif atau negatif (Saraswati, 2011). Hal ini memungkinkan pengguna untuk mencari informasi tentang: 1) Deteksi Flame (rants buruk), 2) Persepsi produk baru, 3) Persepsi merek, 4) Manajemen reputasi.

Sentimen atau opini mengacu pada fokus topik tertentu, pernyataan pada satu topik mungkin akan berbeda makna dengan pernyataan yang sama pada *subject* yang berbeda. Alasan tersebut menyebabkan beberapa penelitian terutama pada *review* produk didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai proses *opinion mining* (Manalu, 2014)*.*

## ***Text Preprocessing***

Tahap *text* *pre-processing* adalah tahap awal dari *text mining.* Tahap ini mencakup semua rutinitas, dan proses untuk mempersiapkan data yang akan digunakan pada operasi *knowledge discovery* sistem *text mining* (Fieldman & Sanger, 2007). Langkah yang dilakukan setelah sebuah kalimat diinputkan adalah menyetarakan karakter kalimat tersebut menjadi karakter huruf kecil semua, proses ini disebut *ToLowerCase*. Kemudian dilakukan tahap *Tokenizing, Tokenizing* yaitu proses penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan *delimiter*-*delimiter* seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kata tersebut (Weiss et al, 2005). Tahap selanjutnya yaitu *stopword* yang merupakan kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen sehingga kata-kata ini perlu dihapus. Dan terakhir adalah *stemming* yang merupakan tahapan pengembalian kata ke dalam kata dasarnya.

## **Seleksi Fitur *Chi-square***

Seleksi fitur *Chi-square* merupakan metode statistika pengujian hipotesis data kontinyu yang mengevaluasi korelasi antar dua variabel dan menentukan apakah variabel tersebut tidak berkaitan atau saling terkait. Pada tes ketertarikan, ketika diterapkan pada populasi suatu subjek, menentukan apakah subjek terkait atau tidak (Thabah, et al., 2009). Fungsi dari seleksi fitur *Chi-square* dalam memilih fitur kata pada analisis sentimen dapat dilihat Persamaan 1.

(1)

Keterangan:

*t* : Kata

*c* : Kelas/Kategori

*N* : Jumlah data training

*A* : Jumlah banyaknya dokumen pada kategori *c* yang memuat *t*

*B* : Jumlah banyaknya dokumen bukan kategori *c* yang memuat *t*

*C* : Jumlah banyaknya dokumen pada kategori *c* yang tidak memuat *t*

*D* : Jumlah banyaknya dokumen bukan kategori *c* yang tidak memuat *t*

Setelah nilai *Chi-square* setiap kata diketahui, kata tersebut akan diurutkan dari kata yang memiliki nilai *Chi-square* tertinggi hingga kata yang memiliki nilai *Chi-square* terendah. Semakin besar nilai *Chi-square* suatu kata atau fitur, maka semakin penting fitur tersebut digunakan dalam proses klasifikasi (Suharno et al, 2017).

## ***Boolean Multinomial Naïve Bayes***

Metode *Boolean Multinomial Naïve Bayes* merupakan salah satu model varian dari *Naïve Bayes Classifier* selain dari model *Multinomial Naïve Bayes.* Kedua model tersebut merupakan model yang biasanya digunakan dalam *text mining*, namun memiliki perbedaan dalam memperhatikan *term* kata. Model *Multinomial Naïve Bayes* memperhatikan *frequence of term* atau frekuensi kemunculan kata tersebut dalam suatu dokumen yang di klasifikasi, sedangkan pada model *Boolean Multinomial Naïve Bayes* tidak memperhatikan frekuensi kemunculan kata dalam dokumen dan hanya menghitung satu kali frekuensi kemunculan katanya. Sehingga model *Boolean Multinomial Naïve Bayes* lebih cocok digunakan untuk melakukan analisis sentimen karena pada analisis sentimen frekuensi suatu kata pada dokumen tidak diperhatikan (Jurafsky, 2017).

Sama seperti model *Naïve Bayes* lainnya, model *Boolean Multinomial Naïve Bayes* menghitung probabilitas setiap kelas yang ada pada dokumen lalu membanding hasilnya, misalnya suatu dokumen memiliki probabilitas kelas a yang lebih besar daripada probabilitas kelas b, maka dokumen tersebut di klasifikasikan dokumen kelas a. *Boolean Naïve Bayes* dapat dihitung dengan rumus:

(2)

Keterangan:

*P*(*c*) : Prioritas *prior* dari kelas *c*

*P*(*c* | *term* dokumen d) : Probabilitas suatu dokumen yang termasuk kelas *c*

*P*(*t*n | *c*) : Probabilitas kata ke-n dengan diketahui kelas *c*

Probabilitas prior kelas c ditentukan dengan rumus :

(3)

Keterangan:

*Nc* : Jumlah dokumen kelas c yang ada pada seluruh dokumen

*N* : Jumlah seluruh dokumen

Probabilitas kata ke-n ditentukan dengan teknik *laplacian smoothing* pada persamaan 4.

(4)

Keterangan:

*count*(*tn,c*) : Jumlah *term tn* yang ditemukan di seluruh data *training* dengan kategori c

*count*(*c*) : Jumlah *term* di seluruh data *training* dengan kategori c

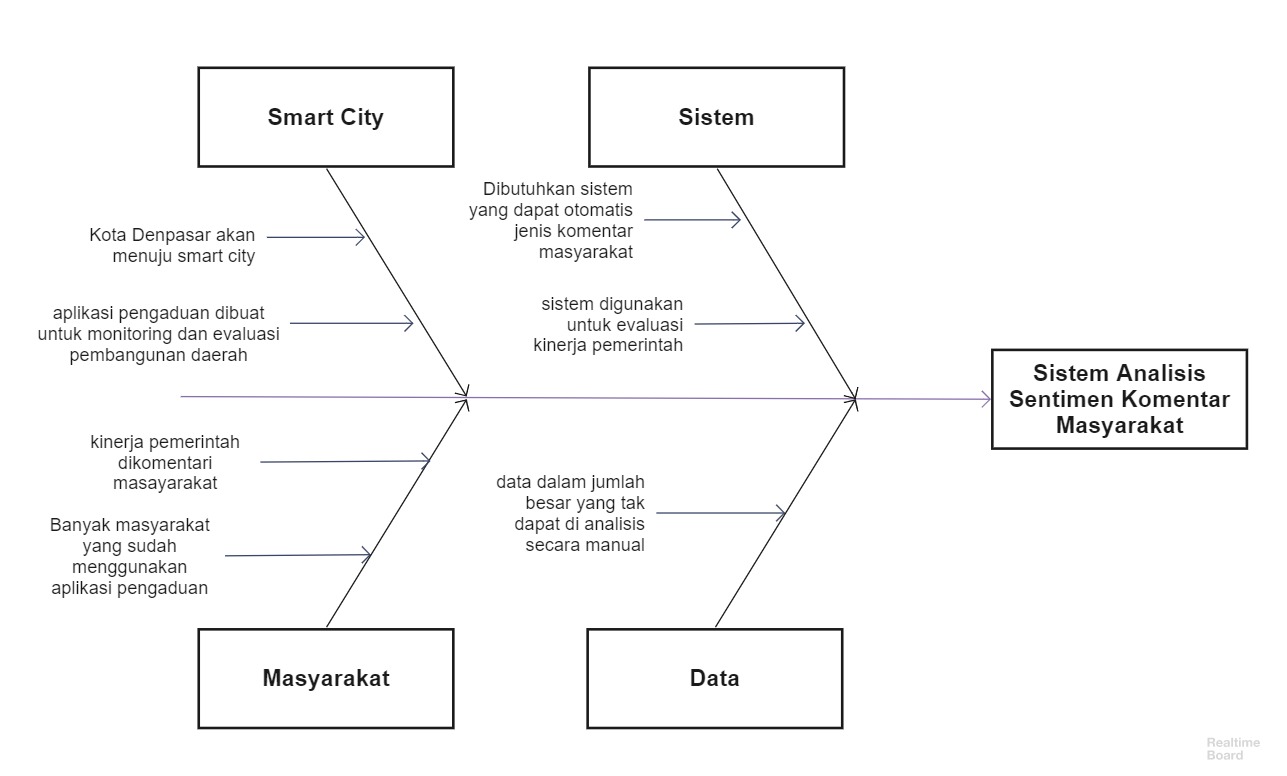
*V* : Jumlah seluruh *term* pada data *training*

Sebelum memasuki semua perhitungan diatas, pada metode *Boolean Multinomial Naïve Bayes* akan menghilangkan kata yang sama pada satu dokumen artinya nilai *term tn* pada setiap dokumen akan menjadi satu.

# **BAB III**

# **ANALISIS DAN PERANCANGAN**

## ***Fishbone* Penelitian**

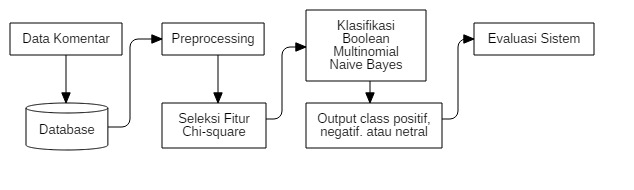


**Gambar 3.1 Fishbone Penelitian**

Pada *fishbone* ini dijelaskan gambaran mengenai hubungan sebab-akibat penelitian ini dilakukan, dimana kepala ikan merupakan akibat, dan tulang-tulang ikan merupakan sebab terjadinya penelitian ini dilakukan. Pada penelitian ini akibat yang di timbulkan adalah dibuatkannya sistem Analisis sentimen komentar masyarakat. Sebab dibuatkannya sistem ada 4 faktor yaitu *Smart city,* masyarakat, sistem, dan data. Pada faktor *smart city,* yang menjadi penyebab karena kota Denpasar akan menuju kota *smart city*, lalu untuk faktor masyarakat karena saat ini sudah banyak masyarakat yang menggunakan aplikasi dan mengomentari kinerja pemerintah, kemudian untuk data karena data dalam jumlah besar tidak dapat dihitung secara manual. Untuk sistem, karena dibutuhkan sistem yang dapat menganalis sentimen komentar secara otomatis.

## **Alur Pengolahan Data**

Pada bagian ini akan digambarkan alur metodologi penelitian secara umum,yaitu dimulai dari pengumpulan data komentar yang diekstrak ke dalam *database*, lalu data tersebut akan memasuki tahap *preprocessing*, lalu memasuki seleksi fitur untuk mengurangi fitur yang tidak penting, setelah itu data dokumen yang sudah diseleksi fiturnya akan dilakukan proses klasifikasi dan akan menghasilkan kelas positif atau negatif, dan terakhir akan di evaluasi kinerja metode yang digunakan. Untuk penjelasan lebih detail untuk setiap langkahnya akan dijelaskan selanjutnya. Adapun gambar alur metodologi penelitian ini sebagai berikut:



**Gambar 3.2 Alur Metodologi Penelitian**

## **Pengumpulan Data**

Pada pengumpulan data, jenis pengambilan data yang digunakan adalah data sekunder yaitu komentar dari aplikasi PRO Denpasar yang diperoleh langsung dari Dinas Komunikasi dan Informasi Kota Denpasar yang berjumlah sekitar ±2000 data, dimana data tersebut adalah data *training* dan data *testing* yang akan dilabeli apakah komentar tersebut positif, negatif, atau netral oleh pakar linguistik. Setelah data dikumpulkan lalu akan dimasukkan kedalam *database*. Berikut merupakan sampel data komentar positif, negatif, dan netral pada Aplikasi Pro Denpasar.

**Tabel 3.1 Contoh Komentar**

|  |  |  |
| --- | --- | --- |
| No. | Komentar | Sentimen |
| 1 | Sangat cepat tanggap, Suksma ! | Positif |
| 2 | Kasus sama, banjir di hulu, gak ada solusi | Negatif |
| 3 | Jam berapa mengalir airnya ? | Netral |

## ***Preproccessing***

Data komentar yang sudah dikumpulkan akan di olah terlebih dahulu sebelum masuk dalam tahap klasifikasi, pengolah data awal merupakan *preprocessing* yang meliputi *ToLowerCase,* Pengembalian kata, *Tokenizing,* *Stopword,* dan *Stemming.*

### ***ToLowerCase***

Tahap ini merupakan proses dimana karakter dari data kometar akan diubah menjadi karakter huruf kecil semuanya. Berikut merupakan contoh tahap *ToLowerCase.*

**Tabel 3.2 Contoh ToLowerCase**

|  |  |
| --- | --- |
| **Komentar** | **Hasil** |
| Sangat cepat tanggap, Suksma ! | sangat cepat tanggap, suksma ! |

### **Pengembalian kata**

Tahap ini merupakan proses pengembalian kata yang sebelumnya disingkat (anomali) menjadi ke bentuk awalnya. Selain itu karena aplikasi PRO Denpasar mayoritas digunakan oleh orang Bali, maka diperlukan pengembalian kata berbahasa Bali ke dalam bahasa Indonesia, karena terdapat komentar yang bercampur dengan bahasa Bali. Berikut contoh tahap pengembalian kata.

**Tabel 3.3 Contoh Pengembalian Kata**

|  |  |
| --- | --- |
| **Komentar** | **Hasil** |
| suksma, kondisi jalan sdh bersih dan sdh bisa dilalui kendaraan | terima kasih, kondisi jalan sudah bersih dan sudah bisa dilalui kendaraan |

### ***Tokenizing***

Tahap ini merupakan proses penguraian deskripsi yang semula berupa kalimat menjadi kata – kata serta menghilangkan *delimiter* – *delimiter* seperti tanda titik (.), koma (,), spasi dan karakter angka yang ada pada kalimat tersebut. Berikut contoh tahap *tokenizing.*

**Tabel 3.4 Contoh Tokenizing**

|  |  |
| --- | --- |
| **Komentar** | **Hasil** |
| terima kasih, kondisi jalan sudah bersih dan sudah bisa dilalui kendaraan | terima  kasih  kondisi  jalan  sudah  bersih  dan  sudah  bisa  dilalui  kendaraan |

### ***Stopword***

*Stopword* merupakan kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen. Misalnya “di”, “oleh”, “pada” dan lain sebagainya. Sebelum proses penghapusan *stopword* dilakukan maka harus dibuat daftar *stopword* (stoplist). Jika termasuk di dalam *stoplist* maka kata-kata tersebut akan dihapus sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata yang mencirikan isi dari suatu dokumen atau *keyword.*

**Tabel 3.5 Contoh proses stopword**

|  |  |
| --- | --- |
| **Komentar** | **Hasil** |
| terima  kasih  kondisi  jalan  sudah  bersih  dan  sudah  bisa  dilalui  kendaraan | terima  kasih  kondisi  jalan  sudah  bersih  sudah  bisa  dilalui  kendaraan |

### ***Stemming***

Tahap *stemming* dilakukan sebagai proses pemetaan dan penguraian berbagai bentuk (*variant*) dari suatu kata menjadi bentuk kata dasarnya (Tala,2003). Tujuan dari proses ini adalah menghilangkan imbuhan-imbuhan baik itu berupa prefix, sufiks, maupun konfiks yang ada pada setiap kata.

**Tabel 3.6 Contoh Stemming**

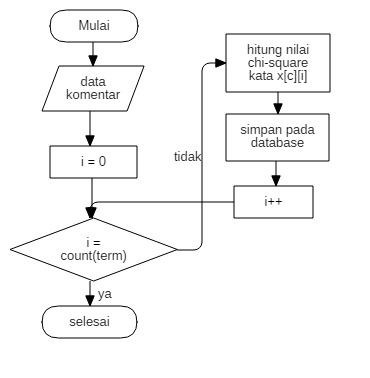
|  |  |
| --- | --- |
| **Komentar** | **Hasil** |
| terima  kasih  kondisi  jalan  sudah  bersih  sudah  bisa  dilalui  kendaraan | terima  kasih  kondisi  jalan  sudah  bersih  sudah  bisa  lalu  kendara |

## **Metode yang digunakan**

Pada penelitian ini terdapat dua metode yang digunakan, yang pertama metode untuk seleksi fitur yaitu *Chi-square,* dan metode untuk melakukan klasifikasi komentar yaitu *Boolean Multinomial Naïve Bayes.*

### **Seleksi Fitur *Chi-square***

Seleksi fitur *Chi-square* digunakan pada data *training* untuk mencari nilai *Chi-square* dari setiap *term* yang ada pada data *training* dan setelah itu akan dimasukkan dalam tabel pengetahuan yang akan digunakan nantinya pada saat memproses *testing*. Dibawah ini merupakan *flowchart* dan penjelasan langkah-langkah seleksi fitur *Chi-square.*



**Gambar 3.3 Flowchart langkah Chi-square**

Langkah-langkah yang dilakukan saat seleksi fitur dengan metode *Chi-square*:

1. Memasukkan data komentar yang diambil dari tabel komentar, pada tabel komentar ini sudah terdapat data *training* yang sudah dilabeli apakah komentar tersebut positif, negatif atau netral.
2. Inisialisasi variabel i sama dengan nol untuk melakukan perulangan.
3. Cek apalah i sama dengan jumlah kata pada komentar.
   1. Jika iya, maka proses selesai.
   2. Jika tidak, lanjutkan ke langkah 4.
4. Hitung nilai *Chi-square* kata ke-i dengan rumus :

(5)

1. Simpan nilai *Chi-square* kedalam tabel pengetahuan.
2. *Increment* i.
3. Kembali ke langkah 3.

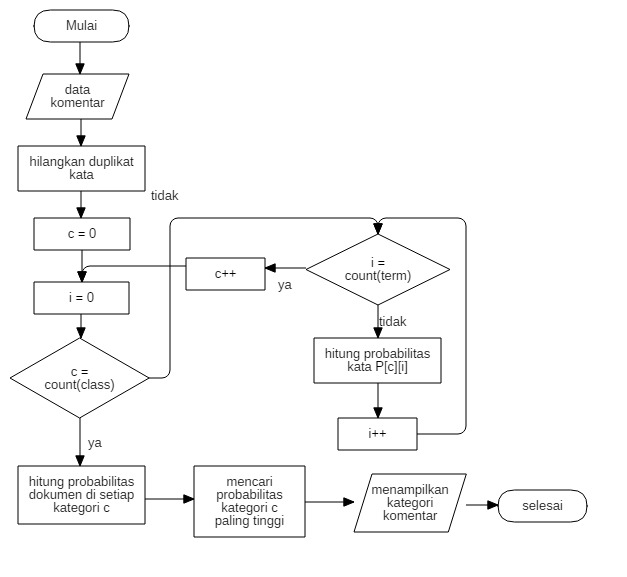
Berikut merupakan contoh tabel pengetahuan hasil seleksi fitur *Chi-square* yang diambil dari penelitian Juen Ling dkk.(2014).

**Tabel 3.7 Contoh Tabel Pengetahuan**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Fitur | Kategori | Frekuensi fitur | Nilai *Chi-square* |
| 1 | *dissapoint* | Negatif | 29 | 40.8 |
| 2 | *great* | Positif | 26 | 35.1 |
| 3 | *love* | Positif | 22 | 28.2 |
| 4 | *good* | Positif | 16 | 15.9 |
| 5 | *us* | Negatif | 10 | 11.1 |
| 6 | *excel* | Positif | 10 | 11.1 |
| 7 | *lot* | positif | 10 | 8.2 |

### ***Boolean Multinomial Naïve Bayes***

Metode *Boolean Multinomial Naïve Bayes* digunakan pada data *testing* untuk menguji data yang masuk memiliki kelas positif, negatif atau netral. Untuk memilih fitur yang digunakan akan dilihat pada tabel pengetahuan, jika ada kata yang tidak ada pada tabel pengetahuan maka akan dilakukan seleksi fitur untuk kata tersebut. Sebelum masuk kedalam metode klasifikasi sesuai dengan karakteristik metode *Boolean Multinomial Naïve Bayes,* bila terdapat kata yang muncul lebih dari satu kali maka kata tersebut hanya akan dihitung satu saja. Setelah itu akan masuk dalam perhitungan metode dan menghasilkan kelas dari sebuah komentar. Berikut merupakan *flowchart* dan langkah-langkah dari metode *Boolean Multinomial Naïve Bayes.*



**Gambar 3.4 Flowchart langkah metode Boolean Multinomial Naïve Bayes**

Langkah-langkah menguji data *testing* dengan metode *Boolean Multinomial Naïve Bayes*:

* + 1. Memasukkan komentar yang dijadikan data *testing*.
    2. Hilangkan duplikasi kata yang terdapat pada komentar
    3. Inisialisasi variabel c sama dengan nol untuk perulangan *class*
    4. Inisialisasi i sama dengan nol untuk perulangan *term*
    5. Cek apakah c sama dengan jumlah *class*

1. Jika iya, dilanjutkan ke langkah 6.
2. Jika tidak, cek apakah i sama dengan jumlah kata pada komentar
3. Jika iya, increment c lalu kembali ke langkah 4
4. Jika tidak, hitung probabilitas kata ke-i pada kelas ke-c dengan rumus:

(5)

1. Increment i, kembali ke langkah 4.
   * 1. Hitung probabilitas setiap c dengan rumus:

(6)

* + 1. Diantara semua probabilitas c, dicari yang tertinggi
    2. Tampilkan kategori komentar
    3. Selesai.

## **Implementasi**

Penelitian ini akan diimplementasikan kedalam bentuk program berbasis web dengan menggunakan bahasa pemrograman web PHP, dan MySQL untuk *database*nya.

## **Pengujian dan Evaluasi**

### **Skenario Pengujian**

Skenario pengujian pada penelitian ini yaitu, data komentar yang sebelumnya sudah dilabeli secara manual dengan bantuan pakar linguistik akan diproses melalui tahapan *preproccesing,* seleksi fitur, dan klasifikasi. Output dari sistem ini adalah label komentar yang sudah di proses sebelumnya merupakan komentar positif, negatif, atau netral. Lalu output tersebut akan dibandingkan dengan label manual yang sudah dilakukan sebelumnya sehingga evaluasi sistem dapat dilakukan.

### **Evaluasi Sistem**

Pada penelitian ini akan dilakukan pengujian terhadap hasil analisis sentimen dari data *testing*. Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasi benar dengan total data testing yang diuji dengan rumus sebagai berikut:

(7)

Evaluasi juga dilakukan pada metode seleksi fitur *Chi-square,* evaluasi dilakukan menggunakan *threshold* dari nilai *Chi-square* yang dimiliki oleh kata, jika suatu kata memiliki nilai *Chi-Square* kurang dari *threshold* maka kata tersebut akan diabaikan. Contoh tabel evaluasi seleksi fitur *Chi-square* adalah sebagai berikut.

**Tabel 3.8 Contoh tabel evaluasi Chi-square**

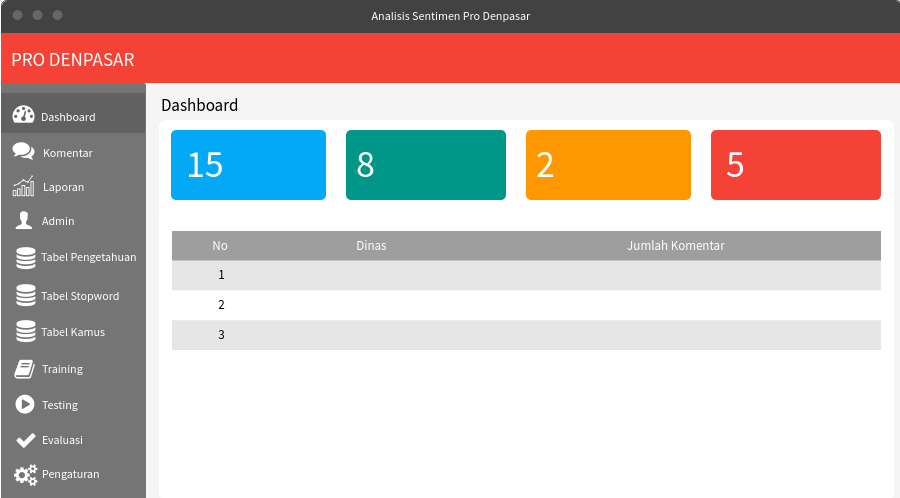
|  |  |  |
| --- | --- | --- |
| **Threshold** | **Jumlah Klasifikasi Benar** | **Akurasi (%)** |
| 0 |  |  |
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  |  |

## **Rancangan Antar Muka Sistem**

Antar muka Sistem penelitian ini dibangun menggunakan bahasa pemrograman HTML dan CSS dengan framework Bootstrap yang akan membuat tampilan menjadi responsive di berbagai platform. Pada sistem analisis sentimen ini terdiri dari beberapa menu yaitu Home atau Dashboard, Menu Admin, Menu Komentar, Menu Laporan, Tabel Pengetahuan, Tabel Stopword, Tabel Kamus, Menu Training, Menu Testing, Menu Evaluasi, Menu Pengaturan dimana semua tersebut dapat diakses oleh Super admin, sedangkan admin biasa hanya dapat mengakses menu Home, Komentar, dan Laporan.

### **Rancangan Tampilan Dashboard (Home)**

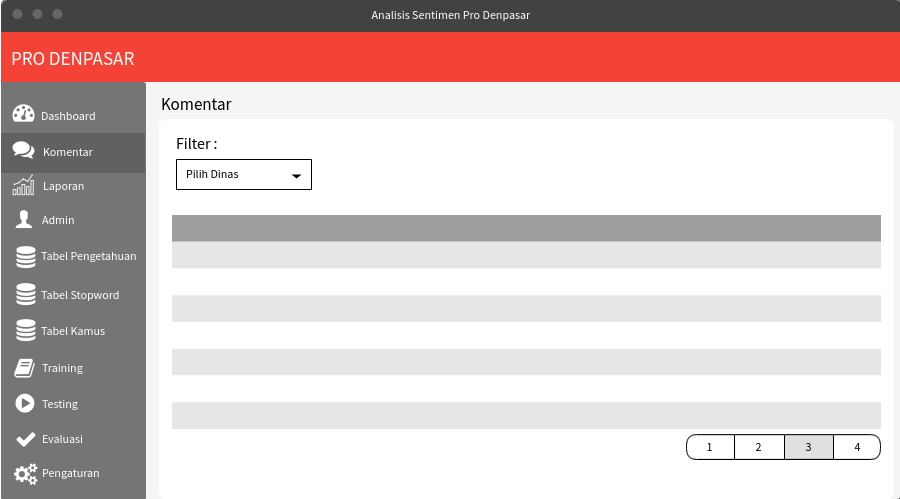
Halaman ini merupakan halaman awal yang akan ditampilkan saat Admin atau Super admin sudah melakukan login. Halaman ini menampilkan simpulan dari informasi-informasi data pada sistem contohnya, jumlah komentar, jumlah komentar positif, netral atau negatif. Rancangan tampilan dashboard dapat dilihat pada gambar 3.5.



**Gambar 3.5 Rancangan Tampilan Dashboard**

### **Rancangan Tampilan Komentar**

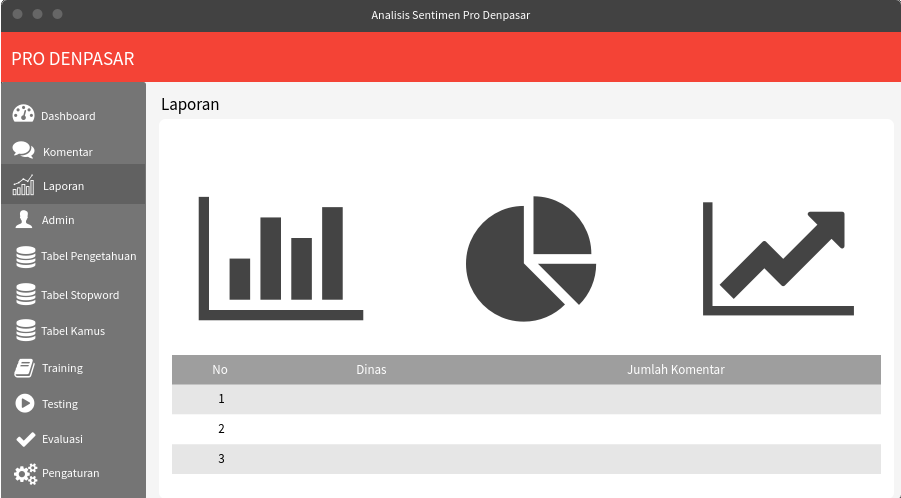
Halaman ini merupakan halaman yang menampilkan semua data komentar, dimana pada halaman ini super admin melakukan pengolahan data terhadap komentar seperti memilih data training dan data testing, memberi label kelas pada komentar, dan menghapus komentar. Rancangan tampilan komentar dapat dilihat pada gambar 3.6.



**Gambar 3.6 Rancangan Tampilan Komentar**

### **Rancangan Tampilan Laporan**

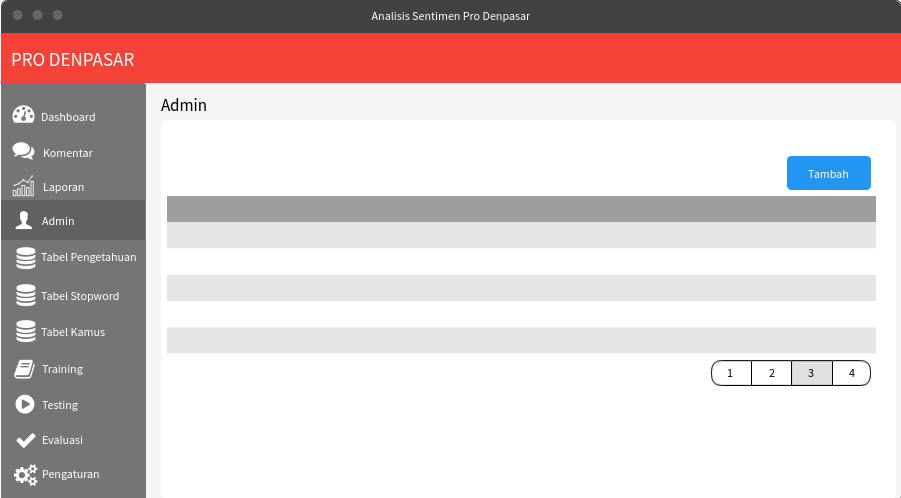
Halaman tampilan laporan menampilkan grafik data hasil olahan proses klasifikasi komentar yang telah dilakukan oleh system untuk dijadikan laporan bagi user dimana user disini adalah pihak pemerintah. Grafik yang ditampilkan seperti grafik kelas komentar dari masing-masing dinas, dan grafik komentar per bulannya. Selain menampilkan grafik pada laporan ini juga menampilkan table yang memuat informasi yang bersifat laporan. Rancangan tampilan laporan dapat dilihat pada gambar 3.7.



**Gambar 3.7 Rancangan Tampilan Komentar**

### **Rancangan Tampilan Tabel**

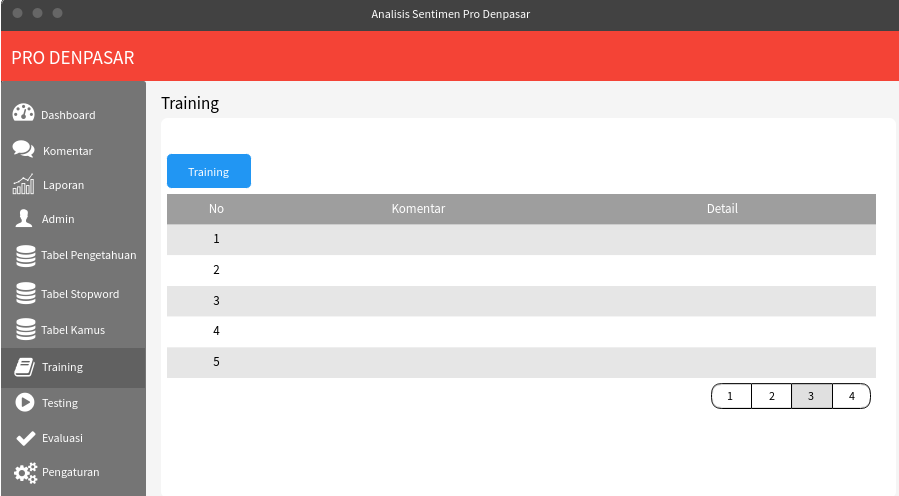
Terdapat 4 menu yang menggunakan tampilan tabel yaitu menu admin, tabel pengetahuan, tabel stopword, dan tabel kamus, dimana pada tampilan tabel ini menampilkan data pengetahuan, stopword, atau data kamus yang dapat di tambahkan, di edit, maupun dihapus. Rancangan tampilan tabel dapat dilihat pada gambar 3.8.



**Gambar 3.8 Rancangan Tampilan Komentar**

### **Rancangan Tampilan Training**

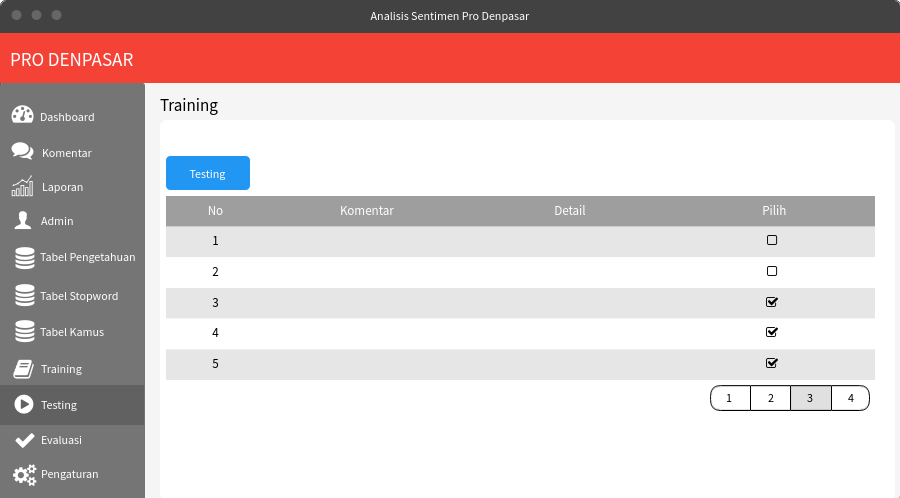
Data komentar yang telah dipilih menjadi data training pada menu komentar selanjutnya ditampilkan pada menu training, yang selanjutnya akan dilakukan proses training dengan menekan button training. Rancangan tampilan training dapat dilihat pada gambar 3.9.



**Gambar 3.9 Rancangan Tampilan Training**

### **Rancangan Tampilan Testing**

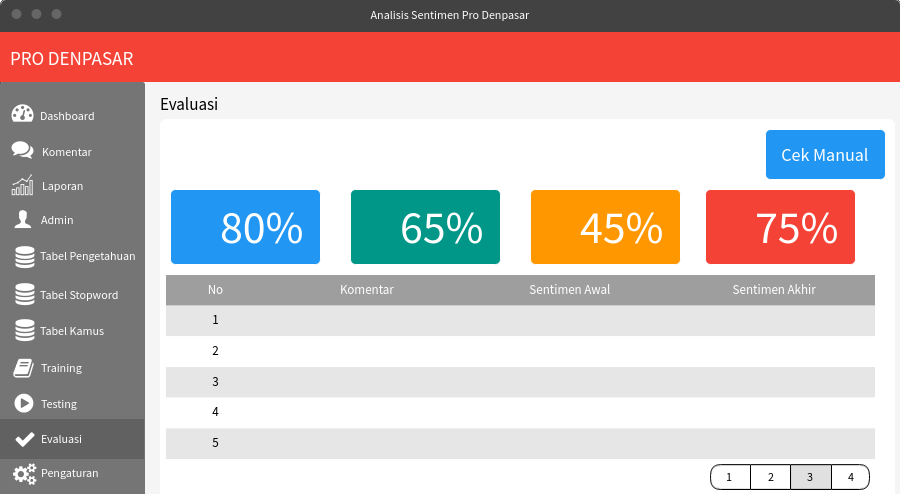
Sama halnya dengan training, pada menu testing juga menampilkan data komentar yang sebelum telah dipilih sebagai data testing pada menu komentar. Saat menekan button testing maka data-data tersebut akan diproses testing untuk mendapatkan kelas komentar dari sistem. Rancangan tampilan testing dapat dilihat pada gambar 3.10.



**Gambar 3.10 Rancangan Tampilan Testing**

### **Rancangan Tampilan Evaluasi**

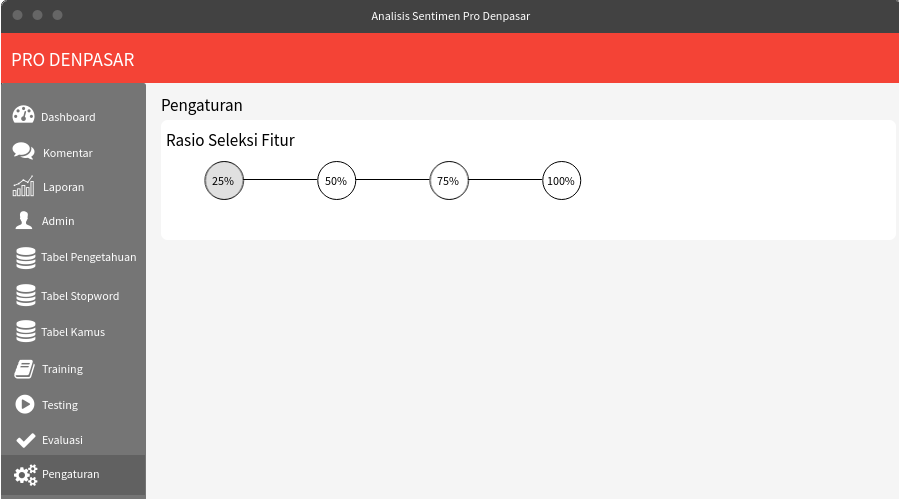
Pada menu evaluasi akan ditampilkan evaluasi dari testing data yang telah dilakukan seperti akurasi dari tiap threshold seleksi fitur *Chi-square*. Rancangan tampilan evaluasi dapat dilihat pada gambar 3.11.



**Gambar 3.11 Rancangan Tampilan Evaluasi**

### **Rancangan Tampilan Pengaturan**

Pada menu pengaturan hanya ditampilkan slider untuk mengubah threshold dari seleksi fitur yang akan dilakukan pada proses testing. Rancangan tampilan pengaturan dapat dilihat pada gambar 3.12.



**Gambar 3.12 Rancangan Tampilan Pengaturan**

# **BAB IV**

# **HASIL DAN PEMBAHASAN**

## **Gambaran Umum Sistem**

Sistem Klasifikasi Sentimen Komentar Masyarakat pada Aplikasi Pro Denpasar telah dibangun. Selain untuk menganalisis dan menguji algoritma yang digunakan untuk mengklasifikasi dan seleksi fitur, pada implementasinya sistem ini dapat digunakan oleh Pemerintah Kota Denpasar untuk melihat ringkasan tanggapan masyarakat dan melakukan evaluasi terhadap kinerja dinas-dinas atau instansi terkait dalam menanggapi pengaduan masyarakat yang telah diadukan di aplikasi Pro Denpasar.

Pada penelitian ini penulis menggunakan data komentar dari aplikasi Pro Denpasar yang diperoleh dan diberi izin oleh Dinas Komunikasi dan Informasi Kota Denpasar. Data tersebut berjumlah 2615 data yang dibagi menjadi data *training* dan data *testing* dimana data tersebut sudah dilabeli bersifat positif, negatif, atau netral dengan bantuan Dosen yang menjadi pakar pada penelitian ini.

Terdapat dua jenis pengguna yang menggunakan sistem ini yaitu admin dan super admin. Admin disini yaitu pihak Pemerintah Kota Denpasar yang dapat melihat data-data komentar yang ada beserta hasil klasifikasinya, dan dapat melihat laporan yang berupa grafik-grafik yang dapat digunakan untuk kegiatan evaluasi, sedangkan super admin disini yaitu penulis, selain dapat mengakses fitur yang dapat diakses oleh admin, super admin disini juga dapat mengakses semua fitur yang digunakan untuk penelitian ini seperti memilih data training dan testing, melakukan proses training dan testing, melabeli komentar, melihat evaluasi dari algoritma. Untuk mendapatkan sentimen komentar ini akan melalui tahapan preprocessing, seleksi fitur, dan klasifikasi.

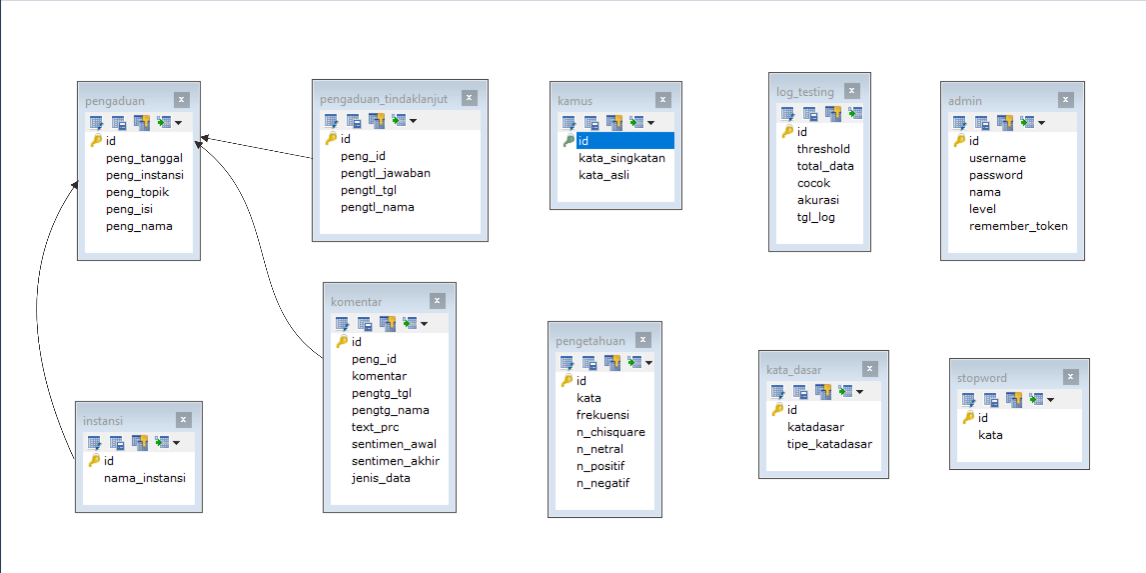
Pengujian algoritma dan seleksi fitur dilakukan pada penelitian ini untuk mengetahui seberapa tepat algoritma dapat mengklasifikasi komentar masyarakat. pengolahan data training, beragamnya data kamus yang ada sangat mempengaruhi untuk mendapatkan tingkat ketepatan algoritma yang tinggi sehingga sistem nantinya dapat digunakan dan diimplementasikan ke lapangan.

## **Lingkungan Perancangan dan implementasi Sistem**

Sistem ini diimplementasikan kedalam sistem berbasis web dengan bahasa pemrograman *PHP* dan menggunakan *Laravel* sebagai *framework* dari *PHP*, sedangkan untuk *Database Management System* (DBMS) menggunakan MySQL. Dalam merancang dan mengimplementasikan sistem ini terdapat beberapa perangkat lunak yang digunakan yaitu:

1. Atom Text Editor
2. XAMPP v3.2.2
3. SQLyog Ultimate
4. Git
5. Google Chrome

## **Implementasi Basis Data**

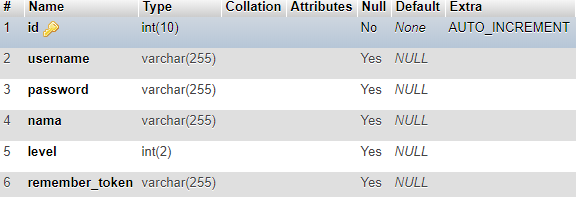


**Gambar 4.1 Skema Basis Data**

Pada gambar 4.1 ditunjukan skema basis data yang digunakan dalam implementasi sistem klasifikasi sentimen komentar masyarakat. Terdapat 10 tabel dalam implementasi basis data tersebut yaitu tabel admin, tabel instansi, tabel kamus, tabel kata dasar, tabel log testing, tabel pengaduan, tabel pengaduan tindaklanjut, tabel pengetahuan, dan tabel stopword yang akan dijelaskan lebih detil dibawah ini.

### **Tabel Admin**

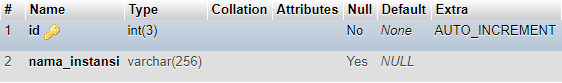
Tabel admin digunakan untuk menyimpan data pengguna sistem yaitu admin dan super admin yang nantinya akan dibedakan pada kolom *level*. Tabel admin ini terdiri dari kolom *id* sebagai *primary key*, kolom *username* untuk username pengguna, kolom password untuk kata sandi pengguna, kolom nama untuk identitas pengguna, kolom level untuk level pengguna apakah admin atau super admin, dan kolom remember token untuk pengingat pengguna. Tabel admin dapat dilihat pada gambar 4.2.



**Gambar 4.2 Tabel Admin**

### **Tabel Instansi**

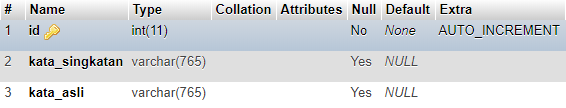
Tabel Instansi digunakan untuk menyimpan data instansi-instansi yang ada di Pemerintahan Kota Denpasar yang sudah terintegrasi dengan aplikasi Pro Denpasar. Tabel instansi ini terdiri dari dua kolom yaitu kolom *id* sebagai *primary key* dan kolom nama untuk nama dari instansi tersebut. Tabel instansi dapat dilihat pada gmabr 4.3.



**Gambar 4.3 Tabel Instansi**

### **Tabel Kamus**

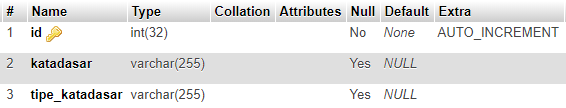
Tabel Kamus digunakan untuk menyimpan data kata-kata singkatan atau kata-kata anomali yang memiliki kata yang sebenarnya, ini berfungsi agar pada tahap preproccessing data kata menjadi seragam sehingga ketepatan algoritma tidak menurun. Selain itu juga terdapat kata-kata berbahasa bali yang diterjamahkan ke dalam bahasa indonesia pada tahap pengembalian kata, karena pada data komentar terdapat kalimat atau kata bahasa bali yang disisipkan. Tabel Kamus terdiri dari 3 kolom yaitu kolom id untuk primary key, kolom kata\_singkatan untuk kata awal, lalu kolom asli untuk hasil terjemahan kata awal. Tabel kamus dapat dilihat pada gambar 4.4.



**Gambar 4.4 Tabel Kamus**

### **Tabel Kata Dasar**

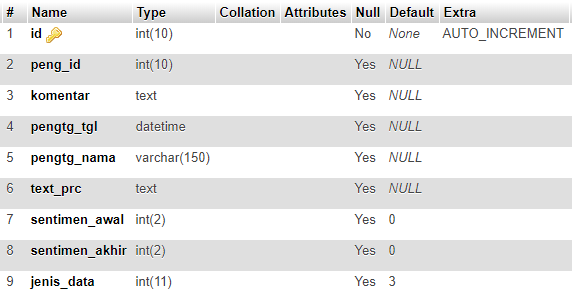
Tabel Kata Dasar digunakan untuk menyimpan data kata dasar yang digunakan pada tahap *Stemming* yaitu merubah kata berimbuhan menjadi kata dasarnya. Terdapat 3 kolom pada tabel kata dasar yaitu kolom id untuk primary key, kolom katadasar untuk menyimpan kata, dan kolom kolom tipe\_katadasar untuk jenis kata dasar tersebut. Tabel Kata Dasar dapat dilihat pada gambar 4.5.



**Gambar 4.5 Tabel Kata Dasar**

### **Tabel Komentar**

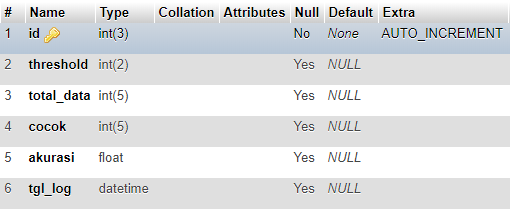
Tabel Komentar digunakan untuk menyimpan data komentar masyarakat yang memiliki label dari pakar, dan label hasil dari proses sistem. Tabel komentar terdiri dari 9 kolom yaitu, kolom id sebagai primary key, kolom peng\_id sebagai foreign key terhadap tabel pengaduan, kolom komentar untuk menyimpan data komentar masyarakat, kolom pengtg\_tgl untuk menyimpan waktu saat komentar diterbitkan, kolom pengtg\_nama untuk menyimpan data nama masyarakat yang melakukan komentar, kolom text\_prc untuk menyimpan hasil *preprocessing* komentar, kolom sentimen\_awal untuk menyimpan label dari pakar, kolom sentimen\_akhir untuk menyimpan label dari hasil proses sistem, dan kolom jenis\_data untuk menyimpan jenis dari data tersebut apakah data *training* atau data *testing*. Tabel Komentar dapat dilihat pada gambar 4.6.



**Gambar 4.6 Tabel Komentar**

### **Tabel Log Testing**

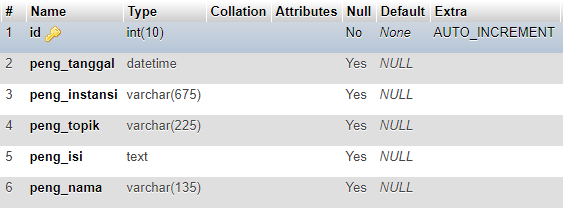
Tabel Log Testing digunakan untuk menyimpan data riwayat penelti melakukan *testing* untuk mendapatkan akurasi yang setinggi-tingginya. Tabel Log Testing terdiri dari 6 kolom yaitu kolom id sebagai primary key, kolom threshold untuk menyimpan threshold yang digunakan saat melakukan *testing*, kolom total\_data untuk menyimpan total data *testing* yang digunakan saat melakukan *testing*, kolom cocok untuk menyimpan berapa data yang sentimen dari sistemnya yang cocok dengan sentimen dari pakarnya, kolom akurasi untuk menyimpan akurasi hasil dari *testing* tersebut yang dihitung dengan cara membagi nilai cocok dengan total data, dan kolom tgl\_log untuk menyimpan waktu saat melakukan *testing* tersebut.



**Gambar 4.7 Tabel Log Testing**

### **Tabel Pengaduan**

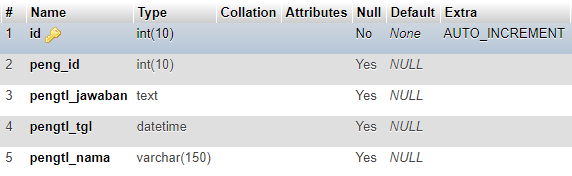
Tabel Pengaduan digunakan untuk menyimpan data pengaduan dari masyarakat, pada pengaduan ini lah nantinya akan ditanggapi oleh dinas terkait dan di komentari oleh masyarakat terkait dengan pengaduan yang sudah diadukan dan kinerja dari dinas terkait menangani pengaduan tersebut. Pada tabel pengaduan ini terdapat 6 kolom yaitu kolom id sebagai primary key, kolom peng\_tanggal untuk menyimpan waktu diterbitkannya pengaduan, kolom peng\_instansi untuk menyimpan instansi yang dituju oleh pengaduan, kolom peng\_topik untuk menyimpan data topik atau judul dari pengaduan yang diterbitkan, kolom peng\_isi untuk menyimpan data isi dari pengaduan, kolom peng\_nama untuk menyimpan pengirim dari pengaduan tersebut. Tabel Pengaduan dapat dilihat pada gambar 4.8.



**Gambar 4.8 Tabel Pengaduan**

### **Tabel Pengaduan Tindak Lanjut**

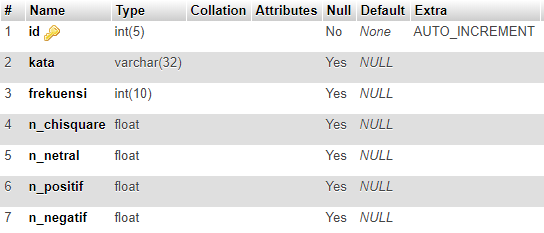
Tabel Pengaduan Tindak Lanjut digunakan untuk menyimpan data tanggapan dari dinas terkait terhadap pengaduan yang diterima. Tabel Pengaduan terdiri dari 5 kolom yaitu kolom id sebagai primary key, kolom peng\_id sebagai foreign key terhadap Tabel Pengaduan, kolom pengtl\_jawaban untuk menyimpan isi dari tanggapan yang diberikan, kolom pengtl\_tgl untuk menyimpan waktu diterbitkannya tanggapan, kolom pengtl\_nama untuk menyimpan instansi yang melakukan tanggapan. Tabel Pengaduan Tindak Lanjut dapat dilihat pada gambar 4.9.



**Gambar 4.9 Tabel Pengaduan Tindak Lanjut**

### **Tabel Pengetahuan**

Tabel Pengetahuan digunakan untuk menyimpan data pengetahuan yang dihasilkan dari proses *training* yang berupa kata yang memiliki frekuensi, nilai *Chi-Square*, nilai netral, nilai positif, dan nilai negatif yang disimpan di setiap kolomnya. Tabel Pengetahuan dapat dilihat pada gambar 4.10.

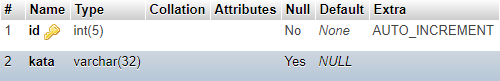


**Gambar 4.10 Pengetahuan**

### **Tabel Stopword**

Tabel Stopword digunakan untuk menyimpan data kata yang menjadi stopword yang nantinya digunakan saat tahap menghilang stopword pada *preprocessing.* Terdapat dua kolom pada tabel ini yaitu kolom id sebagai primary key dan kolom kata untuk menyimpan kata yang menjadi stopword.

.



**Gambar 4.11 Stopword**

## **Fitur Sistem Klasifikasi Komentar Masyarakat Pro Denpasar**

Pada sistem klasifikasi komentar masyarakat ini terdapat dua jenis pengguna yang dapat mengakses fitur-fitur yang ada dalam sistem ini, admin hanya dapat mengakses beberapa fitur saja sedangkan super admin dapat mengakses semua fitur yang ada pada sistem. Fitur-fitur tersebut akan dijelaskan pada tabel 4.1 dibawah ini beserta pengguna yang dapat mengaksesnya.

**Tabel 4.1 Fitur Sistem**

|  |  |
| --- | --- |
| **Pengguna** | **Fitur** |
| Admin | 1. Admin dapat melakukan login 2. Admin dapat melihat beberapa dashboard 3. Admin dapat mengakses menu Pengaduan (detail dan hapus pengaduan) 4. Admin dapat mengakses menu Laporan |
| Super Admin | 1. Super Admin dapat melakukan login 2. Super Admin dapat melihat semua dashboard 3. Super Admin dapat mengakses menu Pengaduan (detail dan hapus pengaduan) 4. Super Admin dapat mengakses menu Laporan 5. Super Admin dapat mengakses menu Komentar (kelola komentar) 6. Super Admin dapat mengelola akun 7. Super Admin dapat mengakses menu Pengetahuan 8. Super Admin dapat mengelola stopword 9. Super Admin dapat mengelola kamus 10. Super Admin dapat melakukan proses Training 11. Super Admin dapat melakukan proses Testing 12. Super Admin dapat mengakses menu Evaluasi |

## **Implementasi Antarmuka**

### **Tampilan Login**

Tampilan ini merupakan tampilan utama ketika sistem pertama dibuka, untuk dapat mengakses semua fitur sistem harus melakukan login terlebih dahulu baik admin maupun super admin. Admin dan super admin dibedakan pada kolom level pada database, sehingga pada saat setelah login admin dan super admin memiliki akses yang berbeda pada beberapa fitur. Untuk melakukan login dibutuhkan username dan password yang telah terdaftar pada database. Untuk tampilan login dapat dilihat pada gambar 4.12.

### **Tampilan Menu Dashboard (Admin)**

Setelah melakukan login, Admin akan dibawa ke halaman menu dashboard ini. Untuk admin, menu dashboard menampilkan informasi-informasi umum dan ringkas seputar komentar-komentar yang ada pada tahun ini dan bulan ini. Untuk tampilan menu dashboard admin dapat dilihat pada gambar 4.13.

### **Tampilan Menu Dashboard (Super Admin)**

Sama halnya dengan Admin, setelah login Super Admin akan dibawa ke halaman menu dashboard, namun menu dashboard pada Super Admin menampilkan informasi yang lebih beragam dari pada Admin, selain yang ditampilkan pada dashboard Admin, menu dashboard pada Super Admin juga menampilkan informasi umu mengenai data training, data testing, pengetahuan, kamus, stopword, dan lainnya yang berhubungan dengan penelitian sistem. Untuk tampilan menu dashboard super admin dapat dilihat pada gambar 4.14.

### **Tampilan Menu Pengaduan**

Pada menu pengaduan pengguna dapat melihat daftar komentar yang disortir berdasarkan komentar terbaru yang masuk. Komentar tersebut ditampilkan dengan judul pengaduan yang dikomentari, instansi yang dituju, waktu komentar, dan sentimen komentar tersebut yang telah diproses oleh sistem, untuk aksi terdapat detail dan hapus, detail untuk melihat detail dari pengaduan dan tanggapan yang dikomentari, hapus untuk menghapus komentar tersebut. Daftar komentar pada menu ini dapat di filter berdasarkan sentimen komentar, indtansi yang dituju, dan kata yang terdapat pada isi komentar. Menu pengaduan ini dapat diakses oleh Admin dan Super Admin. Untuk tampilan menu pengaduan dapat dilihat pada gambar 4.15.

### **Tampilan Menu Laporan**

Pada menu laporan pengguna dapat melihat grafik-grafik yang menunjukkan informasi ringkas yang dapat digunakan sebagai laporan evaluasi kinerja instansi-instansi dalam menangani pengaduan berdasarkan sentimen komentar masyarakat sebulan atau setahun terakhir. Menu laporan dapat diakses oleh admin dan super admin. Untuk tampilan menu laporan dapat dilihat pada gambar 4.16.

### **Tampilan Menu Komentar**

Pada menu komentar, Super admin dapat mengelola data komentar yang digunakan untuk pengembangan sistem. Pada menu komentar ini ditampilkan isi komentar dengan jenis data dan sentimen pakar komentar yang dapat diubah, pada menu ini Super Admin mengelola data komentar tersebut apakah dijadikan data training atau data testing, super admin juga memilih sistem pakar dari komentar tersebut sesuai dengan data yang telah didiskusikan sebelumnya dengan dosen narasumber. Kumpulan komentar tersebut dapat di filter berdasarkan sentimen pakar komentar, jenis data komentar, dan kata dari isi komentar. Menu komentar hanya dapat di akses oleh Super Admin. Untuk tampilan menu komentar dapat dilihat pada gambar 4.17.

### **Tampilan Menu Admin**

Pada menu admin, Super Admin dapat mengelola akun-akun yang dapat mengakses sistem. Super admin dapat mengedit, mengganti password, dan menghapus akun tersebut. Menu admin hanya dapat akses oleh Super Admin. Untuk tampilan menu admin dapat dilihat pada gambar 4.18.

### **Tampilan Menu Pengetahuan**

Pada menu pengetahuan, Super admin dapat melihat hasil dari pengetahuan yang dihasilkan oleh proses training data komentar. Pengetahuan tersebut merupakan kata yang memiliki frekuensi kemunculan, nilai chi-square, nilai netral, nilai positif, dan nilai negatif. Super admin juga dapat mengedit dan menghapus pengetahuan tersebut. Menu Pengetahuan hanya dapat diakses oleh Super Admin. Untuk tampilan menu pengetahuan dapat dilihat pada gambar 4.19.

### **Tampilan Menu Stopword**

Pada menu stopword, Super Admin dapat mengelola stopword, stopword disini merupakan kumpulan kata yang nantinya digunakan pada tahap penghilangan stopword di preproccessing. Super admin dapat menambahkan, mengedit dan menghapus stopword. Menu stopword hanya dapat diakses oleh Super admin. Untuk tampilan menu stopword dapat dilihat pada gambar 4.20.

### **Tampilan Menu Kamus**

Pada menu kamus, Super admin dapat mengelola kamus, kamus disini merupakan kata yang memiliki kata awal dan kata terjemahan yang digunakan saat tahap pengembalian kata pada preprocessing. Super admin dapat menambahkan, mengedit, dan menghapus kamus. Menu kamus hanya dapat diakses oleh Super admin. Untuk tampilan menu kamus dapat dilihat gambar 4.21.

### **Tampilan Menu Training**

Pada menu training, super admin dapat melakukan proses training terhadap data yang telah dipilih sebagai data training pada menu komentar. Super admin dapat menekan tombol training pada kanan atas halaman untuk melakukan proses training. Setelah proses training selesai dilakukan maka kolom text processing yang mulanya kosong akan diisi dengan hasil dari preprocessing data tersebut, dan hasil training lainnya adalah pengetahuan yang memiliki nilai yang dapat dilihat pada menu pengetahuan. Menu training hanya dapat diakses oleh Super Admin. Untuk tampilan menu training dapat dilihat pada gambar 4.22.

### **Tampilan Menu Testing**

Pada menu testing, super admin dapat melakukan proses testing terhadap data yang telah dipilih sebagai data testing pada menu komentar. Super admin dapat menekan tombol testing pada kanan atas halaman untuk melakukan proses testing. Setelah proses testing selesai dilakukan maka kolom text processing dan kolom sentimen sistem yang mulanya kosong akan diisi oleh hasil dari proses testing tersebut, dan log dari testing tersbut akan masuk ke tabel log testing. Menu testing hanya dapat diakses oleh Super admin. Untuk tampilan menu testing dapat dilihat pada gambar 4.23.

### **Tampilan Menu Evaluasi**

Pada menu evaluasi, super admin dapat melihat akurasi dari ketepatan algoritma terhadap hasil dari proses testing. Menu evaluasi menampilkan data hasil dari proses testing yang sebelumnya sudah disimpan pada tabel log testing pada database, data tersebut terdiri dari threshold yang digunakan saat testing, total data testing, kecocokan sentimen pakar data dengan sentimen sistem data, waktu melakukan testing, dan akurasi. Menu evaluasi hanya dapat diakses oleh super admin. Untuk tampilan menu evaluasi dapat dilihat pada gambar 4.24.

## **Implementasi Sistem**

Pada implementasi sistem ini akan dijelaskan bagaimana metode algoritma yang digunakan pada sistem ini diimplementasikan dalam bentuk bahasa kode bahasa pemrograman web. Ada 3 bagian yang akan dibahas yaitu implementasi *text preprocessing,* implementasi seleksi fitur *chi-square,* dan implementasi algoritma *Boolean Multinomial Naïve Bayes.*

### **Implementasi *Text Preprocessing***

*Text Processing* ini dibagi lagi kedalam beberapa tahapan yaitu tahap membersihkan kalimat, tahap *Tokenize,* tahap pengembalian kata, tahap menghilangkan *stopword*, dan tahap *stemming.* *Text Processing* ini digunakan pada saat proses training data maupuan testing data.

Pada tahap membersihkan kalimat, setelah data komentar diambil selanjutnya setiap komentar akan dimasukkan kedalam variabel training untuk melakukan proses membersihkan kalimat yaitu menghilangkan simbol-simbol, angka, dan penggunaan “tab” dan “enter” pada kalimat tersebut. Implementasi tahap membersihkan kata dapat dilihat pada tabel 4.2.

**Tabel 4.2 Tahap Membersihkan Kalimat**

|  |  |
| --- | --- |
| **Baris** | **Kode** |
| 1  2  3  4 | $training=strtolower(preg\_replace('([.,/123457890@!#$%&-\_=+?`~{;:<>}])','',$trainings));  $training= strtolower(preg\_replace('([\n\r])',' ',$training)); |

Setelah kalimat dibersihkan selanjutnya kalimat tersebut akan dipecah menjadi kata dan disimpan kedalam variable array. Implementasi tahap *tokenize* dapat dilihat pada tabel 4.3.

**Tabel 4.3 Tahap Tokenize**

|  |  |
| --- | --- |
| **Baris** | **Kode** |
| 1  2 | $katass = array\_values(array\_filter((explode(' ',$training)))); |

Setalah kalimat dipecah menjadi kata maka selanjut setiap kata tersebut akan di cek pada tabel kamus apakah merupakan kata yang harus diubah atau tidak, kata yang harus diubah adalah anomali kata dan kata berbahasa bali. Implementasi tahap pengembalian kata dapat dilihat pada tabel 4.4.

**Tabel 4.4 Tahap Pengembalian Kata**

|  |  |
| --- | --- |
| **Baris** | **Kode** |
| 1  2  3  4  5  6  7  8 | $kamus=Kamus::select('kata\_asli')  ->where('kata\_singkatan',$katas)->first();  if($kamus){  $kata = $kamus->kata\_asli;  }  else{  $kata = $katas;  } |

Pada perulangan yang sama setelha kata tersebut dikembalikan akan di cek apakah kata tersebut merupakan *stopword* atau tidak, jika tidak kata tersebut akan dimasukkan kedalam variabel array sedangkan jika iya yang dimasukkan ke variabel adalah NULL. Untuk implementasi tahap menghilangkan *stopword* dapat dilihat pada tabel 4.5.

**Tabel 4.5 Tahap Penghilangan Stopword**

|  |  |
| --- | --- |
| **Baris** | **Kode** |
| 1  2  3  4  5  6  7  8 | $stopword=Stopword::select('kata')  ->where('kata',$kata)->first();  if($stopword){  $data2[] = '';  }  else {  $data2[] = $kata;  } |

Tahap terakhir adalah *stemming* yaitu mengubah kata tersebut menjadi kata dasarnya. Setelah itu kata-kata itu akan digabungkan kembali menjadi kalimat yang bersih dan akan disimpan pada kolom text\_prc data masing-masing pada tabel komentar. Implementasi tahap *stemming* dan penyimpanan hasil *text processing* dapat dilihat pada tabel 4.6.

**Tabel 4.6 Tahap Stemming**

|  |  |
| --- | --- |
| **Baris** | **Kode** |
| 1  2  3  4  5  6  7  8  9  10  11 | //stemming  $stemmerFactory=new \Sastrawi\Stemmer\StemmerFactory;  $stemmer = $stemmerFactory->createStemmer();  $sentence=Implode('',array\_values(array\_unique(array\_filter($data2))));  $output = $stemmer->stem($sentence);  //simpan hasil preproccessing  $text\_prc = Komentar::find($id);  $text\_prc->text\_prc = $output;  $text\_prc->save(); |

### **Implementasi *Chi-Square***

Implementasi *Chi-Square* ini digunakan saat melakukan proses training, semua kata yang terdapat pada data training yang sudah melalui proses *Text Processing* akan dicari nilai *Chi-Square*nya, ini berfungsi saat nanti menggunakan seleksi *Chi-Square* saat melakukan proses testing. Proses pencarian nilai *chi-square* dilakukan 3 kali untuk setiap kata yaitu untuk nilai *chi-square* saat positif, negatif, dan netral yang nantinya ketiga nilai *chi-square* tersebut akan dijumlahkan lalu disimpan pada tabel pengetahuan. Untuk implementasi *Chi-Square* dapat dilihat pada tabel 4.7.

**Tabel 4.7 Implementasi Chi-Square**

|  |  |
| --- | --- |
| **Baris** | **Kode** |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40 | $N = Komentar::where('jenis\_data','0')->count(); //seluruh data training  $A = Komentar::where('jenis\_data','0')  ->where('sentimen\_awal',$k)->where('text\_prc','REGEXP','[[:<:]]'.$t.'[[:>:]]')->count(); // data ber kelas $k yg memuat kata $t  $B = Komentar::where('jenis\_data','0')  ->where('sentimen\_awal','!=',$k)->where('text\_prc','REGEXP','[[:<:]]'.$t.'[[:>:]]')->count(); // data selain kelas $k yg memuat kata $t  $C = Komentar::where('jenis\_data','0')  ->where('sentimen\_awal',$k)  ->where('text\_prc','NOT REGEXP','[[:<:]]'.$t.'[[:>:]]')->count(); // data ber kelas $k yg tidak memuat kata $t  $D = Komentar::where('jenis\_data','0')  ->where('sentimen\_awal','!=',$k)  ->where('text\_prc','NOT REGEXP','[[:<:]]'.$t.'[[:>:]]')->count(); // data selain kelas $k yg tidak memuat kata $t  $AD = $A\*$D;  $BC = $B\*$C;  $AC = $A+$C;  $BD = $B+$D;  $AB = $A+$B;  $CD = $C+$D;  //pembilang  $ADBC = $AD-$BC;  $ADBC2 = $ADBC\*$ADBC;  $pembilang = $N\*$ADBC2;  //penyebut  $penyebut = $AC\*$BD\*$AB\*$CD;  //Chi-square  $x[$key][$k] = $pembilang/$penyebut; |

### **Implementasi *Boolean Multinomial Naïve Bayes***

Implementasi *Boolean Multinomial Naïve Bayes* yang merupakan algoritma untuk melakukan klasifikasi terhadap data komentar digunakan saat proses testing dilakukan. Setelah melalui proses *text processing*, data akan diklasifikasi dengan *threshold* yang sudah ditentukan sebelumnya, jika terdapat kata pada data yang memiliki nilai dibawah *threshold* maka kata tersebut akan diabaikan. Proses perhitungan akan dilakukan 3 kali untuk setiap data komentar yaitu untuk mencari nilai probabilitas untuk kelas positif, kelas negatif, dan kelas netral. Sentimen sistem untuk data tersebut adalah yang memiliki nilai probabilitas tertinggi diantara nilai positif, negatif, dan netral tersebut. Implementasi dari metode *Boolean Multinomial Naïve Bayes* dapat dilihat pada tabel 4.8.

**Tabel 4.8 Implementasi Boolean Multinomial Naïve Bayes**

|  |  |
| --- | --- |
| **Baris** | **Kode** |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40  41  42 | for ($k=0; $k<3 ; $k++) {  $nc= Komentar::where('jenis\_data','0')  ->where('sentimen\_awal',$k)->count();  $prior = $nc/$n;  $probKatas[$k] = 1;  foreach ($ts as $key2 => $t) {  $pengetahuan=Pengetahuan::where('kata',$t)  ->first();  if($pengetahuan){  $chisquare = $pengetahuan->n\_chisquare;  }  if($pengetahuan AND $chisquare >= $threshold){  if($k==0){  $probKata[0][$key2] = $pengetahuan  ->n\_netral;  }  else if ($k==1) {  $probKata[1][$key2] = $pengetahuan  ->n\_positif;  }  else if ($k==2) {  $probKata[2][$key2] = $pengetahuan  ->n\_negatif;  }  }  else {  $probKata[$k][$key2] = 1;  }  $probKatas[$k]=  $probKatas[$k] \*$probKata[$k][$key2];  }  $probKomentar[$k]= $prior\*$probKatas[$k];  }  $data2[$key] = $probKomentar;  $updateAkhir = Komentar::find($id);  $updateAkhir->sentimen\_akhir= array\_search(max($probKomentar),$probKomentar);  $updateAkhir->save(); |

## **Pengujian Sistem**

Pada penelitian ini pengujian sistem dilakukan ada dua yaitu mengetahui akurasi dari metode *Boolean Multinomial Naïve Bayes* dalam mengklasifikasi data komentar, dan pengujian terhadap pengaruh seleksi fitur *chi-square* dalam menyeleksi fitur yang ada sehingga fitur yang tidak penting dapat direduksi sebelum dilakukannya proses klasifikasi.

### **Akurasi Sistem**

### **Pengaruh Seleksi Fitur**

# **BAB V**

# **PENUTUP**

## **Kesimpulan**

## **Saran**

# **DAFTAR PUSTAKA**

Fieldman, R., & Sanger, J. (2009). *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data.* New York: Cambridge University Press.

Jurafsky, D. (2017). *Speech and Language Processing 3rd edition.*

Kominfo. (2013). *Informasi Standar Pelayanan Pengaduan Rakyat Online (PRODENPASAR).* Kota Denpasar.

Ling, J., Kencana, I., & Bagus Oka, T. (2014). ANALISIS SENTIMEN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER DENGAN SELEKSI FITUR CHI SQUARE . *E-Jurnal Matematika*, 92-99.

Liu, B. (2017, November 19). *Sentiment Analysis And Opinion Mining. Chicago: Morgan & Claypool Publisher.* Diambil kembali dari http://www.dcc.ufrj.br/~valeriab/DTMSentimentAnalysisAndOpinionMining-BingLiu.pdf

Manalu, B. (2014). Analisis Sentimen Pada Twitter Mennggunakan Text Mining.

Nayak, A., & Natarajan. (2016). Comparative study of Naïve Bayes, Support Vector Machine and Random Forest Classifiers in Sentiment Analysis of Twitter feeds. *International Journal of Advanced Studies in Computer Science and Engineering*, Volume 5.

Rahman, A., Wiranto, & Doewes, A. (2017). Online News Classification Using Multinomial Naive Bayes. *ITSMART: Jurnal Ilmiah Teknologi dan Informasi*, Vol. 6 No.1.

Saraswati. (2011). Text Mining dengan Metode Naïve Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis. *Tesis. Bali: Universitas Udayana*.

Snedecor, W., & Cochran, W. (1989). *Statistical Methods 8th penyunt.* s.l.: State University Press.

Tala, F. (2017, 11 20). *A Study of Stemming Efects on Information Retrieval in Bahasa Indonesia. Institute for Logic, Language and Computation Universiteit van Amsterdam The Netherlands. .* Retrieved from http://www.illc.uva.nl/Research/Reports/MoL-2003-02.text.pdf

Thabtah, F., Eljinini, M., & Zamzeer, M. (2009). Naïve Bayesian Based on Chi Square to Categorize Arabic Data. *Communications of the IBIMA*, 158-163.

Weiss, S., Indurkhya, N., Zhang, T., & Damerau, F. (2005). *Text Mining : Predictive Methods for Analyzing Unstructured Information.* New York: Springer.