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

**KOMPETENSI KOMPUTASI**

**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**

[LEMBAR PENGESAHAN PROPOSAL TUGAS AKHIR ii](#_Toc503958687)

[KATA PENGANTAR iii](#_Toc503958688)

[DAFTAR ISI iv](#_Toc503958689)

[DAFTAR TABEL vi](#_Toc503958690)

[DAFTAR GAMBAR vii](#_Toc503958691)

[1. Latar Belakang 1](#_Toc503958692)

[2. Rumusan Masalah 3](#_Toc503958693)

[3. Tujuan Penelitian 3](#_Toc503958694)

[4. Batasan Masalah 3](#_Toc503958695)

[5. Manfaat Penelitian 4](#_Toc503958696)

[6. Tinjauan Pustaka 4](#_Toc503958697)

[6.1 Tinjauan Studi 4](#_Toc503958698)

[6.2 Aplikasi PRO Denpasar 6](#_Toc503958699)

[6.3 Analisis Sentimen 6](#_Toc503958700)

[6.4 *Text Preprocessing* 7](#_Toc503958701)

[6.5 Seleksi Fitur *Chi-square* 7](#_Toc503958702)

[*6.6* *Boolean Multinomial Naïve Bayes* 8](#_Toc503958703)

[7. Metodologi Penelitian 10](#_Toc503958704)

[7.1 *Fishbone* Penelitian 11](#_Toc503958705)

[7.2 Alur Metodologi Penelitian 11](#_Toc503958706)

[7.3 Pengumpulan data 12](#_Toc503958707)

[7.4 *Preproccessing* 13](#_Toc503958708)

[7.5 Metode yang digunakan 16](#_Toc503958709)

[7.6 Implementasi 19](#_Toc503958710)

[7.7 Pengujian dan Evaluasi 19](#_Toc503958711)

[8. Jadwal Pelaksanaan Penelitian 20](#_Toc503958712)

[DAFTAR PUSTAKA 21](#_Toc503958713)

# **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.

## **Tinjauan Pustaka**

Pada bagian ini akan dijelaskan teori-teori dan tinjauan studi yang dijadikan landasan dalam merancang dan membangun sistem analisis sentimen menggunakan metode *Boolean Multinomial Naïve Bayes.*

### **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 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

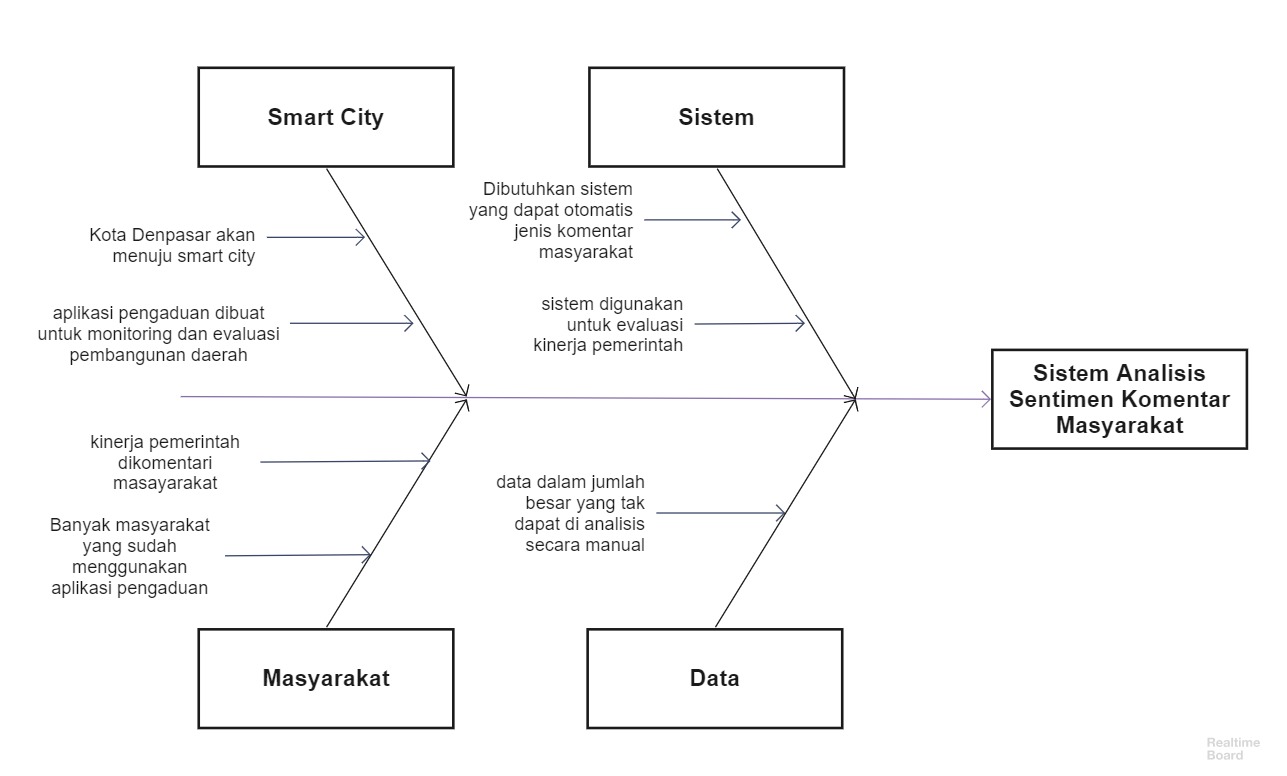
*V* : Jumlah seluruh *term* pada data *training*

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.

## **Metodologi Penelitian**

Pada bagian metodologi penelitian akan dijelaskan mengenai gambaran langkah-langkah yang akan dilakukan dalam menjalankan penelitian ini, langkah-langkah tersebut meliputi *fishbone* penelitian, alur metodologi penelitian, pengumpulan data, pengolahan data awal, metode yang digunakan, dan evaluasi.

### ***Fishbone* Penelitian**

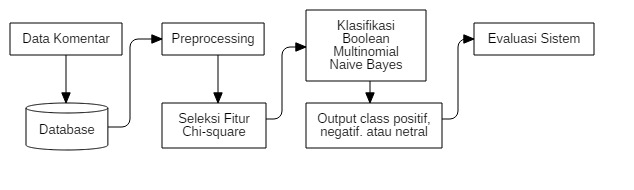


**Gambar 1 Fishbone Penelitian**

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 Metodologi Penelitian**

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 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 contoh data komentar positif, negatif, dan netral.

**Tabel 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 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 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 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 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 6 Contoh Stemming**

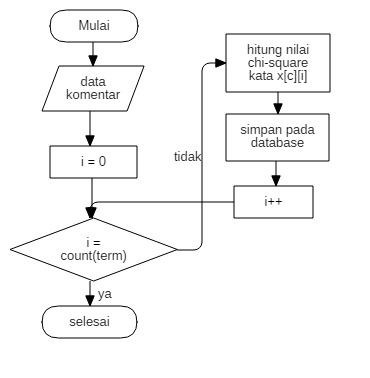
|  |  |
| --- | --- |
| **Komentar** | **Hasil** |
| terima  kasih  kondisi  jalan  sudah  bersih  sudah  bisa  dilalui  kendaraan | terima  kasih  kondisi  jalan  sudah  bersih  sudah  bisa  lalu  kendara |

### **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 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.

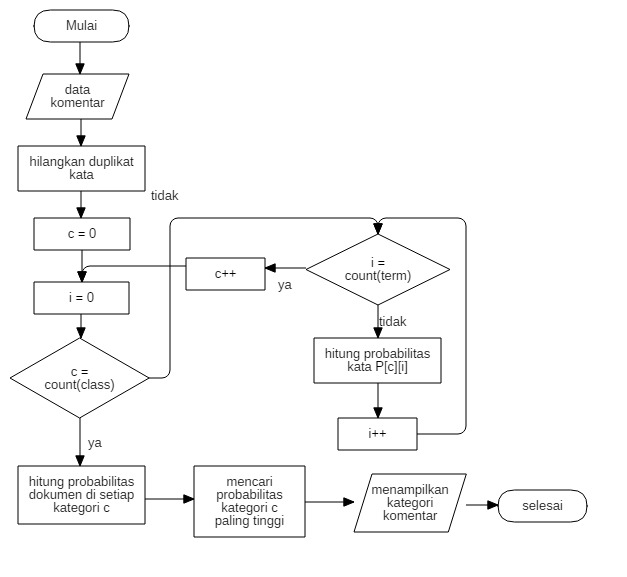
Berikut merupakan contoh tabel pengetahuan hasil seleksi fitur *Chi-square* yang diambil dari penelitian Juen Ling dkk.(2014).

**Tabel 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 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,* rasio jumlah penggunaan fitur mana yang menghasilkan akurasi yang paling tinggi. Contoh tabel evaluasi seleksi fitur *Chi-square* adalah sebagai berikut.

**Tabel 8 Contoh tabel evaluasi Chi-square**

|  |  |  |
| --- | --- | --- |
| **Rasio Fitur (%)** | **Jumlah Klasifikasi Benar** | **Akurasi (%)** |
| 25 |  |  |
| 50 |  |  |
| 75 |  |  |
| 100 |  |  |

## **Jadwal Pelaksanaan Penelitian**

Berikut merupakan jadwal pelaksanaan penelitian yang akan dilakukan.

**Tabel 9 Jadwal Pelaksanaan Penelitian**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Kegiatan | Bulan | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | Studi literatur |  |  |  |  |  |  |
| 2 | Analisis kebutuhan sistem |  |  |  |  |  |  |
| 3 | Desain sistem |  |  |  |  |  |  |
| 4 | Implementasi desain sistem |  |  |  |  |  |  |
| 5 | Pengujian sistem |  |  |  |  |  |  |
| 6 | Penyusunan laporan |  |  |  |  |  |  |

## **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.