

UNIVERSITAS GUNADARMA

FAKULTAS TEKNOLOGI INDUSTRI



Penulisan Ilmiah

**ANALISIS SENTIMEN TERHADAP ULASAN TEKS BAHASA
INDONESIA PADA SITUS PEMESANAN HOTEL ONLINE
PEGIPEGI.COM DENGAN KLASIFIKASI NAÏVE BAYES**

Nama : Dida Adams A
NPM : 53414056
Fakultas : Teknologi Industri
Jurusan : Teknik Informatika
Pembimbing : Dr. Miftah Andriansyah, SSi., MMSI

Diajukan Guna Melengkapi Sebagian Syarat
Dalam Mencapai Setara Sarjana Muda

DEPOK

2017

Pernyataan Orisinalitas dan Publikasi

Saya yang bertanda tangan di bawah ini:

Nama : Dida Adams A
NPM : 53414056
Judul PI : Analisis Sentimen Terhadap Ulasan Teks
Bahasa Indonesia pada Situs Pemesanan
Hotel Online Pegipegi.com dengan Klasifikasi
Naïve Bayes
Tanggal Sidang :
Tanggal Lulus :

Menyatakan bahwa tulisan di atas merupakan hasil karya saya sendiri dan dapat dipublikasikan sepenuhnya oleh Universitas Gunadarma. Segala kutipan dalam bentuk apapun telah mengikuti kaidah dan etika yang berlaku. Semua hak cipta dari logo serta produk yang disebut dalam buku ini adalah milik masing-masing pemegang haknya, kecuali disebutkan lain. Mengenai isi dan tulisan merupakan tanggung jawab Penulis, bukan Universitas Gunadarma.

Demikianlah pernyataan ini dibuat dengan sebenarnya dan dengan penuh kesadaran.

Depok, 28 Agustus 2017

(Dida Adams A)

Lembar Pengesahan

Judul PI : Analisis Sentimen Terhadap Ulasan Teks
Bahasa Indonesia pada Situs Pemesanan
Hotel Online Pegipegi.com dengan Klasifikasi
Naïve Bayes

Nama : Dida Adams A

NPM : 53414056

Tanggal Sidang :

Tanggal Lulus :

MENGETAHUI

Pembimbing

Koordinator PI

(Dr. Miftah Andriansyah, SSi., MMSI)

(Meilani B. Siregar, S.Kom., MMSI)

Ketua Jurusan

(Prof. Dr. -Ing. Adang Suhendra, S.Si., S.Kom., M.Sc)

Abstraksi

Dida Adams A. 53414056

ANALISIS SENTIMEN TERHADAP ULASAN TEKS BAHASA INDONESIA PADA SITUS PEMESANAN HOTEL ONLINE PEGIPEGI.COM DENGAN KLASIFIKASI NAÏVE BAYES.

Penulisan Ilmiah, Fakultas Teknologi Industri, Jurusan Teknik Informatika, Universitas Gunadarma, 2017.

Kata Kunci : Analisis Sentimen, Ulasan, Bahasa Indonesia, Klasifikasi, Naive Bayes

(xiii + 54 + lampiran)

Pada situs pemesanan hotel *online* permasalahan terletak pada pemberian penilaian yang kurang memuaskan menurut calon pelanggan. Penulisan ini menggunakan teknik analisis sentimen untuk menentukan klasifikasi dokumen ulasan termasuk dalam sentimen positif atau sentimen negatif. Analisis sentimen merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan suatu informasi sentimen yang terkandung dalam suatu kalimat opini. Data yang digunakan diambil dari situs pemesanan hotel *online* Pegipegi.com pada 5 hotel di satu wilayah di bulan Mei-Juli tahun 2017, sebanyak 156 dokumen ulasan berbahasa Indonesia. Penulisan ini menggunakan metode *Naïve Bayes Classification (NBC)*, dengan tahapan dimulai dari proses pengambilan data, pengolahan data *training*, dan pembuatan aplikasi. Bahasa pemrograman yang digunakan yaitu Python 2.7.13 dan aplikasi Rapidminer 6.3.

Dari pengujian hasil total yang didapatkan menunjukkan nilai klasifikasi dokumen ulasan sentimen positif diatas 72% dan nilai klasifikasi dokumen ulasan sentimen negatif dibawah 28%.

Daftar Pustaka (1987-2017)

Kata Pengantar

Puji syukur kehadirat Allah SWT, karena atas rahmat dan hidayah-Nya penulis dapat menyelesaikan penulisan ilmiah ini dengan judul “**Analisis Sentimen Terhadap Ulasan Teks Bahasa Indonesia pada Situs Pemesanan Hotel Online Pegipegi.com dengan Klasifikasi Naïve Bayes**”. Penulisan ilmiah ini dibuat untuk melengkapi salah satu syarat mencapai jenjang setara sarjana muda pada program pendidikan di Universitas Gunadarma.

Keberhasilan penulisa dalam menyelesaikan penulisan ilmiah ini tidak lepas dari bantuan dan dukungan dari berbagai pihak, walaupun banyak hambatan dan kesulitan yang dihadapi, namun hal tersebut menjadi pelajaran dan pengalaman yang berharga yang berkesan bagi penulis. Oleh sebab itu, dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Ibu Prof. Dr. E.S. Margianti, S.E., M.M, selaku Rektor Universitas Gunadarma.
2. Bapak Prof. Dr. Ir. Bambang Suryawan, M.T, selaku Dekan Fakultas Teknologi Industri, Universitas Gunadarma.
3. Bapak Prof. Dr. -Ing. Adang Suhendra, S.Si., S.Kom., M.Sc, selaku Ketua Jurusan Teknik Informatika, Universitas Gunadarma.
4. Ibu Meilani B. Siregar, S.Kom., MMSI, selaku Koordinator PI Jurusan Teknik Informatika.
5. Bapak Dr. Miftah Andriansyah, SSi., MMSI, Selaku Dosen Pembimbing yang telah banyak memberikan waktu, tenaga, pikiran dan arahan untuk membantu penulis dalam menyelesaikan Penulisan Ilmiah ini..
6. Keluarga yang selalu memberikan doa, dukungan, materil serta kepercayaan kepada penulis untuk dapat menyelesaian Penulisan Ilmliah ini.

7. Yusi Nur Awalia selaku teman seperjuangan semasa lulus SMA hingga sekarang yang selalu menemani, membantu, serta memberikan dukungan selama penyusunan Penulisan Ilmiah ini.
8. Keluarga Laboratorium Manajemen Dasar terutama angkatan 28 dan 29 yang telah memberikan doa dan dukungan sehingga dapat memudahkan penulis dalam penyusunan Penulisan Ilmiah ini.
9. Teman-teman dari 3IA13 terutama teman satu kelompok bimbingan yang telah berjuang bersama hingga Penulisan Ilmiah ini dapat terselesaikan.
10. Seluruh rekan seperjuangan di Universitas Gunadarma yang telah banyak membantu penulis.
11. Semua pihak yang tidak tersebutkan yang telah membantu penyelesaian Tugas Akhir ini, penulis ucapkan juga terima kasih atas segala bantuan dan sarannya.

Sebagai manusia biasa yang tak luput dari kesalahan, maka penulis meminta maaf atas segala kekurangan dan keterbatasan dalam penyusunan penulisan ilmiah ini.

Akhir kata, hanya kepada Tuhan jualah segalanya dikembalikan dan penulis sadari bahwa penulisan ini masih jauh dari sempurna, disebabkan karena berbagai keterbatasan yang penulis miliki. Untuk itu penulis mengharapkan kritik dan saran yang bersifat membangun untuk menjadi perbaikan di masa yang akan datang.

Semoga apa yang ada pada penulisan ilmiah ini dapat bermanfaat bagi kita semua. Amin.

Depok, Agustus 2017

(Dida Adams A)

Daftar Isi

Halaman Judul	i
Lembar Orisinalitas	ii
Lembar Pengesahan	iii
Abstraksi	iv
Kata Pengantar	v
Daftar Isi	ix
Daftar Gambar	xi
Daftar Tabel	xii
Daftar Lampiran	xiii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Batasan Masalah	2
1.3 Tujuan Penulisan	2
1.4 Metode Penelitian	2
1.5 Sistematika Penulisan	3
2 LANDASAN TEORI	4
2.1 Analisis Sentimen	4
2.2 Kecerdasan Buatan (AI)	5
2.3 <i>Machine Learning</i>	6
2.3.1 <i>Text mining</i>	6
2.3.2 <i>Text Pre-processing</i>	7
2.3.2.1 <i>Stopwords</i>	7

2.3.2.2	Standarisasi Bahasa Prokem	7
2.3.2.3	<i>Stemming</i>	8
2.3.2.4	<i>Tokenizing</i>	8
2.4	<i>Web Scraping</i>	8
2.4.1	<i>Scraping</i> Konten	8
2.4.2	<i>Feature Selection</i>	9
2.4.3	Dokumen Bahasa Indonesia	9
2.4.3.1	Ciri-Ciri Keilmuan Linguistik	9
2.5	<i>E-business</i>	9
2.5.1	Pelaku <i>E-business</i>	10
2.5.2	Alat/Media/Sumber Daya yang Digunakan	10
2.5.3	Objek/Kegiatan Sasaran	10
2.5.4	Tujuan	10
2.5.5	Keuntungan	10
2.5.6	Penyebab Kegagalan <i>E-business</i> Secara Umum	11
2.5.7	Kiat Membangun <i>E-business</i>	11
2.5.8	Konsep <i>E-business</i>	12
2.5.9	Situs Pemesanan Hotel	12
2.6	<i>Naïve Bayes Classifier</i>	14
2.7	Terminologi Aplikasi dalam Pengolahan	16
2.7.1	Bahasa Pemrograman <i>Python</i> (versi : 2.7.13)	16
2.7.1.1	Fitur-Fitur <i>Python</i>	16
2.7.1.2	Hubungan <i>Python</i> dengan Analisis Sentimen	17
2.7.2	Platform <i>Rapidminer</i> (versi : 6.3)	17
2.7.2.1	User Interface	18
2.7.2.2	Dukungan Basis Data	18
2.7.2.3	Dukungan Import File Beragam Format	19
3	ANALISIS DAN PEMBAHASAN	20
3.1	Gambaran Umum Analisis	20
3.2	Flowchart	20
3.2.1	Flowchart <i>Stemming</i>	20
3.2.2	Flowchart <i>Web Scraping</i>	21
3.2.3	Flowchart Arsitektur Sistem Analisis Sentimen	22
3.3	Rancangan Sistem	22
3.3.1	Metode Pengumpulan Data	22
3.3.2	Metode Pemrosesan Awal Data	23
3.3.3	Model yang Diusulkan	23

3.4	Eksperimen dan Pengujian	23
3.4.1	Text <i>Pre-processing</i>	25
3.4.1.1	Filter <i>Tokenizing</i>	25
3.4.1.2	<i>Stemming</i>	26
3.4.1.3	Filter <i>Stopwords</i> (Indonesia)	27
3.5	Rancangan Kamus Data	28
3.5.1	Struktur Kamus Data	28
3.5.1.1	Tabel Kamus Data KBBI	28
3.5.1.2	Tabel Kamus Data Formal	28
3.5.1.3	Tabel Kamus Data Non-formal	29
3.6	Rancangan Sistem	29
3.6.1	Library <i>Python</i>	30
3.6.2	Proses <i>Scraping</i>	30
3.6.3	Rancangan <i>Tokenizing</i>	32
3.6.4	Rancangan <i>Stemming</i>	32
3.6.5	Rancangan Filter (<i>Stopwords</i>) dan <i>Topic Model</i>	33
3.6.6	Rancangan Kamus Data	35
3.7	Pengujian	36
3.7.1	Pengujian Aplikasi	36
3.7.2	Pembuatan Aplikasi dengan <i>Python 2.7.13</i>	46
4	PENUTUP	51
4.1	KESIMPULAN	51
4.2	SARAN	51
DAFTAR PUSTAKA		53
LAMPIRAN		54

Daftar Gambar

2.1	Penetrasi & Perilaku Pengguna Internet di Indonesia	5
2.2	Halaman Utama Situs PegiPegi.com	13
2.3	Halaman Ulasan Hotel di PegiPegi.com	13
2.4	Halaman Pencarian Hotel di PegiPegi.com	14
2.5	Halaman Data Pesanan Hotel di PegiPegi.com	14
2.6	Aplikasi <i>Python 2.7.13</i>	17
2.7	Platform <i>Rapidminer 6.3</i>	18
2.8	Fitur <i>Rapidminer 6.3</i>	19
3.1	Flowchart <i>Stemming Kata Baku</i>	21
3.2	Flowchart <i>Web Scraping</i>	21
3.3	Flowchart Arsitektur Sistem Analisis Sentimen	22
3.4	Contoh Sentimen Positif Berbahasa Indonesia	22
3.5	Contoh Sentimen Negatif Berbahasa Indonesia	22
3.6	Proses Analisis Teks dengan <i>Rapidminer 6.3</i>	25
3.7	Package pada Library <i>Python 2.7.13</i>	30
3.8	Package pada NLTK Library <i>Python</i>	30
3.9	Output <i>Scraping</i>	31
3.10	Output <i>Tokenizing</i>	32
3.11	Output <i>Stemming</i>	33
3.12	Output <i>Stopwords</i>	34
3.13	Output <i>Topic Model</i>	35
3.14	Tampilan Mulai Project Baru	37
3.15	Operator <i>Read Excel</i>	37
3.16	<i>Icon Import Configuration Wizard</i>	37
3.17	Penambahan Operator <i>Validation</i>	38
3.18	Penambahan Operator <i>Select Attributes</i>	38
3.19	Konfigurasi Operator <i>Select Attribute</i>	39
3.20	Proses Relasi	39

3.21 Import Algoritma Naïve Bayes	40
3.22 Penambahan Operator <i>Apply Model & Performance</i>	40
3.23 Statistik Penjabaran Atribut	42
3.24 <i>ExampleSet</i> Data Rapidminer	42
3.25 <i>Performance Vector (Confusion Matrix)</i>	43
3.26 Diagram <i>Scatter</i> dengan Sumbu Kata dan sumbu Sentimen . .	44
3.27 Diagram <i>Pie</i> Sebelum Dikalikan dengan Frekuensi	44
3.28 Diagram <i>Pie</i> Sesudah Dikalikan dengan Frekuensi	45
3.29 Diagram <i>Pie</i> Pembagian Kata Setiap Hotel	45
3.30 Proses Download Library Image	47
3.31 Tampilan Awal Aplikasi Sederhana	47
3.32 Tampilan Output Aplikasi Hotel Horison	48
3.33 Tampilan Output Aplikasi Hotel Harris	48
3.34 Tampilan Output Aplikasi Hotel Red Planet	49
3.35 Tampilan Output Aplikasi Hotel Fave	49
3.36 Tampilan Output Aplikasi Hotel Santika Premiere	50

Daftar Tabel

3.1	Contoh Perhitungan Sebagian Data Hasil <i>Pre-processing</i>	24
3.2	Ilustrasi <i>Tokenizing</i>	25
3.3	Ilustrasi <i>Stemming</i>	27
3.4	Ilustrasi <i>Stopwords</i>	27
3.5	Kamus Data KBBI	28
3.6	Kamus Data Formal (Sastrawi)	29
3.7	Tabel Kamus Data Non-formal	29
3.8	Sebagian Kamus Data Hasil <i>Stopwords & Topic Model</i>	35
3.9	POS-Tag	36
3.10	Tipe Atribut Data Import	38
3.11	Perhitungan Naive Bayes pada Ulasan Hotel Horison	41
3.12	Perhitungan <i>Naive Bayes</i> pada Ulasan Hotel Harris	41
3.13	Perhitungan <i>Naive Bayes</i> pada Ulasan Hotel Red Planet	41
3.14	Perhitungan <i>Naive Bayes</i> pada Ulasan Hotel Fave	41
3.15	Perhitungan <i>Naive Bayes</i> pada Ulasan Hotel Santika Priemere	41
3.16	Akurasi Total Perhitungan <i>Naïve Bayes</i> 63,24%	41

Daftar Lampiran

LAMPIRAN LISTING PROGRAM	54
LAMPIRAN OUTPUT PROGRAM	60

Bab 1

PENDAHULUAN

1.1 Latar Belakang

Situs pemesanan hotel online termasuk dalam kemajuan teknologi dan bisnis yang berkembang cepat. Banyaknya ulasan-ulasan yang memberikan opini terhadap penyedia jasa yang meliputi opini positif maupun opini negatif. Dalam tiap produk yang dijual oleh penjual, terdapat kolom ulasan dari konsumen. Sebuah hotel dapat memiliki puluhan, ratusan, bahkan ribuan ulasan. Dalam membaca keseluruhan ulasan, tentu akan memakan waktu. Oleh karena itu dibutuhkan suatu analisis sentimen dalam menyimpulkan berbagai ulasan tersebut.

Berdasarkan uraian di atas, maka penulis membuat penelitian analisis sentimen terhadap situs pemesanan hotel online Pegipegi.com karena situs tersebut masuk urutan ke 2 berdasarkan situs pemesanan hotel online terpercaya. Hasil yang diharapkan berfokus pada sentimen positif atau negatif dari ulasan hotel tersebut secara umum. Perhitungan dilakukan berdasarkan probabilitas kemunculan kata bernilai sentimen positif dan kata bernilai sentimen negatif. Analisis sentimen tersebut diharapkan dapat mengklasifikasikan dokumen ulasan menjadi sentimen positif dan sentimen negatif.

Kamus data yang digunakan adalah Kamus Besar Bahasa Indonesia (KBBI) dan tambahan kata nonformal menurut penulis yang umum digunakan dalam pemberian ulasan. Penulis menggunakan KBBI untuk menganalisis bahasa formal dan tambahan kata nonformal secara manual pada dokumen ulasan yang ada. Metode yang digunakan yaitu klasifikasi *Naïve Bayes* dengan Bahasa pemrograman *Python 2.7.13* dan aplikasi *Rapidminer 6.3*.

1.2 Batasan Masalah

Adapun ruang lingkup penulisan ilmiah ini dibatasi beberapa batasan masalah, yaitu:

1. Pengambilan sampel yang digunakan yaitu ulasan dari 5 hotel yang terdapat pada situs pemesanan hotel online Pegipegi.com sebanyak 156 dokumen pada bulan Mei-Juli 2017.
2. Proses *Stopword Removal* menggunakan kamus data untuk kata-kata formal dalam Kamus Besar Bahasa Indonesia (KBBI) dengan tambahan beberapa kata nonformal yang dibuat oleh penulis secara manual berdasarkan kata yang umum digunakan dalam pemberian ulasan.
3. Metode yang digunakan adalah klasifikasi Naive Bayes.
4. Hasil analisis yaitu persentase sentimen positif dan negatif ulasan secara umum pada tiap hotel.

1.3 Tujuan Penulisan

Tujuan dari penulisan ilmiah ini adalah untuk mengklasifikasikan dokumen ulasan kedalam sentimen positif dan sentimen negatif.

1.4 Metode Penelitian

Metode penelitian terbagi dalam beberapa tahap, yaitu:

1. Tahap analisis Masalah

Pada tahap ini, penulis melakukan pencarian dan pengumpulan informasi mengenai analisis sentimen dan *Text Mining* dengan melakukan studi pustaka dari beberapa buku, serta mengumpulkan informasi lainnya yang diperlukan dari internet.

2. Tahap Pengambilan dan Penyaringan Data

Pada tahap ini, dilakukan proses *Web Scraping* dan *Text Preprocessing* untuk mendapatkan data yang dapat diolah. Adapun proses *Text Processing* sendiri terdiri dari tahap *Parsing*, *Lexical Analysis*, *Stemming*, dan *Stopword Removal*.

3. Tahap Analisis dan Perhitungan

Pada tahap ini, dilakukan perancangan sistem dalam pengolahan data menggunakan algoritma Naive Bayes dengan aplikasi Rapidminer 6.3. Dalam proses analisis sentimen, tahap pertama adalah pelatihan terhadap data latih. Sedangkan tahap kedua adalah proses klasifikasi data uji.

4. Tahap Pembuatan Aplikasi

Pada tahap ini, dibuat aplikasi untuk menampilkan *output* antarmuka dalam menerapkan hasil penelitian.

1.5 Sistematika Penulisan

Sistematika penulisan ini terdiri dari 4 bab diantaranya :

BAB I PENDAHULUAN

Bab ini berisikan latar belakang masalah, batasan penulisan, metode penulisan dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisikan teori-teori yang dapat menunjang tulisan yang berhubungan dengan *Machine Learning*.

BAB III analisis DAN PEMBAHASAN

Bab ini berisikan tentang pembahasan masalah, yang menjelaskan langkah-langkah dimulai dari perancangan hingga pemakaian program *Machine Learning*.

BAB IV PENUTUP

Bab ini berisikan tentang kesimpulan dan saran-saran yang merupakan tindak lanjut dari pembahasan yang bermanfaat bagi penulis selanjutnya.

Bab 2

LANDASAN TEORI

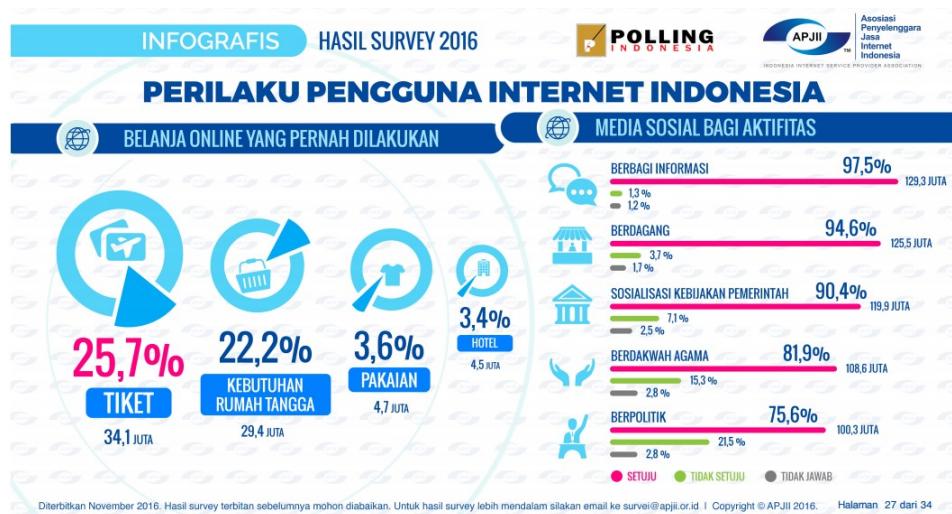
2.1 Analisis Sentimen

Analisis sentimen atau opinion *mining* merupakan proses memahami, mengekstrak, dan mengolah data teksual secara otomatis untuk mendapatkan suatu informasi sentimen yang terkandung dalam suatu kalimat opini. Secara umum, analisis sentimen bertujuan untuk menentukan sikap pembicara atau penulis terhadap suatu topik atau keseluruhan polaritas kontekstual pada suatu dokumen. Sikap dapat berupa penilaian atau evaluasi, sisi emosional penulis pada saat menulis atau efek komunikasi emosional yang penulis inginkan terhadap pembacanya. Situs Website sebagai salah satu jejaring sosial yang interaktif memungkinkan penggunanya untuk mengkritisir suatu barang maupun sebuah fasilitas pelayanan secara langsung setelah menggunakannya. Masyarakat yang semula membutuhkan waktu lama untuk menyampaikan aspirasinya kini dapat melakukannya dengan mudah berkat kehadiran teknologi ini.

Analisis sentimen dapat digunakan dalam berbagai kemungkinan domain, dari produk konsumen, jasa kesehatan, jasa keuangan, peristiwa sosial dan politik pada pemilu. Kecendrungan penelitian tentang analisis sentimen berfokus pada pendapat yang menyatakan atau menyiratkan suatu sentimen positif atau negatif. Pendapat mewakili hampir semua aktivitas manusia, karena pendapat dapat mempengaruhi terhadap perilaku seseorang. Setiap kali kita perlu membuat keputusan, kita ingin tahu pendapat orang lain. Dalam dunia nyata, bisnis dan organisasi selalu ingin melihat opini publik tentang suatu produk atau jasa [15]. Salah satu bisnis yang paling banyak mengadopsi manfaat dari internet atau sistem *online* adalah industri hotel [7].

Peran internet sebagai media reservasi *online* pada beberapa industry akomodasi telah menunjukkan tren meningkatkan kinerja pemasaran, secara umum internet telah menjadi sumber utama reservasi pada hotel-hotel [1]. Kondisi ini menyebabkan para pembisnis *online* harus memberikan pelayanan terbaiknya, salah satunya dengan memperhatikan kualitas *Web* (*Website quality*), sebab kualitas *Web* akan mempengaruhi kepuasan konsumen (customer satisfaction) terhadap suatu perusahaan. [19].

Berikut statistik perilaku pengguna internet di Indonesia :



Gambar 2.1: Penetrasi & Perilaku Pengguna Internet di Indonesia

Analisis sentimen pada suatu kalimat menggambarkan bagian pertimbangan penilaian terhadap entitas atau kejadian tertentu [17]. Entitas adalah produk, layanan, topik, isu, orang, organisasi, atau peristiwa yang menjadi objek target pada kalimat sentimen. Suatu hal yang penting, untuk memahami permasalahan dalam membedakan kalimat sentimen [15].

2.2 Kecerdasan Buatan (AI)

Kecerdasan Buatan atau Artificial Intelligence (AI) merupakan salah satu bagian ilmu komputer yang membuat agar mesin (komputer) dapat melakukan pekerjaan seperti dan sebaik yang dilakukan manusia. Pada awal diciptakannya, komputer hanya difungsikan sebagai alat hitung saja. Namun seiring dengan perkembangan jaman, maka peran komputer semakin mendominasi kehidupan umat manusia. Komputer tidak lagi hanya digunakan sebagai alat hitung, lebih dari itu, komputer diharapkan untuk dapat diberdayakan untuk

mengerjakan sesuatu yang bisa dikerjakan oleh manusia. Bidang yang menggunakan kecerdasan buatan pada penulisan ini adalah penggunaan *Machine Learning*.

2.3 *Machine Learning*

Machine Learning (Pembelajaran Mesin) Mesin disini adalah mesin dalam pengertian lebih mendekati ‘sistem’ bukan mesin mekanik. Istilah pembelajaran pertama kali muncul dalam disiplin ilmu kecerdasan buatan. Pembelajaran berarti menambah pengetahuan, memahami dengan belajar, mengikuti perintah. Pembelajaran mesin merupakan salah satu cabang dari kecerdasan buatan yang membahas mengenai pembangunan sistem yang didapat berdasarkan pada pembelajaran data, atau sebuah studi yang mempelajari cara untuk memprogram sebuah komputer untuk belajar. Inti dari pembelajaran mesin adalah representasi dan generalisasi. Pada tahun 1959, Arthur Samuel mendefinisikan bahwa pembelajaran mesin adalah bidang studi yang memberikan kemampuan untuk belajar tanpa diprogram secara eksplisit. Kemampuan belajar yang menjadi dominan ditentukan oleh kemampuan perangkat lunak atau algoritmanya. Implementasi kemampuan belajar dapat dicapai dengan berbagai teknik, ada yang menggunakan kaidah (aturan), ada yang menggunakan statistika, ada yang menggunakan pendekatan fisiologi yaitu sistem saraf manusia atau disebut dengan ANN (*Artificial Neural Network*) atau jaringan saraf tiruan. Pembelajaran mesin dapat berfungsi untuk beradaptasi dengan suatu keadaan yang baru, serta untuk mendeteksi dan memperkirakan suatu pola.

2.3.1 *Text mining*

Text mining, yang juga disebut text data *mining* (TDM) atau knowledge discovery in the text. *Text mining* merupakan teknik yang digunakan untuk menangani masalah klasifikasi, pengelompokan, ekstraksi informasi, dan pencarian informasi. Untuk data penginapan, *text mining* dilakukan pada *unstructured text* untuk mengekstraksi teks yang tidak tersusun menjadi sebuah informasi yang berguna [5].

Text mining, mengacu pada proses mengambil informasi berkualitas tinggi dari teks. Informasi berkualitas tinggi biasanya diperoleh melalui peramalan pola dan kecenderungan melalui sarana seperti pembelajaran pola statistik. *Text mining* biasanya melibatkan proses penataan teks input, menentuk-

an pola dalam data terstruktur, dan akhirnya mengevaluasi dan menginterpretasi *output*. 'Berkualitas tinggi' di bidang text *mining* biasanya mengacu ke beberapa kombinasi relevansi, kebaruan, dan *interestingness*. Proses text *mining* yang khas meliputi kategorisasi teks, text clustering, ekstraksi konsep/entitas, sentimen analysis, penyimpulan dokumen, dan pemodelan relasi entitas [11]. Tahap-tahap text *mining* secara umum adalah *text preprocessing* dan *feature selection* [9, 5].

2.3.2 Text Pre-processing

Pada tahap ini mencakup seluruh rutinitas, dan mempersiapkan data yang akan dilakukan analisis. Semakin banyak informasi yang dikenali sistem, maka semakin akuratlah sebuah text *mining* itu. Dalam text *Pre-processing* adalah mengubah semua karakter menjadi huruf kecil dan menguraikan kalimat menjadi kata dan menghilangkan karakter spesial seperti titik (.), koma (,), spasi, dan karakter yang tidak diperlukan. Data *Pre-processing* mengurangi dokumen teks input secara signifikan. Pada tahap ini dilakukan *Tokenizing*, *Stopwords*, dan *Stemming* [5].

2.3.2.1 Stopwords

Dapat diartikan sebagai menghilangkan karakter, tanda baca, serta kata-kata umum yang tidak memiliki makna atau informasi yang dibutuhkan. *Stopwords* umumnya digunakan dalam pengambilan informasi salah satu contohnya adalah mesin pencari Google. Pengurangan ukuran indeks dalam teks dengan penghilangan beberapa kata kerja, kata sifat, dan kata keterangan lainnya dapat dimasukkan ke dalam daftar *Stopwords*. *Stopwords* sering di temukan pada kata secara umumnya seperti contoh kata: "dan", "kamu", dll. Selain itu perlu di standarisasi emoticon untuk menangkap informasi tambahan [12].

2.3.2.2 Standarisasi Bahasa Prokem

Prokem adalah bahasa bahasa gaul yang sering dipakai oleh masyarakat Indonesia. Peneliti menambahkan fitur ini karena pada sosial media itu hampir semua kata itu berisi kata – kata gaul. Jika tidak dilakukan standarisasi kata slang maka banyak makna kata yang bisa hilang.

2.3.2.3 *Stemming*

Stemming akan dilakukan pada tahap ini untuk menghilangkan imbuhan sehingga siap untuk di analisis. Proses *stemming* adalah teknik yang digunakan untuk mengurangi kata-kata untuk bentuk kata dasar dengan menghapus derivasi dan afiks menurut Wahiba Ben. Alasan dilakukan *stemming* adalah untuk membuat data menjadi mudah untuk dianalisis. [14].

2.3.2.4 *Tokenizing*

Tokenisasi adalah proses untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi token-token/bagian-bagian tertentu. Sebagai contoh, tokensisasi dari kalimat “Hotel ini sangat bagus sekali” menghasilkan enam token, yaitu : “Hotel”, ”ini”, ”sangat”, ”bagus”, ”sekali”. Dengan mengambil dalih spasi dan tanda baca ([id.wikipedia.org/wiki TokenName](https://id.wikipedia.org/wiki	TokenName)).

2.4 *Web Scraping*

Secara umum, *Web Scraping* adalah pengambilan sebuah dokumen semi-terstruktur dari internet, umumnya berupa halaman-halaman *Web* dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut. Istilah gampangnya yaitu pengambilan konten dari situs satu ke situs lain.

Untuk penulisan ini *Web Scraping* mengacu pada definisi *Web Scraping* adalah teknik yang digunakan untuk mengambil isi sebuah halaman situs PegiPegi hotel *online* secara spesifik, yaitu ulasan dari ulasan seseorang terhadap suatu pelayanan atau sejenisnya. Data yang digunakan teks berbahasa Indonesia diambil dari situs pemesanan hotel *online* yaitu, pegipegi.com dengan menggunakan bahasa pemrograman *Python* 2.7.13

2.4.1 *Scraping* Konten

Ada dua cara untuk melakukan *Scraping* konten, yaitu :

1. Manual – ini adalah cara yang paling sederhana, dengan copy paste dari situs lain.
2. Otomatis – ini membutuhkan *tools* atau program (biasa disebut “bot”) yang dibuat untuk *Python* atau sebuah program yang sistematis menelusuri *World Wide Web* untuk membuat indeks data dan mengambil semua konten yang sesuai dengan parameter tertentu.

2.4.2 *Feature Selection*

Feature selection adalah suatu kegiatan yang umumnya bisa dilakukan secara berdampingan dengan text *Pre-processing*. Tahap seleksi fitur (*feature selection*) bertujuan untuk mengurangi dimensi dari suatu kumpulan teks, atau dengan kata lain menghapus kata-kata yang dianggap tidak penting atau tidak menggambarkan isi dokumen sehingga proses pengklasifikasian lebih efektif dan akurat [9, 5].

2.4.3 Dokumen Bahasa Indonesia

Pada penelitian ini penulis menggunakan dokumen bahasa Indonesia dengan pendekatan Linguistik Umum (*general linguistics*) yang merumuskan secara umum semua bahasa manusia yang bersifat alamiah itu. Linguistik umum memberikan gambaran umum tentang suatu bahasa sehingga menghasilkan teori bahasa yang bersangkutan dan ilmu yang mengambil bahasa sebagai objek kajiannya. Pada linguistik umum diberikan ciri umum bahasa manusia, diuraikan secara sederhana, umum, tepat dan objektif. Linguistik umum memberikan informasi umum mengenai teori, prosedur kerja, dan faham-faham yang berkembang dalam linguistik [16].

2.4.3.1 Ciri-Ciri Keilmuan Linguistik

Linguistik mempunyai 3 ciri yaitu:

1. Eksplisit Adalah jelas, menyeluruh, tidak mempunyai dua makna, pasti / konsisten Contoh: Men + sikat = menyikat Men + sapu = menyapu.
2. Sistematis Adalah berpola dan beraturan.
3. Objektif Adalah sesuai keadaan atau apa adanya.

2.5 *E-business*

E-business adalah praktek pelaksanaan dan pengelolaan proses bisnis utama seperti perancangan produk, pengelolaan pasokan bahan baku, manufaktur, penjualan, pemenuhan pesanan, dan penyediaan servis melalui penggunaan teknologi komunikasi, komputer, dan data yang telah terkomputerisasi [3]. Berdasarkan beberapa definisi e-bisnis yang dikemukakan di atas, kita dapat menggabungkannya ke dalam suatu definisi *E-business* yang utuh dengan

melihat kesamaan dari setiap definisi tersebut dan menggabungkannya. Kesamaan tersebut dapat kita lihat dari beberapa sudut pandang, yaitu pelaku *E-business*, alat atau media atau sumber daya yang digunakan, objek atau kegiatan yang menjadi sasaran, tujuannya, dan keuntungan yang diberikan. Hasilnya sebagai berikut:

2.5.1 Pelaku *E-business*

Organisasi, konsumen, perusahaan/pengembang, supplier, pekerja, rekan bisnis.

2.5.2 Alat/Media/Sumber Daya yang Digunakan

1. Teknologi informasi dan komunikasi.
2. Komputer, data yang telah terkomputerisasi.
3. Internet.

2.5.3 Objek/Kegiatan Sasaran

1. Kegiatan bisnis.
2. Proses bisnis utama.
3. Pembelian, penjualan, pelayanan, transaksi.
4. Operasi bisnis utama.

2.5.4 Tujuan

1. Koordinasi, Komunikasi, dan Pengelolaan organisasi.
2. Transformasi proses bisnis.
3. Sharing informasi.

2.5.5 Keuntungan

1. Pendekatan yang aman, fleksibel, dan terintegrasi.
2. Memberikan nilai bisnis yang berbeda.
3. Efisien.

4. Peningkatan produktivitas dan keutungan.

2.5.6 Penyebab Kegagalan *E-business* Secara Umum

1. Tidak ada komitmen yang utuh dari manajemen puncak.
2. Penerapan *E-business* tidak diikuti proses change management.
3. Tidak profesionalnya vendor teknologi informasi yang menjadi mitra bisnis
4. Buruknya infrastruktur komunikasi.
5. Tidak selarasnya strategi TI dengan strategi perusahaan.
6. Adanya masalah keamanan dalam bertransaksi.
7. Kurangnya dukungan finansial.
8. Belum adanya peraturan yang mendukung dan melindungi pihak-pihak yang bertransaksi (*cyberlaw*).
9. Menggunakan target jangka pendek sebagai pijakan investasi *E-business*.

2.5.7 Kiat Membangun *E-business*

1. Membenahi terlebih dahulu sistem pengelolaan sumber daya perusahaan secara terpadu.
2. Membuat perencanaan investasi teknologi secara mendetail dan komprehensif.
3. Menentukan arah investasi teknologi untuk menjawab kebutuhan jangka panjang.
4. Membentuk struktur organisasi yang fleksibel dan adaptif terhadap perubahan.
5. Melakukan kerjasama kondusif dengan berbagai mitra bisnis (vendor, pemasok barang, lembaga keuangan, dan lain sebagainya).

2.5.8 Konsep *E-business*

E-business memiliki karakteristik tujuan yang sama dengan bisnis secara konvensional, hanya saja *E-business* memiliki scope yang berbeda. Bisnis mengandalkan pertemuan antar pebisnis seperti halnya rapat ditempat khusus, atau sekedar untuk berkenalan dengan partner bisnis, sedangkan *E-business* mengandalkan media Internet sebagai sarana untuk memperoleh tujuannya. Menurut Turban, *E-business* atau bisnis elektronik merujuk pada definisi *e-commerce* yang lebih luas, tidak hanya pembelian dan penjualan barang serta jasa, tetapi juga pelayanan pelanggan, kolaborasi dengan mitra bisnis, e-learning, dan transaksi elektronik dalam perusahaan.

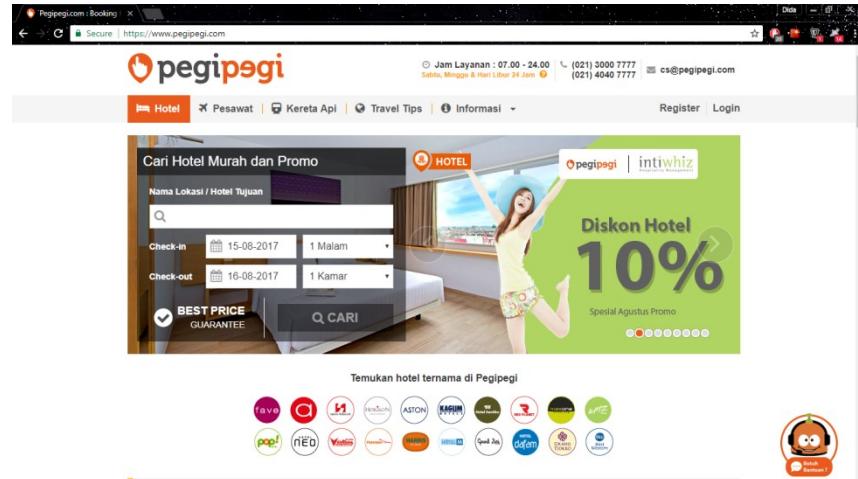
2.5.9 Situs Pemesanan Hotel

Berikut daftar urutan 5 situs pilihan untuk pemesanan kamar hotel terpercaya [10]:

1. Agoda.com
2. PegiPegi.com
3. Wego.co.id
4. Traveloka.com
5. booking.com

Pada penulisan kali ini penulis akan mengambil beberapa sample untuk membahas sebuah situs *booking hotel online* yang terbilang cukup populer untuk ukuran pendatang baru di dunia maya, yaitu PegiPegi.com.

Tampilan awal situs PegiPegi.com:



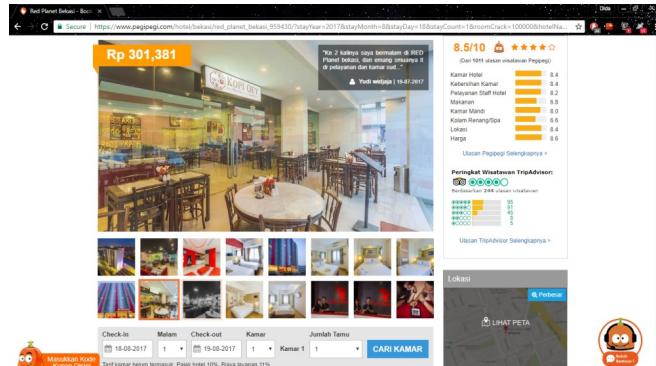
Gambar 2.2: Halaman Utama Situs PegiPegi.com

Tampilan ulasan masyarakat terhadap beberapa hotel:

10	Good service	
Shella gabriella / Wanita / 20 - 24 tahun / Perjalanan bersama Pasangan		
11-07-2017	Kamar Hotel	★★★★★ 10 Kamar Mandi
	Kebersihan Kamar	★★★★★ 10 Kolam Renang/Spa
	Pelayanan Staff Hotel	★★★★★ 10 Lokasi
	Makanan	★★★★★ 10 Harga
Over all baik		
10	Hotel Horison Ultima Bekasi	
Lukmanul Hakim / Pria / 20 - 24 tahun / Perjalanan bersama Teman		
10-07-2017	Kamar Hotel	★★★★★ 10 Kamar Mandi
	Kebersihan Kamar	★★★★★ 10 Kolam Renang/Spa
	Pelayanan Staff Hotel	★★★★★ 10 Lokasi
	Makanan	☆☆☆☆☆ 0 Harga
Hotel Horison Ultima bekasi sangat nyaman, selain letaknya yang strategis karena letaknya berada di jantung kota bekasi pelayanannya pun sangat baik dan hotel ini sangat recomended bagi pelancong-pelancong yang ingin mendapatkan kenyamanan yang lebih		
10	Hotel menyenangkan	
Aditya / Pria / 20 - 24 tahun / Perjalanan Keluarga		
30-06-2017	Kamar Hotel	★★★★★ 10 Kamar Mandi
	Kebersihan Kamar	★★★★★ 10 Kolam Renang/Spa
	Pelayanan Staff Hotel	★★★★★ 10 Lokasi
	Makanan	★★★★★ 10 Harga
Sangat baik dari semua segi baik pelayanan, kebersihan dll sehingga perjalanan dengan keluarga bisa sangat berkesan		

Gambar 2.3: Halaman Ulasan Hotel di PegiPegi.com

Tampilan ketika user mencari hotel yang diinginkan:



Gambar 2.4: Halaman Pencarian Hotel di PegiPegi.com

Berikut halaman untuk mengisi data saat Anda memesan kamar:

Gambar 2.5: Halaman Data Pesanan Hotel di PegiPegi.com

2.6 Naïve Bayes Classifier

Naïve Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam *Naïve Bayes* model yang digunakan adalah “model fitur independen” [18].

Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk *Machine Learning* dan data *mining*. Performa *Naïve Bayes* yang kompetitif dalam proses klasifikasi walaupun menggunakan asumsi keindependennan atribut (tidak ada kaitan antar atribut). Asumsi keindependennan atribut ini pada data sebenarnya jarang terjadi, namun walaupun asumsi keindependennan atribut tersebut dilanggar performa pengklasifikasian *Naïve Bayes* cukup tinggi, hal ini dibuktikan pada

berbagai penelitian empiris [20].

Dari definisi di atas dapat diambil kesimpulan bahwa *Naïve Bayes* adalah sebuah teknik klasifikasi probabilistik yang berdasarkan teorema bayes yang menggunakan asumsi keindependen atribut (tidak ada kaitan antar atribut) dalam proses pengklasifikasiannya.

Dalam sebuah aturan yang mudah, sebuah klasifikasi *Naïve Bayes* diasumsikan bahwa ada atau tidaknya ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Untuk contohnya, buah akan dianggap sebagai sebuah apel jika berwarna merah, berbentuk bulat dan berdiameter sekitar 6 cm. Walaupun jika ciri-ciri tersebut bergantung satu sama-lainnya, dalam Bayes hal tersebut tidak dipandang sehingga masing-masing fitur seolah tidak memiliki hubungan apapun. Berdasarkan ciri alami dari sebuah model probabilitas, klasifikasi *Naïve Bayes* bisa dibuat lebih efisien dalam bentuk pembelajaran. Dalam beberapa bentuk praktiknya, parameter untuk perhitungan model *Naïve Bayes* menggunakan metode *maximum likelihood*, atau kemiripan tertinggi.

Prediksi *Naïve Bayes* didasarkan pada teorema Bayes dengan formula untuk klasifikasi sebagai berikut [18]:

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \quad (2.1)$$

Sedangkan *Naïve Bayes* dengan fitur kontinu memiliki formula:

$$P(Y|X) = \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.2)$$

Keterangan:

$P(Y|X)$ = Probabilitas data dengan vektor X pada kelas Y.

$P(Y)$ = Probabilitas awal kelas Y

$P(Y) \prod_{i=1}^q P(X_i|Y)$ = Probabilitas independen kelas Y dari semua fitur dalam vektor X

μ = Mean atau nilai rata-rata dari atribut dengan fitur kontinu

σ = Deviasi standar

2.7 Terminologi Aplikasi dalam Pengolahan

2.7.1 Bahasa Pemrograman *Python* (versi : 2.7.13)

Python adalah bahasa pemrograman interpretatif multiguna. Tidak seperti bahasa lain yang susah untuk dibaca dan dipahami, *Python* lebih menekankan pada keterbacaan kode agar lebih mudah untuk memahami sintaks. Hal ini membuat *Python* sangat mudah dipelajari baik untuk pemula maupun untuk yang sudah menguasai bahasa pemrograman lain. Sampai saat ini *Python* masih dikembangkan oleh *Python perangkat lunak Foundation*. Bahasa *Python* mendukung hampir semua sistem operasi, bahkan untuk sistem operasi Linux, hampir semua distronya sudah menyertakan *Python* di dalamnya.

Dengan kode yang simpel dan mudah diimplementasikan, seorang programmer dapat lebih mengutamakan pengembangan aplikasi yang dibuat, bukan malah sibuk mencari *Syntax error* [21].

2.7.1.1 Fitur-Fitur *Python*

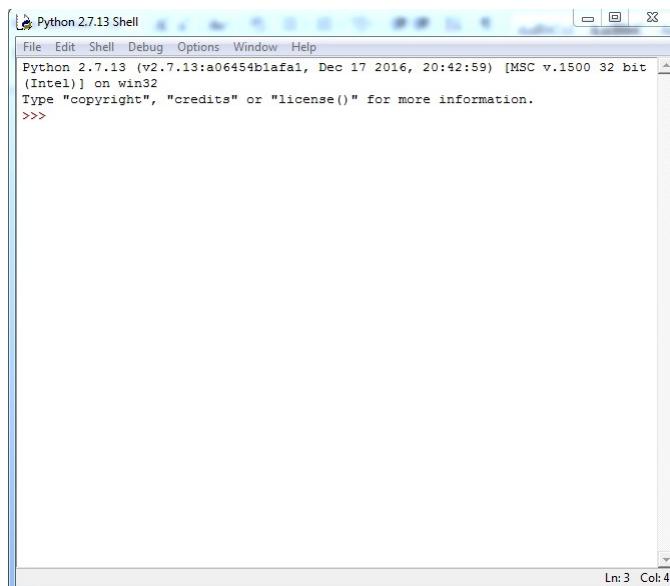
Python memiliki beberapa fitur yang menjadikan bahasa pemrograman ini berbeda dari bahasa lain antara lain:

1. Memiliki kepustakaan yang luas, dalam distribusi *Python* telah disediakan modul-modul ‘siap pakai’ untuk berbagai keperluan.
2. Memiliki tata bahasa yang jernih dan mudah dipelajari.
3. Memiliki aturan *layout* kode sumber yang memudahkan pengecekan, pembacaan kembali dan penulisan ulang kode sumber.
4. Berorientasi obyek.
5. Memiliki sistem pengelolaan memori otomatis (*garbage collection*, seperti java) Modular, mudah dikembangkan dengan menciptakan modul-modul baru; modul-modul tersebut dapat dibangun dengan bahasa *Python* maupun C/C++.
6. Memiliki fasilitas pengumpulan sampah otomatis, seperti halnya pada bahasa pemrograman Java, *Python* memiliki fasilitas pengaturan penggunaan memory komputer sehingga para pemrogram tidak perlu melakukan pengaturan memory komputer secara langsung.

7. Memiliki banyak fasilitas pendukung sehingga mudah dalam pengoperasiannya.

2.7.1.2 Hubungan *Python* dengan Analisis Sentimen

1. *Pre-processing* sederhana teks data di *Python*.
2. *Stopwords*, *Regular Expression* “sederhana”, *stemming*, *Scraping*, dsb.
3. *Read* dan *Write* data ke sebuah file di *Python* (dictionary, list, & set).



Gambar 2.6: Aplikasi *Python* 2.7.13

2.7.2 Platform *Rapidminer* (versi : 6.3)

Rapidminer adalah *perangkat lunak* data analisis yang memiliki banyak fitur untuk menyelesaikan berbagai permasalahan dengan proyek data *mining*. Dikembangkan oleh Rapid-I di Jerman, dengan lisensi AGPL yang membuatnya dapat diunduh siapa saja secara gratis dari Website resminya di <http://www.Rapidminer.com>.

Penulis lebih memilih menggunakan platform *Rapidminer* karena pekerjaan yang dilakukan oleh *Rapidminer* (*Data mining Tools*) adalah berkisar dengan analisis teks, mengekstrak pola-pola dari data set yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Tujuan dari analisis teks ini adalah untuk mendapatkan informasi bermutu tertinggi dari teks yang diolah.

Rapidminer menyediakan prosedur data *mining* dan *Machine Learning*, di dalamnya termasuk: ETL (*extraction, transformation, loading*), data *Pre-processing*, visualisasi, modelling dan evaluasi. Proses data *mining* tersusun atas operator-operator yang nestabel, dideskripsikan dengan XML, dan dibuat dengan GUI (*Graphical User Interface*).

2.7.2.1 User Interface

Secara umum tampilan antar muka *Rapidminer* terdiri atas tiga kolom utama. Pada kolom sebelah kiri terdapat panel *overview* yang memiliki tampilan *bird-eye* dari *Process Window* yang bisa memudahkan pengguna dalam memantau dan mengeksplorasi lebih mudah, terlebih jika kamu memiliki proses dalam ukuran besar. Berikutnya adalah panel *Repositori* yang menampung beragam menu fungsi repositori data. Terakhir adalah panel operator yang menampung hampir seluruh menu fungsional seperti *Process Control, Utility, Repository Access, Import, Export, Data Transformation, Modeling and Evaluation*. Selanjutnya kita akan melihat kolom kedua di posisi tengah yang memiliki ukuran paling lebar dibandingkan dua kolom lainnya. Namun, pengguna dapat mengatur lebar kolom.

2.7.2.2 Dukungan Basis Data

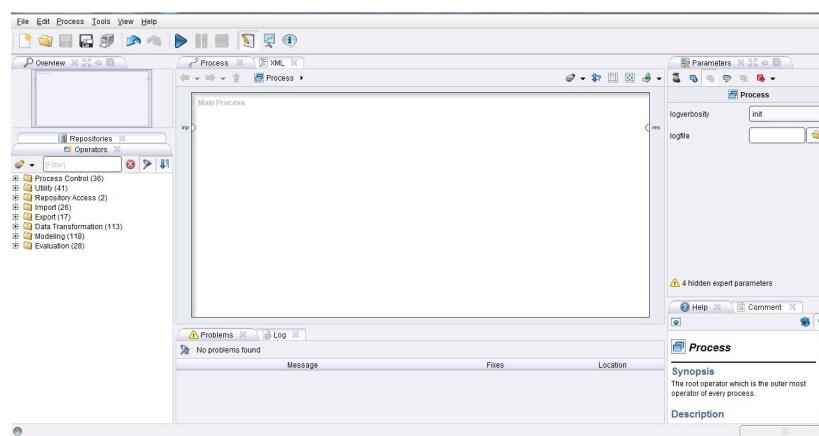


Gambar 2.7: Platform *Rapidminer* 6.3

Dalam sebuah proyek data *mining* atau data analisis, kebanyakan menggunakan basis data sebagai sumber data. *Rapidminer* menyediakan dukungan untuk beberapa *RDBMS* (*Relational Database Management System*), beberapa yang terkenal diantaranya *HSQLDB, Ingres, Microsoft SQL Server, MySQL, Ms. Access, Oracle, PostgreSQL, Sybase*.

2.7.2.3 Dukungan Import File Beragam Format

CSV Format, *Spreadsheet XLS*, *AccessDatabase*, XML File, *Binary File*, *Database* tabel. Disamping beragam fungsi pemrosesan data yang dimiliki instalasi bawaan *Rapidminer*, kamu juga bisa menambahkan ekstensi tambahan untuk menambahkan fitur ke *Rapidminer*.



Gambar 2.8: Fitur *Rapidminer* 6.3

Bab 3

ANALISIS DAN PEMBAHASAN

3.1 Gambaran Umum Analisis

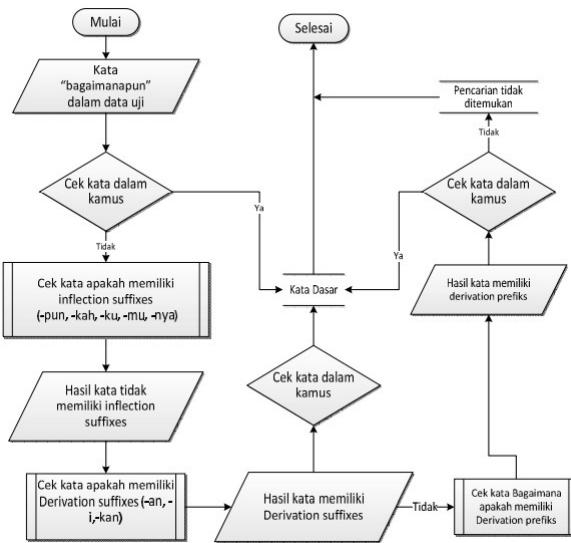
Penulis membuat sebuah analisis sentimen publik terhadap suatu hotel dari situs PegiPegi.com dengan mengambil data kumpulan ulasannya sebagai data. Pembuatan analisis sentimen ini dibagi menjadi 3 tahap yaitu tahap *Scraping* menggunakan perangkat lunak *Python*, platform *Rapidminer* dalam implementasi algoritma yaitu Klasifikasi *Naïve Bayes* dan Aplikasi sederhana dengan hasil output data yang sudah diolah.

3.2 Flowchart

Pada sub-bab ini akan dijabarkan mengenai perancangan berupa Flowchart yaitu Flowchart *Web Scraping*, Flowchart *stemming*, dan Flowchart Arsitektur Sistem Analisis Sentimen pada bahasa pemrograman *Python*.

3.2.1 Flowchart *Stemming*

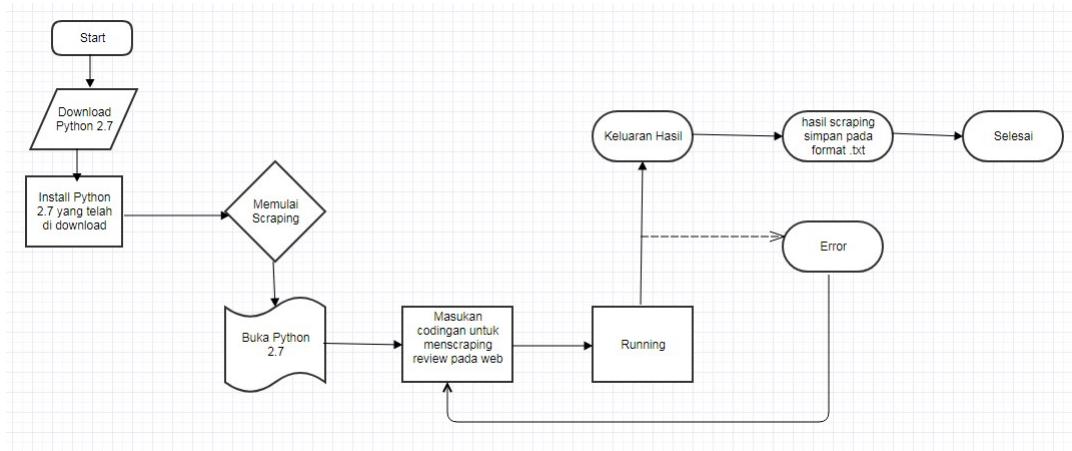
Berikut proses *stemming* atau proses dimana pembuangan imbuhan kata. Dengan mempunyai alur:



Gambar 3.1: Flowchart *Stemming Kata Baku*

3.2.2 Flowchart *Web Scraping*

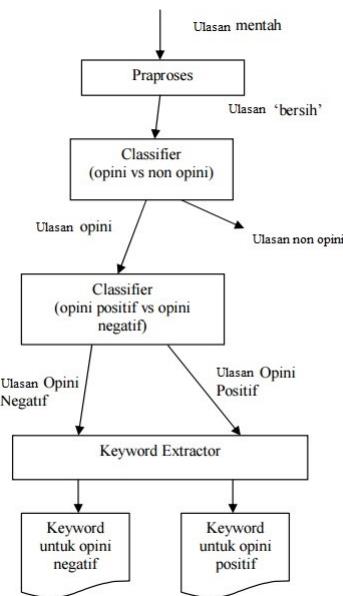
Flowchart *Web Scraping* ini merupakan konsep alur *Scraping Web* pada aplikasi *Python 2.7* yang telah digunakan oleh user. Disini menggunakan fitur Editor dengan penyorotan sintaks dan output. Setiap menyelesaikan kodingan *Scraping* akan langsung mendapatkan hasil outputnya pada *Python shell*. Dan jika sudah menyelesaikan secara keseluruhan datanya maka dapat menyimpan file dengan format .txt.



Gambar 3.2: Flowchart *Web Scraping*

3.2.3 Flowchart Arsitektur Sistem Analisis Sentimen

Sistem ini menerima masukan berupa data komentar atau ulasan mentah. Data ini akan dipra-proses terlebih dulu sebelum diproses lebih lanjut.



Gambar 3.3: Flowchart Arsitektur Sistem Analisis Sentimen

3.3 Rancangan Sistem

3.3.1 Metode Pengumpulan Data

Untuk memperoleh data tersebut penulis mengambil data dari user ulasan di <https://www.PegiPegi.com> dengan memilih daerah Bekasi untuk hotel yang akan diulasan dengan bahasa Indonesia. Data yang diambil berjumlah kurang lebih ratusan data sentimen baik data sentimen positif dan data sentimen negatif.

Ramah bersih tenang itu yg sy rasakan selama nginep di Red Planet sehingga sy dan psngan sy merasa nyaman honeymoon disitu...

Gambar 3.4: Contoh Sentimen Positif Berbahasa Indonesia

Hotel yg bagus kamar nya tapi sayang kuning air nya... Breakfast nya krg cepet buat restock makanan nya mungkin karena rame ya

Gambar 3.5: Contoh Sentimen Negatif Berbahasa Indonesia

3.3.2 Metode Pemrosesan Awal Data

Untuk memudahkan dalam mengelola data maka data perlu kita berikan analisis sentimen secara manual dengan membaca maksud dari kalimat yang ada dalam sentimen tersebut, sehingga dapat diberikan penilaian bahwa sentimen tersebut merupakan sentimen negatif atau positif. Setelah itu data kita pisahkan antara sentimen positif dan negatif masing masing kedalam folder yang berbeda, yaitu untuk sentimen positif kita masukkan ke folder positif sedangkan sentimen negatif kita masukkan ke folder negatif hal ini dilakukan guna memudahkan saat pengambilan, penambahan, pengurangan maupun saat pemrosesan terhadap data tersebut.

3.3.3 Model yang Diusulkan

Pada penelitian ini penulis mengusulkan suatu metode dalam analisis sentimen guna mengetahui sentimen masyarakat dalam ruang lingkup pemesanan hotel *online* di PegiPegi.com, dengan menggunakan metode Algoritma *Naïve Bayes Classifier* merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Dalam penelitian ini yang menjadi data uji adalah dokumen opini. Ada dua tahap pada klasifikasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya. Dalam algoritma *Naïve Bayes Classifier* setiap dokumen direpresentasikan dengan pasangan atribut “x₁, x₂, x₃, ...x_n” dimana x₁ adalah kata pertama, x₂ adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori opini. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (Vmap).

3.4 Eksperimen dan Pengujian

Dengan menggunakan rumus klasifikasi *Naïve Bayes* berikut:

Keterangan:

$$P(K|X) = \frac{nK+1}{nX+|Vocab|}$$

P(K|X) = Probabilitas data dengan vektor K pada kelas X.

K = Kata X = Kelas kata (positif/negatif).

nK = Jumlah munculnya kata yg sama

nX = Jumlah munculnya kata yg sama

$|Vocab|$ = Mutlak jumlah seluruh kata dari berbagai kelas kata

Penulis mengambil sebagian data hasil *Pre-processing* sebagai contoh perhitungan manual:

Tabel 3.1: Contoh Perhitungan Sebagian Data Hasil *Pre-processing*

Kata	POS-Tag	Frekuensi	Sentimen
bagus	adj	1	positif
favorit	n	7	positif
lambat	adj	1	negatif
lumayan	adj	3	positif
masalah	adv	1	negatif

Jawab:

$$P(kata|sentimen) = \frac{nkata+1}{nsentimen+|TotalKata|}$$

- $P(lumayan|positif) = \frac{1+1}{9+|13|} = 0,090$
- $P(bagus|positif) = \frac{7+1}{9+|13|} = 0,363$
- $P(lambat|negatif) = \frac{1+1}{4+|13|} = 0,117$
- $P(masalah|negatif) = \frac{3+1}{4+|13|} = 0,235$
- $P(favorit|positif) = \frac{1+1}{9+|13|} = 0,090$

Probabilitas kelas kata positif dari total 5 kelas kata sebesar 0,6 sedangkan probabilitas kelas kata negatif dari total 5 kelas kata sebesar 0,4. Jadi, $P(\text{positif}) > P(\text{negatif})$ kelas kata positif lebih dominan dibanding kelas kata negatif.

Dikarenakan banyaknya data maka penulis menggunakan bantuan *Tools* yang digunakan dalam eksperimen ini adalah *Rapidminer* yang diupdate dengan penambahan *plugin text processing* yang telah memiliki fitur pemrosesan teks di antaranya:



Gambar 3.6: Proses Analisis Teks dengan *Rapidminer* 6.3

3.4.1 Text *Pre-processing*

Dengan fitur text *Pre-processing* adalah mengubah semua karakter menjadi huruf kecil dan menguraikan kalimat menjadi kata dan menghilangkan karakter spesial seperti titik (.), koma (,), spasi, dan karakter yang tidak diperlukan. Data *Pre-processing* mengurangi dokumen teks input secara signifikan.

Pre-processing

3.4.1.1 Filter Tokenizing

Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca serta memfilter berdasarkan panjang teks. Untuk metode pengujian atau evaluasi dilakukan pengujian terhadap model-model yang diteliti untuk mendapatkan informasi model diusulkan.

Tabel 3.2: Ilustrasi *Tokenizing*

Contoh Data			
Data Latih hasil <i>tokenizing</i>		Tahapan <i>stemming</i>	
Input	Data Latih hasil <i>tokenizing</i>	Output	Tahapan <i>stemming</i>
	“mau bagaimanapun juga pelayanan hotel harus bagus agar lebih nyaman”		“mau” “bagaimana” “juga” “layan” “hotel” “harus” “bagus” “agar” “lebih” “nyaman”

3.4.1.2 *Stemming*

Pada tahap ini dilakukan pembuangan imbuhan kata. Misalnya kata "bagaimanapun" maka langkah yang dilakukan *stemming* dijalankan selanjutnya mengecek kata "bagaimanapun" pada kamus kata dasar, jika tidak ada maka akan masuk ke tahap penghapusan partikel yaitu "-pun" pada akhiran kata dasar selanjutnya pengecekan awalan jika tidak ada maka algoritma mengembalikan kata menjadi "bagaimana" dan *stemming* berhenti. Untuk *stemming* kata "jg" dan "jd" maka sistem akan mengecek pada kamus yang berisi singkatan dari kata-kata dasar anomali yang sudah didaftarkan contoh yang terdapat pada kamus kata "jg"="juga", "jd"="jadi". Sehingga semua kata tidak yang terdaftar pada kamus yang berisi singkatan dari kata-kata dasar akan diganti ke kata dasar yang baku. Pada flowchart *stemming* di atas tahapan yang dilakukan adalah sebagai berikut:

1. Kata dicari terlebih dahulu pada kamus. Jika ditemukan, maka kata tersebut adalah kata dasar, jika tidak maka langkah 2 yang dilakukan.
2. Pada kata dalam data uji akan dicek apakah memiliki *inflectional suffixes*, yaitu akhiran ("lah", "kah", "tah", "pun") dan kata ganti kepunyaan atau *possessive pronoun PP* ("ku", "mu", "nya") ternyata pada kata dalam data uji tidak terdapat *inflectional suffixes* dan *possessive pronoun* kemudian sistem melakukan proses selanjutnya *Derivation prefiks*.
3. Pada kata dalam data uji akan dicek apakah memiliki *Derivation suffixes*, yaitu akhiran ("an", "i", "kan") ternyata pada kata dalam data uji terdapat *Derivation suffixes* maka sistem menghapus akhiran *derivation suffixes* lalu sistem mengecek ke kamus kata dasar jika kata hasil *derivation suffixes* ada di dalam kamus kata dasar maka sistem berhenti jika tidak ditemukan akan dilakukan proses selanjutnya.
4. Kata dalam data uji akan dicek, apakah memiliki *derivation prefixes* (awalan), ternyata kata dalam data uji mengandung *derivation prefixes* kemudian sistem mencari kata ke kamus kata dasar jika ditemukan maka kata dalam data uji adalah kata dasar dan proses berhenti, jika tidak maka kata dikembalikan dan proses berhenti.

Tabel 3.3: Ilustrasi *Stemming*

Contoh Data			
Data Latih hasil <i>tokenizing</i>		Tahapan <i>stemming</i>	
Input	“mau”	Output	“mau”
	“bagaimanapun”		“bagaimana”
	“juga”		“juga”
	“pelayanan”		“layan”
	“hotel”		“hotel”
	“harus”		“harus”
	“bagus”		“bagus”
	“agar”		“agar”
	“lebih”		“lebih”
	“nyaman”		“nyaman”

3.4.1.3 Filter Stopwords (Indonesia)

Dengan fitur ini maka teks sebelum diklasifikasikan, dihilangkan teks yang tidak berhubungan dengan analisis sentimen sehingga lebih efisien. Filter *Stopwords* bahasa indonesia ini penulis ambil dari internet yang dibuat oleh Sastrawi dengan penulis menambahkan beberapa kata yang memiliki arti sama dengan kata-kata yang sudah ada dalam daftar *Stopwords* tersebut.

Tabel 3.4: Ilustrasi *Stopwords*

Contoh Data			
Data Latih hasil <i>stemming</i>		Tahapan <i>Stopwords</i>	
Input	“mau”	Output	
	“bagaimana”		“layan”
	“juga”		“hotel”
	“layan”		“bagus”
	“hotel”		“lebih”
	“harus”		“nyaman”
	“bagus”		
	“agar”		
	“lebih”		
	“nyaman”		

3.5 Rancangan Kamus Data

3.5.1 Struktur Kamus Data

Dalam sistem Analisis Sentimen pada ulasan situs pemesanan hotel *online* dengan teks bahasa Indonesia. Diperlukan sebuah kamus data berupa “kata” bahasa Indonesia umum yang biasa digunakan pada Pengolahan Teks (*Text Processing*). Berikut penulis sajikan beberapa kata dari kamus data berupa tabel, yang terdiri dari 3 tabel yaitu tabel Kamus Besar Bahasa Indonesia (KBBI), kamus data Formal (Sastrawi) dan tabel kamus data Non-formal (Penulis):

3.5.1.1 Tabel Kamus Data KBBI

Dalam tabel ini berisi kata-kata menurut Kamus Besar Bahasa Indonesia (KBBI):

Tabel 3.5: Kamus Data KBBI

A	B	C	D
adapun	betul	cadas	demi
adem	biadab	cahaya	denah
aku	biasa	cangkir	desak
alhasil	biaya	cantik	detak
ampuh	bising	cenderung	diam
asumsi	buang	ceroboh	diskon

Kamus data KBBI diambil penulis yang digunakan pada proses *Stopwords* sebagai data untuk menyaring hasil *scraping* agar seluruh kata menjadi hanya kata Bahasa Indonesia saja.

3.5.1.2 Tabel Kamus Data Formal

Dalam tabel ini berisi kata-kata Formal bersumber dari Sastrawi:

Tabel 3.6: Kamus Data Formal (Sastrawi)

A	B	C	D
akankah	bersuara	cara	dan
alur	berapapun	cukup	dari pada
antara	berkali-kali	cuma	diibaratkan
berasal	bermula	caranya	dikatakan
atau	bertanya	carikan	dimintai
bawasanya	belakangan	cukup	diperlukan

Kamus data formal diambil penulis dari berkas Sastrawi yang digunakan pada proses *stemming* dari hasil data *scraping* untuk menghilangkan kata imbuhan menjadi kata dasar (contoh: bersuara diubah menjadi suara).

3.5.1.3 Tabel Kamus Data Non-formal

Dalam tabel ini berisi kata-kata Non-Formal campuran yang dibuat oleh Penulis:

Tabel 3.7: Tabel Kamus Data Non-formal

rsk	top	gue	mantep
yg	tak	ya	iya
gitu	justru	amat	makin
bgs	nyaris	telanjur	sekedar
emang	pw	bgt	jlk

Kamus data Non-formal ditambahkan penulis berdasarkan pengamatan diri sendiri secara manual terhadap Bahasa kehidupan sehari-hari tentang penggunaan kata-kata seperti ini. Kamus data ini digunakan pada proses *Stopwords* untuk dihilangkan. (rsk memiliki arti rusak, bgs memiliki arti bagus, pw memiliki posisiwenak, bgt memiliki arti banget, jlk memiliki arti jelek).

3.6 Rancangan Sistem

Pada sub-bab ini akan dijelaskan mengenai perancangan *Scraping* data, text *Pre-processing* dan *feature selection* menggunakan *Python* 2.7.13 serta cara kerja algoritma pada aplikasi pengolah data *Machine Learning Rapidminer* hingga menghasilkan output yang signifikan. Perancangan sistem tersebut meliputi:

3.6.1 Library Python

Sebelum memulai memasukan *Syntax Scraping*, terlebih dahulu install modul-modul yang diperlukan seperti: *nltk*, *corpora*, *gensim*, *sastrawi* dll yang semuanya merupakan package pada library *Python* untuk proses *Scraping*.

```
C:\Users\hp>pip list
DEPRECATION: --legacy-format is deprecated. Use --format=legacy|columns
or --format=columns under the [list] section
boto <2.47.0>
bz2file <0.98>
certifi <2017.4.17>
chardet <3.0.4>
ConfigParser <1.0>
gensim <2.2.0>
idna <2.5>
nltk <2.0.4>
numpy <1.13.0+mkl>
pip <9.0.1>
PyYAML <3.12>
requests <2.18.1>
Sastrawi <1.0.1>
scipy <0.19.1>
setuptools <28.8.0>
six <1.10.0>
smart-open <1.5.3>
urllib3 <1.21.1>
```

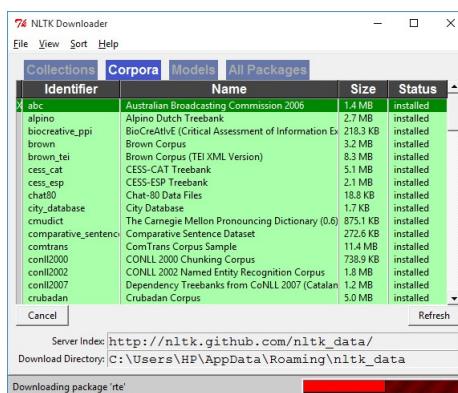
Gambar 3.7: Package pada Library *Python* 2.7.13

Syntax untuk download modul NLTK (Natural Language Toolkit) pada *Python* 2.7.13 Shell:

Algoritma 3.1 Download *nltk*

```
>>> import nltk
>>> nltk.download()
```

Pada *Syntax* baris pertama untuk memanggil modul *nltk* dan pada baris kedua untuk mendownload nya.



Gambar 3.8: Package pada NLTK Library *Python*

3.6.2 Proses *Scraping*

Pada bagian ini merupakan proses dimana data akan diambil dari isi sebuah halaman situs PagiPagi.com secara spesifik, yaitu ulasan dari ulasan sese-

orang terhadap suatu pelayanan atau sejenisnya. Berikut adalah sebagian *Syntax* pada proses *Scraping* (hanya “Hotel Horison”):

Algoritma 3.2 Scraping

```
import urllib
import re
htmlfile = urllib.urlopen("https://www.pegipegi.com/hotel/bekasi/
hotel_horison_ultima_bekasi_989768/")
htmlltext = htmlfile.read()
regex = '<div class="contentComment"
itemprop="description">(.*?)</div>'
pattern = re.compile(regex)
hasil = re.findall(pattern, htmlltext)
print(hasil)
```

Syntax di atas merupakan proses dimana awal dari proses *Scraping* dimulai yaitu teknik yang digunakan untuk mengambil isi sebuah halaman situs pemesanan hotel *online* secara spesifik. Baris ke-3 berupa input dari link pada situs tersebut dan Baris ke-5 adalah bagian dimana ulasan dari masyarakat mengenai hotel tersebut. Berikut adalah hasil output dari *Syntax* di atas:

```
Python 2.7.13 Shell
File Edit Shell Debug Options Window Help
=====
RESTART: C:\Users\USER\Documents\PI\Scrapfor.py =====
['Tempat nya oke banget, strategis, kebersihan ok', 'Over all baik', 'Hotel Horison Ultima bekasi sangat nya man, selain letaknya yang strategis karena letaknya berada di jantung kota bekasi pelayanannya pun sangat baik dan hotel ini sangat recommended bagi pelancong-pelancong yang ingin mendapatkan kenyamanan yang lebih', 'Sangat baik dari semua segi baik pelayanan, kebersihan dll sehingga perjalanan dengan keluarga bisa sangat berkesan', 'Horison Hotel Bekasi memberikan kenyamanan utk berlibur bersama keluarga.selain kamarnya yg cukup nyaman dan bersih,pelayanannya pun ramah.So tidak akan mengecewakan deh...', 'Happy to be back again and again to this place. I've stayed near on Bekasi. But if i needed to quality time, no need to go to far away. Just stay and try the facilities. I will be back as soon.', 'Nyaman banget, fasilitas mantap, makanannya top banget. Intinya betah dah ', 'Mantap untuk bisnis dan Cocok Untuk keluar Anda.. di Weekend..atau liburan.. '+'element.review+''] ['Ke 2 Kalinya saya bermalam di RED Planet bekasi, dan emang smuanya it dr pelayanan dan kamar sudah baik. Intinya pas dan nyaman bwt kt yg jenuh dan mau beristirahat dg tenang di RED Planet aj'. 'Ramah bersih tenang itu yg sy rasakan selama nginep di Red Planet sehingga sy dan psngan sy merasa nyaman honeymoon disitu... ', 'Clean and nice places wt affordable prices on pegipegi and strategic hotel next to carefour and siloam hospital, many food&#039;s cafe around hotel area, recommended', 'overall memuaskan', 'Akses sangat deket dengan akses tol, sangat mudah di jangkau.', 'kamar bersih, staff semua ramah. next holiday nginep disini lagi', 'Rekomend deh mantap fasilitas dan pelanannya Top bwt saya pribadi krn saya baru pertama x ke Red Planet bekasi', 'Dengan harga yg cukup murah saya mendapatkan kamar yg bagus dan nyaman.', 'Hotel yg bersih dan sangat nyaman untuk menginap serta di dukung staff hotel yg terbaik,juga dekat dengan lokasi saya berkerja pada saat itu.serta di dukung koneksi wifi yang cepat memudahkan saya untuk berkomunikasi dan jejaring sosial', 'sangat nyaman menginap di hotel red planet bekasi', "'+'element.review+''] ['hotel yg cukup nyaman baik kamar maupun kmr mandi nya, lokasi sgt strategis krn bersebelahan dgn mall summarecon jd mudah utk cari variasi kuliner & hiburan...swimming poll jd sgt nyaman utk balita..karyawannya jd ramah..breakfastnya lumayan enak & byk sekali variasi menu tradisionalnya..mantap pokoknya lah...', 'saya dan keluarga puas dg harga, pelayanan dan fasilitas hotel', 'Kamar luas, bersih, dan fasilitas lengkap, lokasi strategis, pelayanan staff ramah cuma sedikit lambat.. ', 'Pengalaman pertama menginap di hotel harris summarecon bekasi sangat mengesankan. Kami memutuskan untuk menginap di harris krn disana terdapat kolam renang anak, kids station dan program2 fun untuk anak2. Kamar hotel juga sangat nyaman,bersih&ckp besar.Variasi sarapannya jd sangat byk. Pelayanan staff juga sangat ramah.Lokasi hotel yg berdekatan dgn mall membuat suasana semakin menyenangkan.Bravo Hotel Harris Bekasi&thank u for excellent service!', 'Hotel ini merupakan salah satu hotel yang sangat direkomendasikan untuk menginap di bekasi. Lokasi yang strategis, fasilitas lengkap, dan pelayanan yang baik. Hotel ini cocok untuk liburan keluarga atau bisnis. Jaraknya dekat dengan pusat kota dan berbagai destinasi wisata. Kamar yang nyaman dan bersih. Staff yang ramah dan profesional. Harga yang kompetitif. Hotel ini menjadi pilihan yang tepat bagi mereka yang mencari penginapan yang nyaman dan memenuhi berbagai kebutuhan.']
```

Gambar 3.9: Output *Scraping*

3.6.3 Rancangan Tokenizing

Pada bagian ini merupakan proses untuk membagi teks yang dapat berupa kalimat, paragraf atau dokumen, menjadi token-token/bagian-bagian / perkata tertentu. Berikut adalah potongan *Syntax* pada proses *tokenizing*:

Algoritma 3.3 Tokenizing

```
from nltk.tokenize import RegexpTokenizer
#memanggil paket dari regexptokenizer dari modul nltk.
regexpTokenizer = reguler expression dari paket nltk untuk
tokenizer
Token = str(hasil)
#Token = variabel yg difungsikan untuk mengubah tipe data
variabel Total menjadi string.
tokenizer = RegexpTokenizer(r'\w+')
#Tokenizer = variabel untuk mendefinisikan Regular Expression
Tokens = tokenizer.tokenize(Token)
#pemanggilan fungsi tokenize beserta output
```

Syntax di atas merupakan proses *tokenizing* dengan mengambil data hasil *Scraping* lalu mengubah semua kalimat menjadi potongan-potongan kata. Berikut adalah hasil output dari Syntax di atas:



Gambar 3.10: Output Tokenizing

3.6.4 Rancangan Stemming

Pada bagian ini merupakan teknik yang digunakan untuk mengurangi kata-kata untuk bentuk kata dasar dengan menghilangkan kata imbuhan dengan kamus data yang telah disiapkan untuk menguji data.

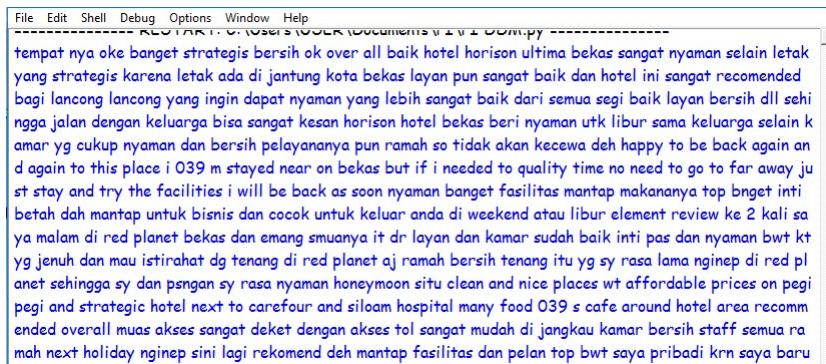
Berikut adalah potongan *Syntax* pada proses stemming:

Algoritma 3.4 stemming

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
#Import paket sastrawi
factory = StemmerFactory()
#Pembuatan stemmer dengan memanggil StemmerFactory dari
Sastrawi
stemmer = factory.create_stemmer() Stemm = str(Tokens)
#Untuk memasukan data Tokens
stemming = stemmer.stem(Stemm)
#Pemanggilan fungsi stemming beserta output
```

Syntax di atas merupakan proses *stemming* dengan mengambil tipe data hasil *tokenizing* lalu mengubahnya menjadi kata dasar.

Berikut adalah hasil output dari Syntax di atas:



```
File Edit Shell Debug Options Window Help
tempat nya oke banget strategis bersih ok over all baik hotel horison ultima bekas sangat nyaman selain letak yang strategis karena letak ada di jantung kota bekas layan pun sangat baik dan hotel ini sangat recommended bagi lancung lancung yang ingin dapat nyaman yang lebih sangat baik dari semua segi baik layan bersih dll sehi ngga jalan dengan keluarga bisa sangat kesan horison hotel bekas beri nyaman utk libur sama keluarga selain k amar yg cukup nyaman dan bersih pelayanannya pun ramah so tidak akan kecewa deh happy to be back again an d again to this place i O39 m stayed near on bekas but if i needed to quality time no need to go to far away ju st stay and try the facilities i will be back as soon nyaman banget fasilitas mantap makananya top bngt inti betah dah mantap untuk bisnis dan cocok untuk keluar anda di weekend atau libur element review ke 2 kali sa ya malam di red planet bekas dan emang smuanya it dr layan dan kamar sudah baik inti pas dan nyaman bwt kt yg jenuh dan mau istirahat dg tenang di red planet aj ramah bersih tenang itu yg sy rasa lama nginep di red pl ant sekingta sy dan psngan sy rasa nyaman honeymoon situ clean and nice places wt affordable prices on pegi pegi and strategic hotel next to carefour and siloam hospital many food O39 s cafe around hotel area recomm ended overall muas akses sangat deket dengan akses tol sangat mudah di jangkau kamar bersih staff semua ra mah next holiday nginep sini lagi rekomen deh mantap fasilitas dan pelan top bwt saya pribadi krn saya baru
```

Gambar 3.11: Output *Stemming*

3.6.5 Rancangan Filter (*Stopwords*) dan *Topic Model*

Pada bagian ini terdapat dua proses sekaligus yaitu pada Filter (*Stopwords*) yang merupakan fitur dimana teks sebelumnya diklasifikasikan, di hilangkan dahulu teks yang tidak berhubungan dengan analisis sentimen sehingga dimensi teks akan berkurang tanpa mengurangi isi sentimen dari teks tersebut. Seperti contoh kata: “dan”, “kamu”,dll.

Berikut adalah potongan Syntax pada proses *Stopwords*:

Algoritma 3.5 Stopwords

```
from nltk.corpus import Stopwordss
#import Stopwordss dari library nltk
stopw = set(Stopwordss.words('kbbi'))
#var stopw untuk menampung Stopwordss file 'kbbi'
datasw = str(stemming) #var datasw mengubah tipe data dari hasil
stemming menjadi string
SWindo = [i for i in datasw.lower().split() if i in stopw]
#var SWindo untuk melakukan Stopwordss yaitu dengan menyaring
lalu mengambil hanya kata yang berada dalam kbbi saja
```

Syntax di atas merupakan proses *stemming* dengan mengambil tipe data hasil *stemming* lalu menyaring semua kata dan mengambil hanya bahasa Indonesia saja. Berikut adalah hasil output dari Syntax di atas:

Gambar 3.12: Output Stopwords

Sedangkan *Topic Model* untuk menghasilkan representasi dokumen berupa kata-kata kunci dari dokumen. Kata-kata kunci tersebut yang akan digunakan dalam proses pengindeksan/frekuensi kata serta pencarian dokumen untuk ditemukan kembali sesuai kebutuhan pengguna.

Berikut adalah potongan Syntax pada proses *Topic Model* dengan menggunakan modul *Counterlist*:

Algoritma 3.6 Topic Model

```
#Counterlist
from collections import Counter
list = SWindo
#untuk menghitung banyak nya frekuensi dari kemunculan kata
counts = Counter(list)
print(counts)
```

Syntax di atas merupakan proses *Topic Model* dengan menggunakan modul counterlist dan mengambil tipe data hasil *Stopwords* lalu menghitung banyaknya frekuensi output dari Syntax di atas:

```
Counter({'hotel': 30, 'dan': 24, 'nyaman': 20, 'sangat': 17, 'di': 17, 'untuk': 16, 'bersih': 15, 'kamar': 14, 'yang': 13, 'baik': 11, 'saya': 10, 'dengan': 10, 'bekas': 9, 'anak': 9, 'kolam': 8, 'inap': 8, 'ada': 8, 'ramah': 8, 'fasilitas': 7, 'strategis': 7, 'renang': 7, 'libur': 7, 'banget': 6, 'layan': 6, 'dapat': 6, 'juga': 6, 'ini': 6, 'lengkap': 5, 'lokasi': 5, 'harga': 5, 'pun': 5, 'tidak': 5, 'luas': 5, 'karena': 5, 'planet': 5, 'variasi': 4, 'cukup': 4, 'keluarga': 4, 'mandi': 4, 'kami': 4, 'mantap': 4, 'bisa': 4, 'enak': 4, 'bagus': 4, 'besar': 4, 'muas': 4, 'kali': 3, 'putus': 3, 'selain': 3, 'rasa': 3, 'dari': 3, 'harap': 3, 'buat': 3, 'jadi': 3, 'murah': 3, 'paling': 3, 'semua': 3, 'dekat': 3, 'itu': 3, 'sehingga': 3, 'mudah': 3, 'tenang': 3, 'keluar': 3, 'cari': 3, 'letak': 3, 'pokok': 3, 'top': 3, 'senang': 3, 'kota': 3, 'puas': 3, 'indah': 3, 'serta': 3, 'banyak': 2, 'dukung': 2, 'tempat': 2, 'sana': 2, 'bawa': 2, 'deh': 2, 'lain': 2, 'pakai': 2, 'sini': 2, 'akese': 2, 'lancang': 2, 'milik': 2, 'via': 2, 'inti': 2, 'masalah': 2, 'makan': 2, 'cocok': 2, 'hibur': 2, 'sarap': 2, 'lebih': 2, 'beri': 2, 'sajad': 2, 'sedikit': 2, 'sama': 2, 'mau': 2, 'atau': 2, 'kenan': 2, 'ke': 2, 'pertama': 2, 'deposit': 2, 'over': 1, 'orang': 1, 'beli': 1, 'segit': 1, 'rupat': 1, 'akhir': 1, 'menu': 1, 'bisnis': 1, 'perlu': 1, 'ain': 1, 'tapi': 1, 'jenak': 1, 'main': 1, 'kloset': 1, 'pelan': 1, 'relaks': 1, 'ingin': 1, 'dah': 1, 'oke': 1, 'dar': 1, 'cepat': 1, 'sahur': 1, 'bahau': 1, 'cuman': 1, 'padat': 1, 'baru': 1, 'jantung': 1, 'terus': 1, 'kim': 1, 'cuma': 1, 'lah': 1, 'sosial': 1, 'saat': 1, 'sudah': 1, 'guna': 1, 'hingga': 1, 'email': 1, 'sayang': 1, 'usah': 1, 'sal': 1, 'lumayan': 1, 'tradisional': 1, 'kurang': 1, 'emang': 1, 'hospital': 1, 'habis': 1, 'jangkau': 1, 'koneksi': 1, 'akin': 1, 'kecil': 1, 'ragam': 1, 'tinggal': 1, 'area': 1, 'pribadi': 1, 'seluncur': 1, 'suara': 1, 'kerja': 1, 'lagi': 1, 'tol': 1, 'kalau': 1, 'akan': 1, 'alam': 1, 'khusus': 1, 'semangat': 1, 'ultima': 1, 'tanpa': 1, 'keren': 1, 'malam': 1, 'hanya': 1, 'satu': 1, 'betah': 1, 'macet': 1, 'jenuh': 1, 'alir': 1, ' tutup': 1, 'lambat': 1, 'langkah': 1, 'utama': 1, 'situ': 1, 'pas': 1, 'banding': 1, 'maupun': 1, 'langsung': 1, 'kaya': 1, 'ya': 1, 'dewasa': 1, 'kedap': 1, 'selalu': 1, 'bagi': 1, 'kuliner': 1, 'kan': 1, 'sauna': 1, 'staf': 1, 'lama': 1, 'susah': 1, 'aneka': 1, 'favorit': 1, 'jalan': 1, 'suka': 1, 'istirahat': 1, 'tidur': 1, 'kecewa': 1})
```

Gambar 3.13: Output Topic Model

3.6.6 Rancangan Kamus Data

Pada bagian ini merupakan proses dimana hasil dari *Stopwords & Topic Model* akan diolah kembali ke dalam *spReadsheet* dalam bentuk tabel yang dimana di klasifikasikan secara spesifik sesuai kategori kamus data analisis sentimen. Sehingga saat proses berikutnya (*Rapidminer*) mudah dipahami oleh aturan kompilator tersebut untuk memunculkan output yang diinginkan.

Kata	POS-Tag	Frekuensi	Sentimen
nyaman	Adj	17	positif
bersih	Adj	16	positif
bekas	n	10	negatif
bagus	Adj	7	positif
lengkap	Adj	7	positif
strategis	Adj	7	positif
dekat	Adj	6	positif
tidak	Adv	5	negatif
senang	Adj	5	positif
enak	Adj	3	positif
cuma	Adv	2	negatif

Tabel 3.8: Sebagian Kamus Data Hasil *Stopwords & Topic Model*

Penulis memindahkan data hasil dari proses *Pre-processing* secara manual kedalam dokumen *spreadsheet* dengan 4 kolom diantaranya:

- Kata = berisi data dari kata-kata yang didapat dari seluruh proses *Text Pre-processing*

- POS-Tag = berisi data kelas-kelas tiap kata termasuk golongan dalam tabel:

Tabel 3.9: POS-Tag

Kelas Kata	Arti
Adj	Adjektiva (kata sifat)
Adv	Adverbia (kata keterangan)
n	Nomina (kata benda)
v	Verba (kata kerja)
p	Partikel

- Frekuensi = berisi data jumlah banyaknya kata yang muncul
- Sentimen = berisi data dari penentuan analisis secara manual penuis dengan meneliti data ulasan, Karena sampai saat ini penulis belum menemukan data tagging kata positif maupun negatif berbahasa indonesia

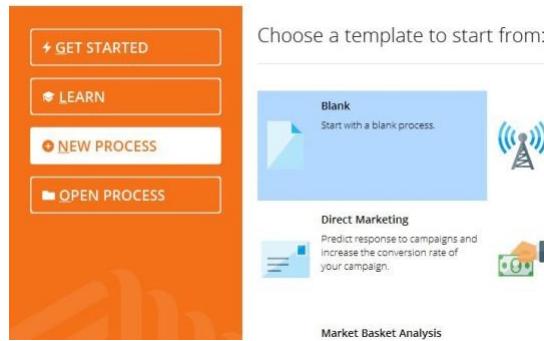
3.7 Pengujian

Pada sub-bab ini akan dijelaskan mengenai pengujian analisis sentimen menggunakan algoritma *Naïve Bayes* pada platform *Rapidminer*. Pengujian yang dilakukan adalah pengujian diagram. Pengujian diagram dilakukan untuk mengetahui sejauh mana hasil dari semua komentar masyarakat akan hotel tersebut apakah dominan positif atau negatif.

3.7.1 Pengujian Aplikasi

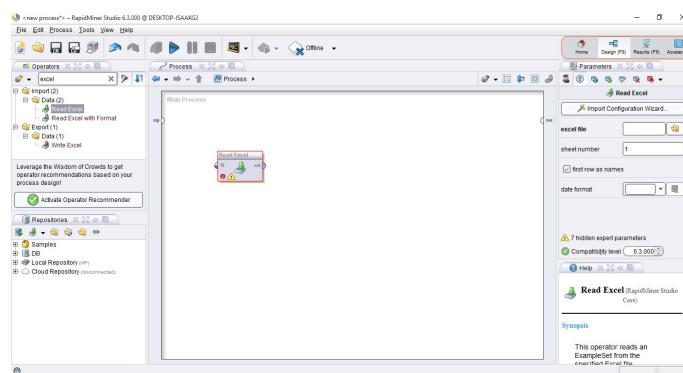
Pada pengujian ini akan dilakukan langkah demi langkah data *mining*, termasuk operator untuk input, output, data *Pre-processing*, dan mengkombinasikannya dengan mengekstrak pola-pola pada metode algoritma, kecerdasan buatan serta database. Tujuannya adalah untuk menciptakan sebuah model yang memprediksi nilai variabel/presentase target berdasarkan beberapa variabel masukan. Berikut langkah-langkah dari pengujian analisis sentimen pada metode algoritma *Naïve Bayes* menggunakan aplikasi *Rapidminer*:

1. Buka aplikasi *Rapidminer Studio*
2. pilih *New Process, Blank*



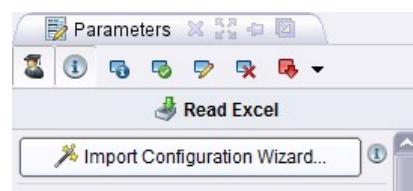
Gambar 3.14: Tampilan Mulai Project Baru

3. Pada *Operators*, ketik *excel* kemudian pilih *Read excel* (drag) yang berfungsi sebagai inputan pertama dalam memasukan data set hasil *Stopwords* dan *Topic Model* yang telah dibuat penulis pada *excel*.



Gambar 3.15: Operator Read Excel

4. Pada *Parameters*, pilih *Import Configuration Wizard*, pilih data Kamus Data *Stopwords* & data *training.xlsx*, ganti *Attribute* kolom sentimen menjadi *Label*, *Finish*

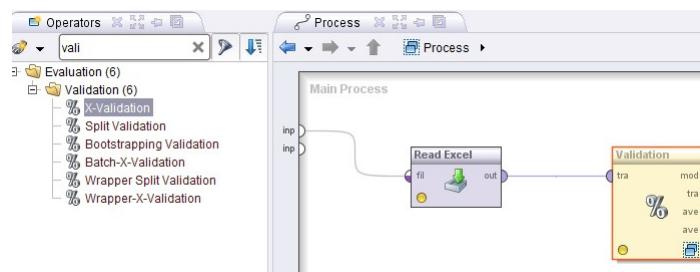


Gambar 3.16: Icon Import Configuration Wizard

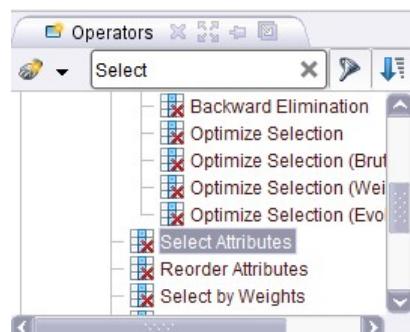
Tabel 3.10: Tipe Atribut Data Import

Kata	Pos Tag	Freq	Sent
polyno...	polyno...	integer	binomi...
attribute	attribute	attribute	label
sangat	adv	17	positif
nyaman	adj	17	positif
bersih	adj	16	positif
baik	adj	10	positif
bekas	n	10	negatif
ada	v	9	positif
layan	v	9	positif
fasilitas	n	9	positif
ramah	adj	9	positif
lennkan	arti	7	nnositif

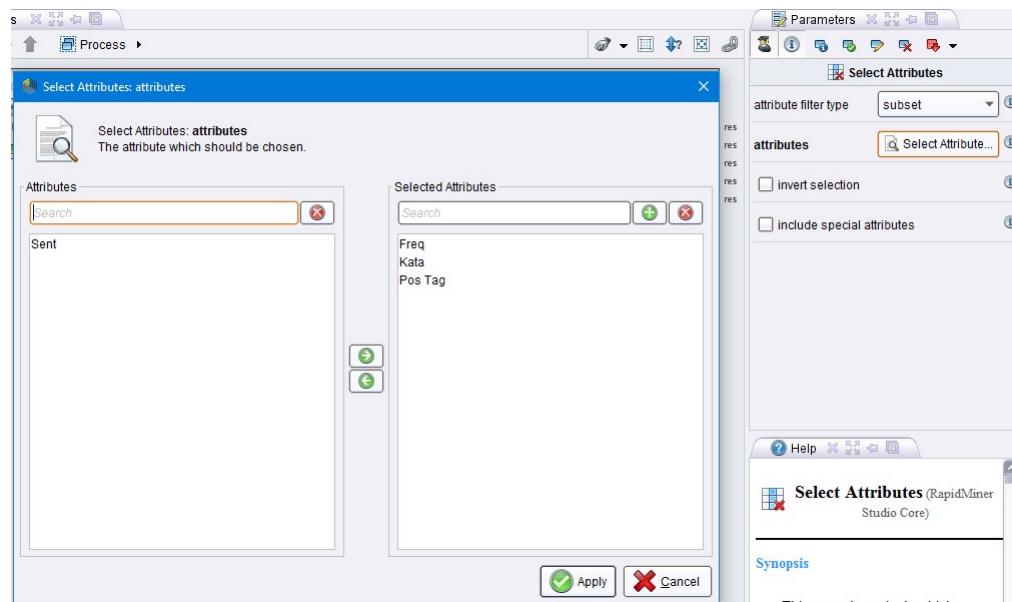
5. Pada *Operators*, pilih *Validation* kemudian *X-Validation*
6. Operator *X-Validation* adalah operator bersarang. Ini memiliki dua subproses yaitu pelatihan (*Training*) dan pengujian (*Testing*). Subproses pelatihan digunakan untuk melatih model. Model yang dilatih kemudian diterapkan pada subproses pengujian. Kinerja model juga diukur selama tahap pengujian.

Gambar 3.17: Penambahan Operator *Validation*

7. Pada *Operators*, pilih *Validation* kemudian tambahkan *Select Attribute*

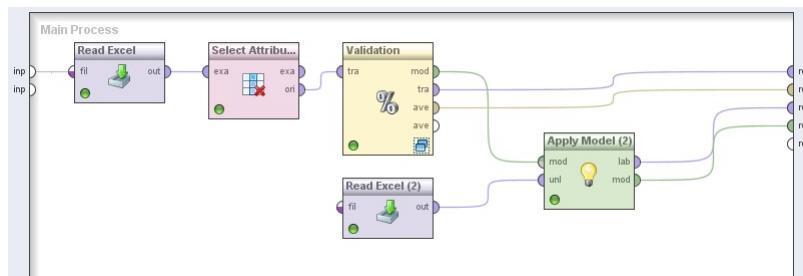
Gambar 3.18: Penambahan Operator *Select Attributes*

8. Click pada *Select Attributes*
9. Kemudian ubah attribute Filter type menjadi subset, dan drag attribute selain Sent (positif / negatif)



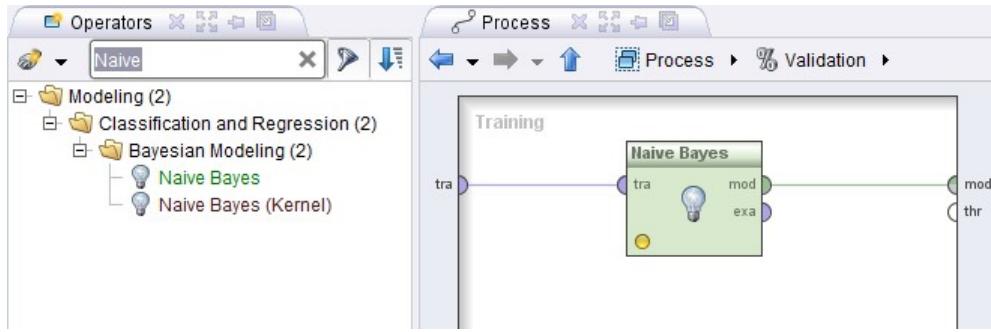
Gambar 3.19: Konfigurasi Operator *Select Attribute*

10. Tambahan Operators *Read excel* dengan input data yang sama dan Hubungkan garis dari mod (model), tra (training), ave (average) pada *Validation* ke res (result set) seperti di bawah ini:



Gambar 3.20: Proses Relasi

11. Double click pada *Validation*
12. Pada Operators, pilih algoritma *Naïve Bayes* kemudian drag ke layout *Training* dan hubungkan proses:

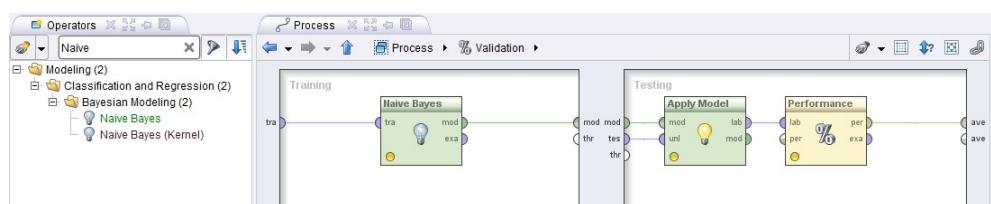


Gambar 3.21: Import Algoritma Naïve Bayes

13. Drag operator “Apply Model” dan “Performance” ke layout Testing:
14. *Apply Model* atau Terapkan Model merupakan operator yang menerapkan model yang sudah dipelajari atau dilatih pada *ExampleSet* (data awal: excel). Semua parameter yang dibutuhkan disimpan dalam objek model. Wajib bahwa kedua *ExampleSet* harus memiliki jumlah, urutan, jenis dan peran atribut. Jika sifat data meta dari *ExampleSet* ini tidak konsisten, hal ini dapat menyebabkan kesalahan serius.

ExampleSet

1. Sedangkan, *Performance* atau Kinerja merupakan operator yang digunakan untuk evaluasi kinerja. Kriteria kinerja ini ditentukan secara otomatis agar sesuai dengan jenis tugas pembelajaran.
2. Secara umum kriteria tersebut ditambahkan untuk tugas klasifikasi:
3. Ketepatan, Presisi/level keberagaman, Penarikan, Optimis, Netral, Pe-simis, *Statistik kappa* (mengukur kesepakatan/keeratan antar-penilai untuk kategoris item).



Gambar 3.22: Penambahan Operator Apply Model & Performance

4. Klik Run untuk mengetahui semua hasil.

Berikut adalah hasil olah data dari aplikasi Rapidminer:

- (a) Berikut hasil akurasi yang di dapat dari evaluasi kinerja model klasifikasi yang berdasarkan pada kemampuan akurasi prediktif suatu model. Akurasi prediktif merupakan parameter untuk mengukur ketepatan aturan klasifikasi yang dihasilkan dalam test set berdasarkan atribut yang ada ke dalam kelasnya.

Tabel 3.11: Perhitungan Naive Bayes pada Ulasan Hotel Horison

	true positif	true negatif	class precision
pred. positif	11	3	78.57%
pred. negatif	8	1	11.11%
class recall	57.89%	25.00%	

Tabel 3.12: Perhitungan *Naive Bayes* pada Ulasan Hotel Harris

	true positif	true negatif	class precision
pred. positif	14	1	93.33%
pred. negatif	12	8	40.00%
class recall	53.85%	88.89%	

Tabel 3.13: Perhitungan *Naive Bayes* pada Ulasan Hotel Red Planet

	true positif	true negatif	class precision
pred. positif	15	3	83.33%
pred. negatif	10	0	0.00%
class recall	60.00%	0.00%	

Tabel 3.14: Perhitungan *Naive Bayes* pada Ulasan Hotel Fave

	true positif	true negatif	class precision
pred. positif	6	0	100.00%
pred. negatif	16	6	27.27%
class recall	27.27%	100.00%	

Tabel 3.15: Perhitungan *Naive Bayes* pada Ulasan Hotel Santika Priemere

	true positif	true negatif	class precision
pred. positif	17	7	70.83%
pred. negatif	8	5	38.46%
class recall	68.00%	41.67%	

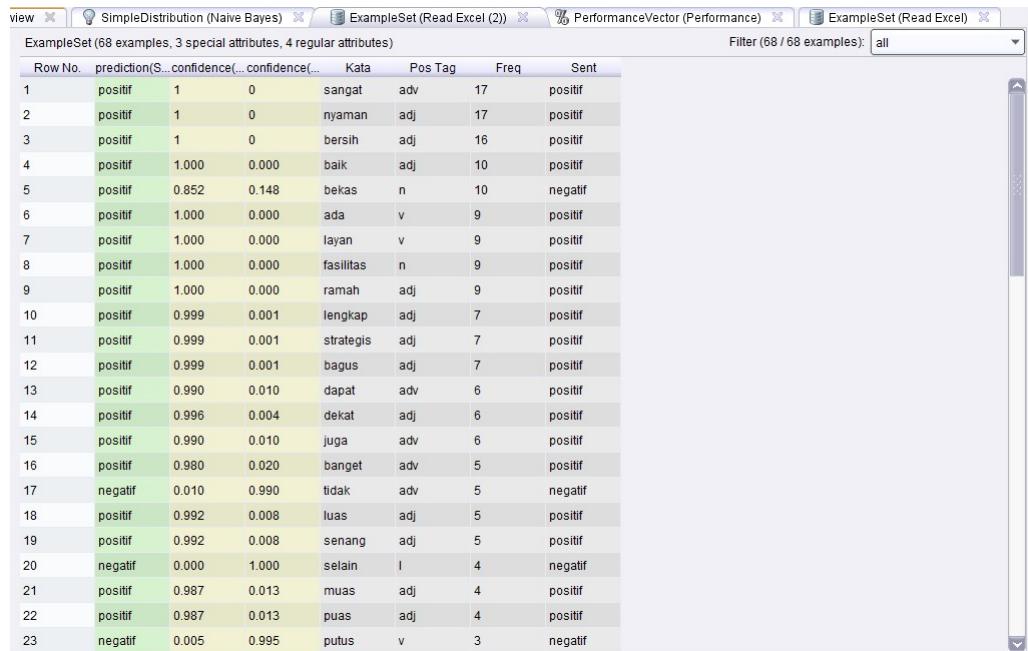
Tabel 3.16: Akurasi Total Perhitungan *Naïve Bayes* 63,24%

accuracy: 64.46% +/- 16.22% (mikro: 64.79%)			
	true positif	true negatif	class precision
pred. positif	37	15	71.15%
pred. negatif	10	9	47.37%
class recall	78.72%	37.50%	

Keterangan:

Adapun hasil akurasi dapat dihitung dengan rumus berikut:

- $Recall = \frac{TP}{(TP+FN)}$
 $Recall$ untuk sentimen positif = $\frac{37}{(37+10)} = 78.72\%$
 $Recall$ untuk sentimen negatif = $\frac{9}{(9+15)} = 37.50\%$
- $Precision = \frac{TP}{(TP+FP)}$
 $Precision$ untuk sentimen positif = $\frac{37}{(37+15)} = 71.15\%$
 $Precision$ untuk sentimen negatif = $\frac{9}{(9+10)} = 47.37\%$
- $Accuracy = \frac{(TP+TN)}{(P+N)}$
 $Accuracy = \frac{(37+9)}{(37+10)+(15+9)} = 64.788\%$



Row No.	prediction(S...confidence(...confidence(...	Kata	Pos Tag	Freq	Sent		
1	positif	1	0	sangat	adv	17	positif
2	positif	1	0	nyaman	adj	17	positif
3	positif	1	0	bersih	adj	16	positif
4	positif	1.000	0.000	baik	adj	10	positif
5	positif	0.852	0.148	bekas	n	10	negatif
6	positif	1.000	0.000	ada	v	9	positif
7	positif	1.000	0.000	layan	v	9	positif
8	positif	1.000	0.000	fasilitas	n	9	positif
9	positif	1.000	0.000	ramah	adj	9	positif
10	positif	0.999	0.001	lengkap	adj	7	positif
11	positif	0.999	0.001	strategis	adj	7	positif
12	positif	0.999	0.001	bagus	adj	7	positif
13	positif	0.990	0.010	dapat	adv	6	positif
14	positif	0.996	0.004	dekat	adj	6	positif
15	positif	0.990	0.010	juga	adv	6	positif
16	positif	0.980	0.020	banget	adv	5	positif
17	negatif	0.010	0.990	tidak	adv	5	negatif
18	positif	0.992	0.008	luas	adj	5	positif
19	positif	0.992	0.008	senang	adj	5	positif
20	negatif	0.000	1.000	selain	I	4	negatif
21	positif	0.987	0.013	muas	adj	4	positif
22	positif	0.987	0.013	puas	adj	4	positif
23	negatif	0.005	0.995	putus	v	3	negatif

Gambar 3.23: Statistik Penjabaran Atribut



Name	Type	Miss.	Statistics	Filter (7 / 7 attributes):
prediction prediction(Sent)	Binomial	0	Least negatif (22) Most positif (46)	Values positif (46), negatif (22)
confidence_positif confidence(positif)	Real	0	Min 0.000 Max 1 Average 0.664 Deviation 0.459	
confidence_negatif confidence(negatif)	Real	0	Min 0 Max 1.000 Average 0.336 Deviation 0.459	
Kata	Polynomial	0	Least variasi (1) Most Cuma (1)	Values Cuma (1), ada (1), ...[66 mo]
Pos Tag	Polynomial	0	Least I (1) Most adj (37)	Values adj (37), adv (12), ...[4 more]
Freq	Integer	0	Min 1 Max 17 Average 3.691 Deviation 3.837	
Sent	Binomial	0	Least negatif (23) Most positif (45)	Values positif (45), negatif (23)

Gambar 3.24: ExampleSet Data Rapidminer

Berdasarkan hasil pada kolom confidence(positif) dan confi-

dence(negatif), terlihat rata-rata probabilitas sentimen positif yaitu 0.664 (66.4%) dan rata-rata probabilitas sentimen negatif yaitu 0.336 (33.6%).

- (b) *Naïve Bayes* mengklasifikasikan 2 kelas yaitu Positif dan Negatif: Akurasi dinyatakan dalam persentase.

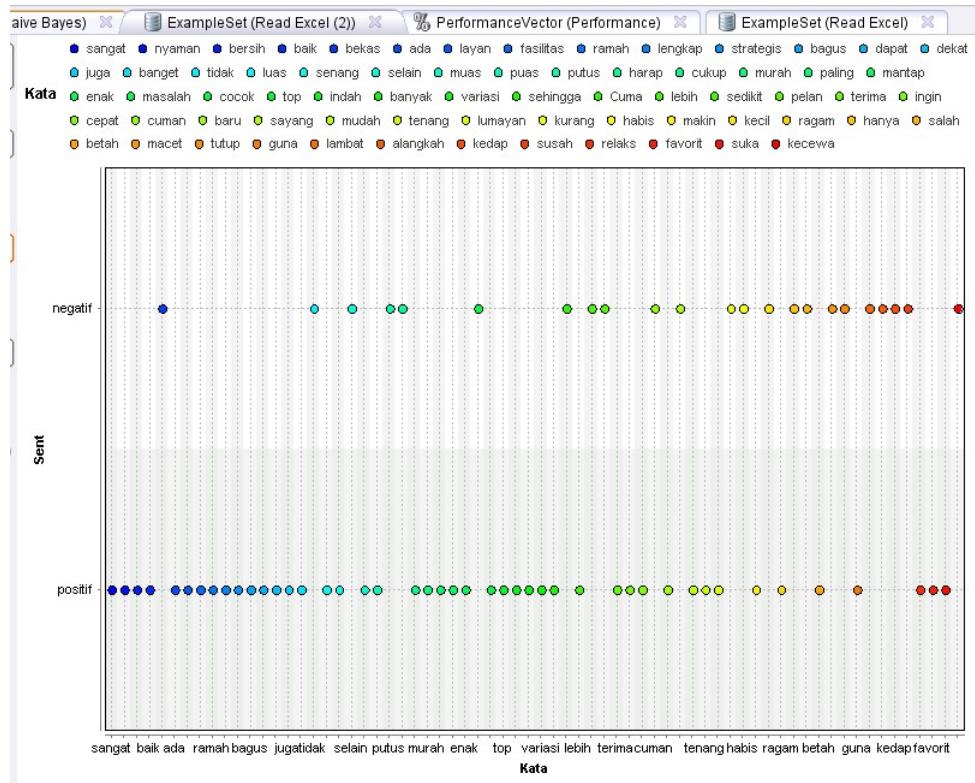
Akurasi dinyatakan dalam persentase. Semua kasus yang tercakup oleh aturan klasifikasi, diklasifikasikan dengan benar ke dalam kelas yang diprediksinya. Untuk mendapatkan nilai akurasi prediktif diperlukan perhitungan jumlah kasus yang diprediksikan dengan benar dan jumlah kasus yang diprediksikan dengan salah. Perhitungan tersebut ditabulasikan ke dalam tabel yang disebut *confusion matrix* (suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data *mining*)

PerformanceVector

```
PerformanceVector:
accuracy: 64.46% +/- 16.22% (mikro: 64.79%)
ConfusionMatrix:
True: positif negatif
positif:      37      15
negatif:      10       9
```

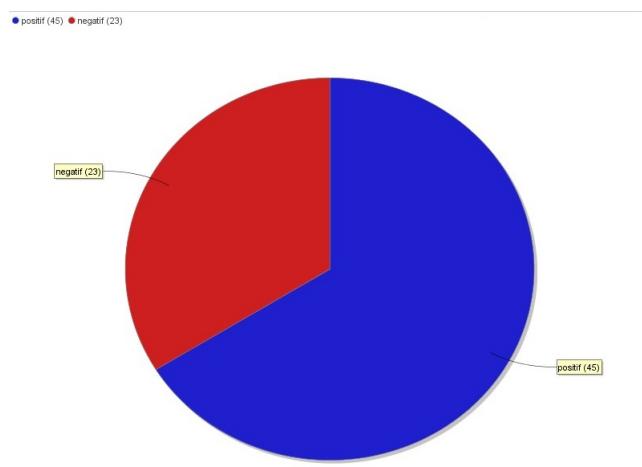
Gambar 3.25: *Performance Vector (Confusion Matrix)*

- (c) Berikut output dari diagram Scatter dari banyaknya kata dan jenis-jenisnya yang berbanding dengan sentimenal apakah kata tersebut masuk kedalam kata positif atau negatif.

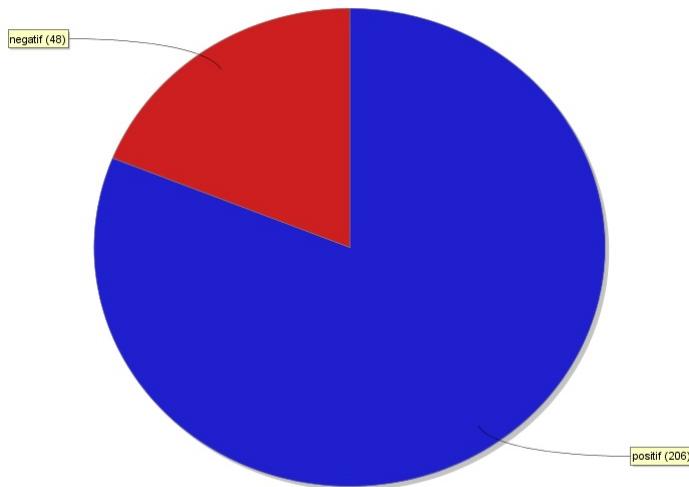


Gambar 3.26: Diagram Scatter dengan Sumbu Kata dan sumbu Sentimen

- (d) Dan berikut output dari semua kesimpulan yang di deskripsikan dalam bentuk diagram pie:

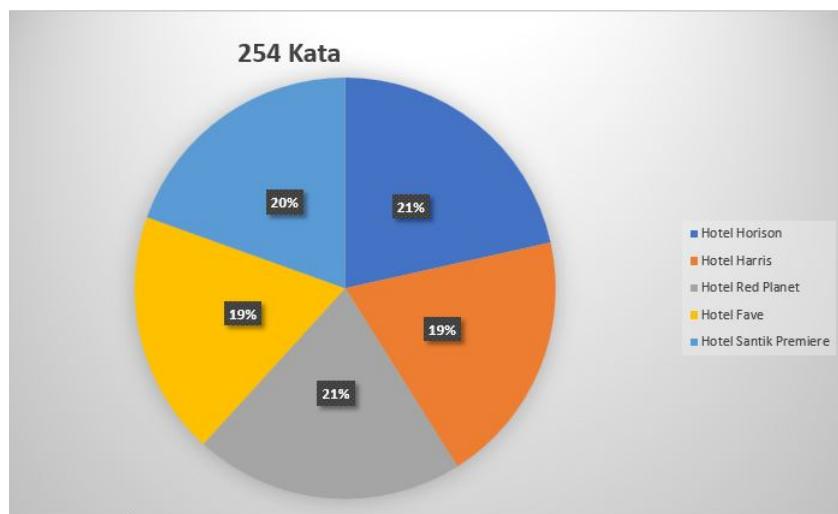


Gambar 3.27: Diagram Pie Sebelum Dikalikan dengan Frekuensi



Gambar 3.28: Diagram *Pie* Sesudah Dikalikan dengan Frekuensi

Dapat diketahui bahwa aplikasi *Rapidminer* dapat berhasil berjalan Dengan hasil Akurasi sebesar 64.79% pada perhitungan klasifikasi *Naïve Bayes*. Terdapat sentimen positif (diagram bar: Biru) dengan jumlah 206 kata dan sentimen negatif (diagram bar: Merah) dengan jumlah 48 kata. Dominasi dimiliki oleh sentimen positif.



Gambar 3.29: Diagram *Pie* Pembagian Kata Setiap Hotel

Dari total 251 kata yang diolah dengan *Rapidminer*, berikut adalah persentase ulasan tiap-tiap hotel yang telah dibagi berdasarkan jumlah total kata dengan dibagi jumlah kata tiap-tiap hotel.

- Hotel Horison = $\frac{43}{254} \times 100\% = 17\%$
Sebanyak 43 kata ulasan pada hotel Horison mempengaruhi 254 total kata sebesar 17%. Dengan sentimen positif 35 kata dan negatif 8 kata.
- Hotel Harris = $\frac{61}{254} \times 100\% = 24\%$
Sebanyak 61 kata ulasan pada hotel Harris mempengaruhi 254 total kata sebesar 24%. Dengan sentimen positif 51 kata dan negatif 10 kata.
- Hotel Red Planet = $\frac{40}{254} \times 100\% = 16\%$
Sebanyak 40 kata ulasan pada hotel Red Planet mempengaruhi 254 total kata sebesar 16%. Dengan sentimen positif 36 kata dan negatif sebesar 4 kata.
- Hotel Fave = $\frac{37}{254} \times 100\% = 14\%$
Sebanyak 47 kata ulasan pada hotel Horison mempengaruhi 254 total kata sebesar 14%. Dengan sentimen positif 31 kata dan negatif 6 kata.
- Hotel Santika Premiere = $\frac{73}{254} \times 100\% = 29\%$
Sebanyak 73 kata ulasan pada hotel Horison mempengaruhi 254 total kata sebesar 29%. Dengan sentimen positif 53 kata dan negatif 20 kata.

3.7.2 Pembuatan Aplikasi dengan Python 2.7.13

Dengan semua hasil yang didapat, penulis akan melanjutkan membuat sebuah aplikasi analisis sentimen ini menggunakan bahasa pemrograman *Python* kembali. Untuk mempermudah user/pengguna dalam mendapatkan hasil output yang lebih fleksibel pada konsep *User Interface*. Sehingga pengguna hanya perlu klik setiap nama hotel saja kemudian muncul foto beserta analisisnya. Sebelum memulai coding, install terlebih dahulu “*pip install image*” pada cmd dengan tujuan untuk menambahkan kemampuan pemrosesan gambar ke penerjemah *Python*. Library ini mendukung banyak format file, dan menyediakan kemampuan pemrosesan gambar dan grafis yang hebat.

```

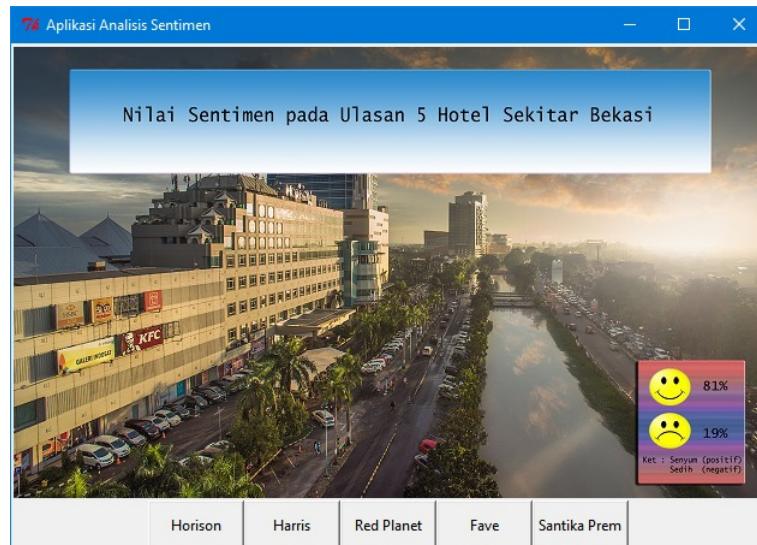
C:\Windows\system32\cmd.exe
Copyright <c> 2009 Microsoft Corporation. All rights reserved.

C:\Users\hp\pip install image
Collecting image
  Downloading image-1.5.11-py2.py3-none-any.whl
Collecting olefile (from image)
  Downloading Pillow-4.2.1-cp27m-win32.whl (1.2MB) 100% [=====] 1.2MB 49kB/s
Collecting django (from image)
  Downloading Django-1.11.4-py2.py3-none-any.whl (6.9MB) 100% [=====] 7.0MB 86kB/s
Collecting olefile (from pillow>image)
  Downloading olefile-0.8.2-py2.py3-none-any.whl (0.1MB) 100% [=====] 81kB 88kB/s
Collecting pytz (from django->image)
  Downloading pytz-2017.2-py2.py3-none-any.whl (494kB) 100% [=====] 491kB 205kB/s
Installing collected packages: olefile, pillow, pytz, django, image
  Running setup.py install for olefile ... done
Successfully installed django-1.11.4 image-1.5.11 olefile-0.8.4 pillow-4.2.1 pytz-2017.2

```

Gambar 3.30: Proses Download Library Image

Pada form utama ini merupakan tampilan awal aplikasi untuk user mendapatkan informasi dari analisis dengan mengetahui hasil dalam bentuk persen yang berada dibawah kanan dengan emoticon senyum yang berarti positif dan sedih yang berarti negatif.



Gambar 3.31: Tampilan Awal Aplikasi Sederhana

Apabila user mengklik tombol “Horison”, Maka akan muncul hasil kedua berupa foto hotel Horison beserta nilai sentimen dan alamat letak hotel itu berada:



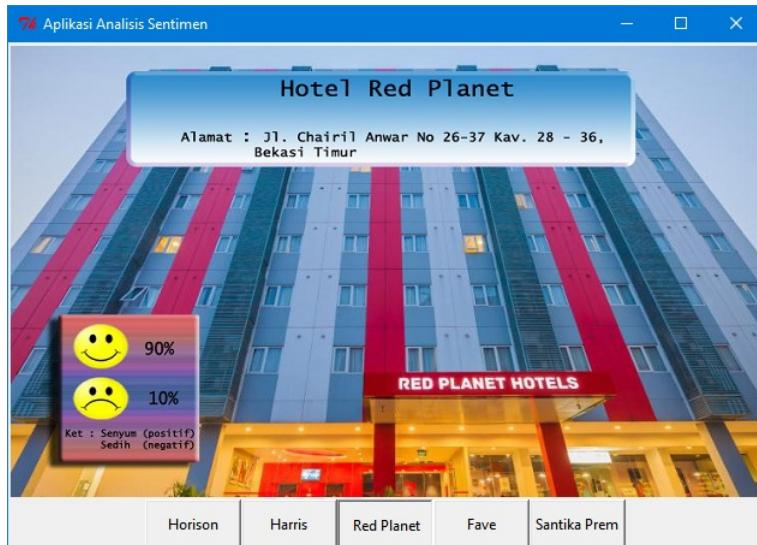
Gambar 3.32: Tampilan Output Aplikasi Hotel Horison

Apabila user mengklik tombol “Harris”, Maka akan muncul hasil ketiga berupa foto hotel Harris beserta nilai sentimen dan alamat letak hotel itu berada:



Gambar 3.33: Tampilan Output Aplikasi Hotel Harris

Apabila user mengklik tombol “Red Planet”, Maka akan muncul hasil keempat berupa foto hotel Red Planet beserta nilai sentimen dan alamat letak hotel itu berada:



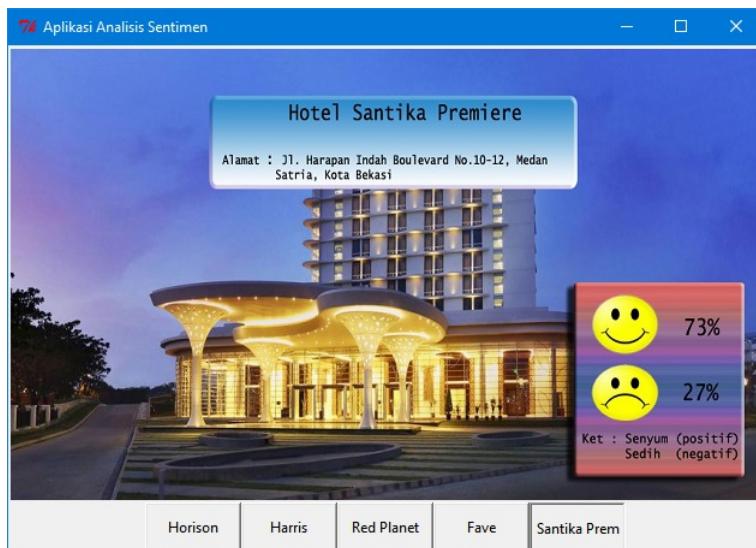
Gambar 3.34: Tampilan Output Aplikasi Hotel Red Planet

Apabila user mengklik tombol “Fave”, Maka akan muncul hasil kelima berupa foto hotel Fave beserta nilai sentimen dan alamat letak hotel itu berada:



Gambar 3.35: Tampilan Output Aplikasi Hotel Fave

Apabila user mengklik tombol “Santika Prem”, Maka akan muncul hasil keenam berupa foto hotel Santika Premiere beserta nilai sentimen dan alamat letak hotel itu berada:



Gambar 3.36: Tampilan Output Aplikasi Hotel Santika Premiere

Berdasarkan dari hasil pengujian di atas, “Analisis Sentimen pada ulasan Hotel yang berada disekitar Bekasi (Horison,Fave,dll) yang dengan Tekst bahasa Indonesia” dapat berjalan dengan sangat baik pada aplikasi *Rapidminer*. Dengan hasil Akurasi sebesar 64.79% pada perhitungan klasifikasi *Naïve Bayes*. Terdapat sentimen positif dengan jumlah 206 kata dan sentimen negatif dengan jumlah 48 kata dalam 156 dokumen. Dominasi dimiliki oleh sentimen positif. Sementara dari total 254 kata, hotel Horison mempengaruhi sebesar 17%, hotel Harris mempengaruhi sebesar 24%, hotel Red Planet mempengaruhi sebesar 16%, hotel Fave mempengaruhi sebesar 14%, hotel Santika Premiere mempengaruhi sebesar 29%.

Bab 4

PENUTUP

4.1 KESIMPULAN

Berdasarkan penulisan ini, dapat disimpulkan bahwa pengolahan data yang sudah dilakukan aplikasi *Rapidminer* dapat berjalan terhadap data hasil *pre-processing* yang mengambil dari situs pemesanan hotel *online* Pegipegi.com pada 5 hotel di satu wilayah di bulan Mei-Juli tahun 2017, sebanyak 156 dokumen ulasan berbahasa Indonesia. Terbukti dengan metode klasifikasi *Naïve Bayes* dapat diterapkan pada proses analisis sentimen dengan menghasilkan akurasi prediktif sebesar 64.79% dengan menunjukan persentase kata sebesar 81% (206 kata) pada sentimen positif, serta 19% (48 kata) pada sentimen negatif dari total 254 kata dalam 156 dokumen. Perhitungan tiap-tiap hotel menghasilkan nilai klasifikasi dokumen ulasan sentimen positif diatas 72% dan nilai klasifikasi dokumen ulasan sentimen negatif dibawah 28%. Sehingga dari hasil perhitungan yang dilakukan oleh aplikasi *Rapidminer* dengan klasifikasi *Naive Bayes* didominasi oleh sentimen positif.

4.2 SARAN

Rancangan analisis sentimen terhadap situs Pemesanan Hotel Online ini masih memungkinkan untuk dilakukan pengembangan dan modifikasi. Pengembangan misalnya penentuan kata tersebut masuk kedalam kelas kata positif atau negatif tidak secara manual, melainkan dengan sistem yang sudah automatis sehingga data lebih akurat. Semoga analisis ini dapat berguna dan lebih baik lagi dalam pengembangan nantinya.

DAFTAR PUSTAKA

- [1] Ac, N. (2009). Keoptimalan naïve bayes dalam klasifikasi. *Seminar Program Studi Ilmu Komputer*. (Diakses pada 11 Juli 2017).
- [2] Aldi, N. and Muslim, M. A. (2016). Analisis sentimen pada opini mahasiswa menggunakan natural language processing.seminar nasional ilmu komputer. *Seminar Nasional Ilmu Komputer 2016*. ISBN: 978-602-1034-40-8.
- [3] Alter's, S. (2002). *Information Systems: Foundation of E-Business*. Prentice Hall.
- [4] Andi, N. H. (2015). Analisis sentimen terhadap wacana politik pada media masa online menggunakan algoritma support vector machine dan naive bayes. *Jurnal Elektronik Sistem Informasi dan Komputer*, 1(1).
- [5] Berry, M. and Kogan, J. (2010). *Text Mining Application and theory*. United Kingdom : Wiley.
- [6] Bolshakov, I. and Gelbuk, A. (2004). *Computational Linguistics*. Intiuitivo Politenico Nacional. 1st ed.
- [7] Carroll, W. J. and Sileo, L. (2007). *Chains gain ground online: Hotels have much to celebrate (electronic version)*.
- [8] Daniel Pakpahan, H. W. (2014). Aplikasi opinion mining dengan algoritma naive bayes untuk menilai berita online. *Program Studi Teknik Informatika Politeknik Negeri Batam*.
- [9] Feldman, R. and Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York : Cambridge University Press. 1st ed.
- [10] Gaptex (Diakses pada 20 Juli 2017). 5 situs booking hotel online terpercaya. <http://gaptex.com/5-situs-booking-hotel-online-terpercaya/>.

- [11] Han, J., Kamber, and Micheline (2006). *Data Mining : Concept and Techniques Second Edition*. Morgan Kaufmann Publishers.
- [12] Imambi, S. S. and T., S. (2011). Pre-processing of medical documents and reducing dimensionality. *Adv. Comput. An Int. J. (ACIJ)*, 2(5):15–24.
- [13] Juen, L., Kencana, I. P. E. N., and Oka, T. B. (2014). Analisis sentimen menggunakan metode naÃ¯ve bayes ckassufuer dengan seleksi fitur chi-square. *E-Jurnal Matematika*, 3(3). ISSN: 2303-1751.
- [14] Karaa, W. B. A. (2013). A new stemmer to improve information retrieval. *International Journal of Network Security and Its Applications (IJNSA)*, 5(4):143–154.
- [15] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synth Lect Human Lang Technol.
- [16] Martinet, A. (1987). *Ilmu Bahasa Pengantar*. Kanisius.
- [17] Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*.
- [18] Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan*. Yogyakarta : Andi.
- [19] Sadeh, I. and Zion, M. (2011). Which type of inquiry project do high school biology. *Student Prefer: Open or Guide*.
- [20] Shadiq, M. A. (2011). The digital media and habits attitudes of south east asianconsumers. *Seminar Program Studi Ilmu Komputer*.
- [21] van Rossum, J. G. (1991). *Linking a Stub Generator (AIL) to a Prototyping Language (Python)*.
- [22] Wikipedia (Diakses pada 20 Juli 2017). Tokenisasi. <https://id.wikipedia.org/wiki/Tokenisasi>.

LAMPIRAN LISTING PROGRAM

Syntax Pre-processing

```
#memanggil module urllib untuk scrapping
import urllib
#memanggil module reguler expression
import re
#memanggil module warning agar menolak warning ketika
memanggil corpora dari module gensim

#Hotel Horison
#htmlfile1 = variabel yg difungsikan untuk membuka url
    hotel Horison ultima bekasi dari web pegipegi
htmlfile1 = urllib.urlopen("https://www.pegipegi.com/
    hotel/bekasi/hotel_horison_ultima_bekasi_989768/")
#htmltext1 = variabel yg difungsikan untuk membaca
    htmlfile1
htmltext1 = htmlfile1.read()
#regex1 = variabel yg difungsikan untuk menaruh link
    html tempat komentar diambil dengan (.+?) bermaksud
    mengambil semua komentar per div disetiap class=
    ContentComment.
regex1='<div class="contentComment" itemprop="'
    description">(.+?)</div>'
#pattern1 = variabel yg difungsikan untuk mengcompile
    reguler expression regex1
pattern1=re.compile(regex1)
#hasil1 = variabel yg difungsikan untuk mencari semua
    komentar pada web var htmlfil1 yg beralamatkan var
    regex1
```

```
hasil1=re.findall(pattern1 ,htmltext1)

#Hotel Red Planet
#htmlfile1 adalah variabel yg difungsikan untuk membuka
url hotel Red Planet bekasi dari web pegipegi
htmlfile2=urllib.urlopen("https://www.pegipegi.com/
    hotel/bekasi/red_planet_bekasi_959430/")
htmltext2=htmlfile2.read()
regex2='<div class="contentComment" _itemprop="'
    description">(.+)</div>'
pattern2=re.compile(regex2)
hasil2=re.findall(pattern2 ,htmltext2)

#Hotel Harris
#htmlfile1 adalah variabel yg difungsikan untuk membuka
url hotel Harris bekasi dari web pegipegi
htmlfile3=urllib.urlopen("https://www.pegipegi.com/
    hotel/bekasi/hotel_harris__bekasi_945858/")
htmltext3=htmlfile3.read()
regex3='<div class="contentComment" _itemprop="'
    description">(.+)</div>'
pattern3=re.compile(regex3)
hasil3=re.findall(pattern3 ,htmltext3)

#Hotel Santika Priemiere
#htmlfile1 adalah variabel yg difungsikan untuk membuka
url hotel Santika Priemere bekasi dari web pegipegi
htmlfile4=urllib.urlopen("https://www.pegipegi.com/
    hotel/bekasi/hotel_santika_premiere_bekasi_906014/")
htmltext4=htmlfile4.read()
regex4='<div class="contentComment" _itemprop="'
    description">(.+)</div>'
pattern4=re.compile(regex4)
hasil4=re.findall(pattern4 ,htmltext4)

#Hotel Fave
#htmlfile1 adalah variabel yg difungsikan untuk membuka
```

```

url hotel Fave bekasi dari web pegipegi
htmlfile5=urllib.urlopen("https://www.pegipegi.com/
    hotel/bekasi/favehotel_ahmad_yani_bekasi_948432/")
htmltext5=htmlfile5.read()
regex5='<div class="contentComment" _itemprop="
    description">(.+)</div>'
pattern5=re.compile(regex5)
hasil5=re.findall(pattern5 , htmltext5)

#Total = Variabel yg difungsikan untuk menampung
#komentar dari semua hasil1–hasil5 yg sudah di
#scrapping
hasil=(hasil1+hasil2+hasil3+hasil4+hasil5)

#Tokenize
#memanggil paket dari regexptokenizer dari modul nltk.
<regexptokenizer = reguler expression dari paket
nltk untuk tokenizer
from nltk.tokenize import RegexpTokenizer
#Token = variabel yg difungsikan untuk mengubah tipe
#data variabel Total menjadi string.
Token      = str(hasil)
#Tokenizer = variabel untuk mendefinisikan regular
#expression
tokenizer = RegexpTokenizer(r '\w+')
#pemanggilan fungsi tokenize beserta output
Tokens     = tokenizer.tokenize(Token)

#Stemming
#import paket sastrawi
from Sastrawi.Stemmer.StemmerFactory import
    StemmerFactory
#pembuatan stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

```

```

#untuk memasukan data
Stemm = str(Tokens)
#pemanggilan fungsi stemming beserta output
Stemming = stemmer.stem(Stemm)

#Stopwords
#import stopwords dari library nltk
from nltk.corpus import stopwords
#var stopw untuk menampung stopwords file 'kbbi'
stopw = set(stopwords.words('kbbi'))
#var datasw menguhb
datasw = str(Stemming)
SWindo = [i for i in datasw.lower().split() if i in
stopw]

#Counterlist
from collections import Counter
list = SWindo
#untuk menghitung banyak nya frekuensi dari kemunculan
kata
counts = Counter(list)
print(counts)

```

Syntax Aplikasi Sederhana

```

import Tkinter as tk
from PIL import ImageTk, Image
import tkMessageBox

def horison(panel):
    path = "3.34.jpg"
    img = ImageTk.PhotoImage(Image.open(path))
    panel.configure(image=img)
    panel.image = img

```

```

def harris(panel):
    path = "3.35.jpg"
    img = ImageTk.PhotoImage(Image.open(path))
    panel.configure(image=img)
    panel.image = img

def redplanet(panel):
    path = "3.36.jpg"
    img = ImageTk.PhotoImage(Image.open(path))
    panel.configure(image=img)
    panel.image = img

def fave(panel):
    path = "3.37.jpg"
    img = ImageTk.PhotoImage(Image.open(path))
    panel.configure(image=img)
    panel.image = img

def santika(panel):
    path = "3.38.jpg"
    img = ImageTk.PhotoImage(Image.open(path))
    panel.configure(image=img)
    panel.image = img

#Membuat main window
window = tk.Tk()
window.title("Aplikasi_Analisis_Sentimen",)

#Membagi frame untuk gambar (diatas) dan tombol ( dibawah)
top = tk.Frame(window)
top.pack(side="top")
bottom = tk.Frame(window)
bottom.pack(side="bottom")

#Memasang gambar

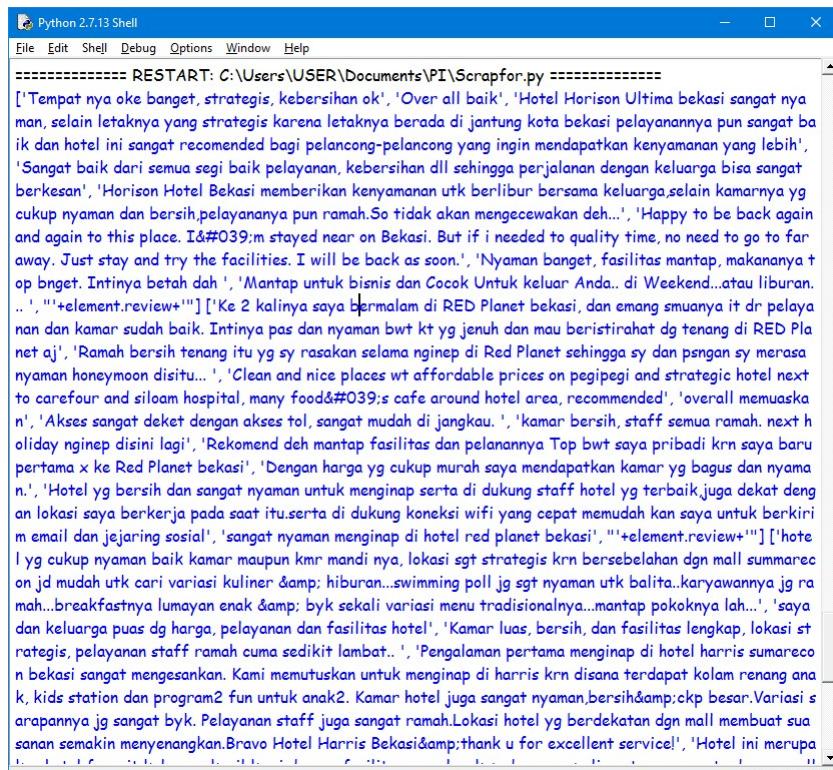
```

```
path = "3.33.jpg"
img = ImageTk.PhotoImage(Image.open(path))
panel = tk.Label(window, image = img)
panel.image = img
panel.pack(side = "top", fill = "both", expand = "yes")

#Membuat tombol
button1 = tk.Button(window, text="Horison", width=10,
                    height=2, command=lambda: horison(panel))
button1.pack(in_=bottom, side="left")
button2 = tk.Button(window, text="Santika_Prem", width
                    =10, height=2, command=lambda: santika(panel))
button2.pack(in_=bottom, side="right")
button3 = tk.Button(window, text="Fave", width=10,
                    height=2, command=lambda: fave(panel))
button3.pack(in_=bottom, side="right")
button4= tk.Button(window, text="Red_Planet", width=10,
                    height=2, command=lambda: redplanet(panel))
button4.pack(in_=bottom, side="right")
button5 = tk.Button(window, text="Harris", width=10,
                    height=2, command=lambda: harris(panel))
button5.pack(in_=bottom, side="bottom")

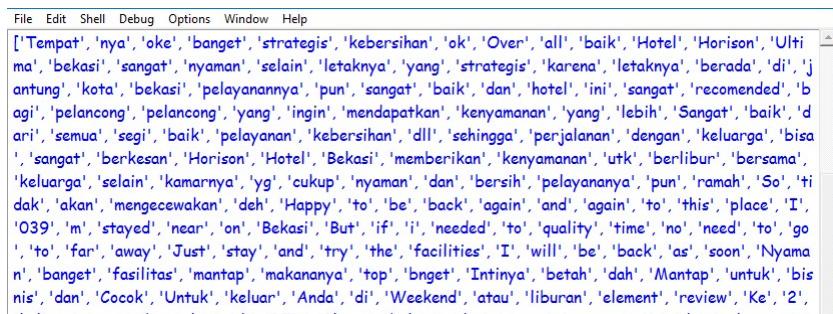
#Menjalankan GUI
window.mainloop()
```

LAMPIRAN OUTPUT PROGRAM



```
Python 2.7.13 Shell
File Edit Shell Debug Options Window Help
=====
RESTART: C:\Users\USER\Documents\PI\Scrapfor.py =====
['Tempat nya oke banget, strategis, kebersihan ok', 'Over all baik', 'Hotel Horison Ultima bekasi sangat nya man, selain letaknya yang strategis karena letaknya berada di jantung kota bekasi pelayanannya pun sangat baik dan hotel ini sangat recomended bagi pelancong-pelancong yang ingin mendapatkan kenyamanan yang lebih', 'Sangat baik dari semua segi baik pelayanan, kebersihan dll sehingga perjalanan dengan keluarga bisa sangat berkesan', 'Horison Hotel Bekasi memberikan kenyamanan utk berlibur bersama keluarga,selain kamarnya yg cukup nyaman dan bersih,pelayanannya pun ramah.So tidak akan mengecewakan deh...', 'Happy to be back again and again to this place. I'm stayed near on Bekasi. But if i need to quality time, no need to go to far away. Just stay and try the facilities. I will be back as soon.', 'Nyaman banget, fasilitas mantap, makanannya t op bnget. Intinya betah dah', 'Mantap untuk bisnis dan Cocok Untuk Keluar Anda.. di Weekend..atau liburan..', "+element.review+"] ['Ke 2 kalinya saya bermalam di RED Planet bekasi, dan emang smuanya it dr pelayan dan kamar sudah baik. Intinya pas dan nyaman bwt kt yg jenuh dan mau beristirahat dg tenang di RED Planet aj', 'Ramah bersih tenang itu yg sy rasakan selama nginep di Red Planet sehingga sy dan psngan sy merasa nyaman honeymoon disitu...', 'Clean and nice places wt affordable prices on pegipegi and strategic hotel next to carefour and siloam hospital, many food&#039;s cafe around hotel area, recommended', 'overall memuaska n', 'Akses sangat deket dengan akses tol, sangat mudah di jangkau.', 'Kamar bersih, staff semua ramah. next h oliday nginep disini lagi', 'Rekomend deh mantap fasilitas dan pelanannya Top bwt saya pribadi krn saya baru pertama x ke Red Planet bekasi', 'Dengan harga yg cukup murah saya mendapatkan kamar yg bagus dan nyaman.', 'Hotel yg bersih dan sangat nyaman untuk menginap serta di dukung staff hotel yg terbaik,juga dekat deng an lokasi saya berkerja pada saat itu.serta di dukung koneksi wifi yang cepat memudah kan saya untuk berkirim email dan jejaring sosial', 'sangat nyaman menginap di hotel red planet bekasi', "+element.review+"] ['hote l yg cukup nyaman baik kamar maupun kmr mandi nya, lokasi sgt strategis krn bersebelahan dgn mall summarecon jd mudah utk cari variasi kuliner & hiburan..swimming poll jg sgt nyaman utk balita..karyawannya jg ramah..breakfastnya lumayan enak & byk sekali variasi menu tradisionalnya...mantap pokoknya lah...', 'saya dan keluarga puas dg harga, pelayanan dan fasilitas hotel', 'Kamar luas, bersih, dan fasilitas lengkap, lokasi strategis, pelayanan staff ramah cuma sedikit lambat..', 'Pengalaman pertama menginap di hotel harris sumarecon bekasi sangat mengesankan. Kami memutuskan untuk menginap di harris krn disana terdapat kolam renang anak, kids station dan program2 fun untuk anak2. Kamar hotel juga sangat nyaman,bersih&ckp besar.Variasi sarapannya jg sangat byk. Pelayanan staff juga sangat ramah.Lokasi hotel yg berdekatan dgn mall membuat suasana semakin menyenangkan.Bravo Hotel Harris Bekasi&thank u for excellent service!', 'Hotel ini merupakan salah satu hotel yang sangat nyaman dan bersih di bekasi', '+element.review+"] ['Ke 2 kalinya saya bermalam di RED Planet bekasi, dan emang smuanya it dr pelayan dan kamar sudah baik. Intinya pas dan nyaman bwt kt yg jenuh dan mau beristirahat dg tenang di RED Planet aj', 'Ramah bersih tenang itu yg sy rasakan selama nginep di Red Planet sehingga sy dan psngan sy merasa nyaman honeymoon disitu...', 'Clean and nice places wt affordable prices on pegipegi and strategic hotel next to carefour and siloam hospital, many food&#039;s cafe around hotel area, recommended', 'overall memuaska n', 'Akses sangat deket dengan akses tol, sangat mudah di jangkau.', 'Kamar bersih, staff semua ramah. next holiday nginep disini lagi', 'Rekomend deh mantap fasilitas dan pelanannya Top bwt saya pribadi krn saya baru pertama x ke Red Planet bekasi', 'Dengan harga yg cukup murah saya mendapatkan kamar yg bagus dan nyaman.', 'Hotel yg bersih dan sangat nyaman untuk menginap serta di dukung staff hotel yg terbaik,juga dekat deng an lokasi saya berkerja pada saat itu.serta di dukung koneksi wifi yang cepat memudah kan saya untuk berkirim email dan jejaring sosial', 'sangat nyaman menginap di hotel red planet bekasi', "+element.review+"] ['hote l yg cukup nyaman baik kamar maupun kmr mandi nya, lokasi sgt strategis krn bersebelahan dgn mall summarecon jd mudah utk cari variasi kuliner & hiburan..swimming poll jg sgt nyaman utk balita..karyawannya jg ramah..breakfastnya lumayan enak & byk sekali variasi menu tradisionalnya...mantap pokoknya lah...', 'saya dan keluarga puas dg harga, pelayanan dan fasilitas hotel', 'Kamar luas, bersih, dan fasilitas lengkap, lokasi strategis, pelayanan staff ramah cuma sedikit lambat..', 'Pengalaman pertama menginap di hotel harris sumarecon bekasi sangat mengesankan. Kami memutuskan untuk menginap di harris krn disana terdapat kolam renang anak, kids station dan program2 fun untuk anak2. Kamar hotel juga sangat nyaman,bersih&ckp besar.Variasi sarapannya jg sangat byk. Pelayanan staff juga sangat ramah.Lokasi hotel yg berdekatan dgn mall membuat suasana semakin menyenangkan.Bravo Hotel Harris Bekasi&thank u for excellent service!', 'Hotel ini merupakan salah satu hotel yang sangat nyaman dan bersih di bekasi', '+element.review+"] ['Ke 2 kalinya saya bermalam di RED Planet bekasi, dan emang smuanya it dr pelayan dan kamar sudah baik. Intinya pas dan nyaman bwt kt yg jenuh dan mau beristirahat dg tenang di RED Planet aj', 'Ramah bersih tenang itu yg sy rasakan selama nginep di Red Planet sehingga sy dan psngan sy merasa nyaman honeymoon disitu...', 'Clean and nice places wt affordable prices on pegipegi and strategic hotel next to carefour and siloam hospital, many food&#039;s cafe around hotel area, recommended', 'overall memuaska n', 'Akses sangat deket dengan akses tol, sangat mudah di jangkau.', 'Kamar bersih, staff semua ramah. next holiday nginep disini lagi', 'Rekomend deh mantap fasilitas dan pelanannya Top bwt saya pribadi krn saya baru pertama x ke Red Planet bekasi', 'Dengan harga yg cukup murah saya mendapatkan kamar yg bagus dan nyaman.', 'Hotel yg bersih dan sangat nyaman untuk menginap serta di dukung staff hotel yg terbaik,juga dekat deng an lokasi saya berkerja pada saat itu.serta di dukung koneksi wifi yang cepat memudah kan saya untuk berkirim email dan jejaring sosial', 'sangat nyaman menginap di hotel red planet bekasi', "+element.review+"] ['hote l yg cukup nyaman baik kamar maupun kmr mandi nya, lokasi sgt strategis krn bersebelahan dgn mall summarecon jd mudah utk cari variasi kuliner & hiburan..swimming poll jg sgt nyaman utk balita..karyawannya jg ramah..breakfastnya lumayan enak & byk sekali variasi menu tradisionalnya...mantap pokoknya lah...', 'saya dan keluarga puas dg harga, pelayanan dan fasilitas hotel', 'Kamar luas, bersih, dan fasilitas lengkap, lokasi strategis, pelayanan staff ramah cuma sedikit lambat..', 'Pengalaman pertama menginap di hotel harris sumarecon bekasi sangat mengesankan. Kami memutuskan untuk menginap di harris krn disana terdapat kolam renang anak, kids station dan program2 fun untuk anak2. Kamar hotel juga sangat nyaman,bersih&ckp besar.Variasi sarapannya jg sangat byk. Pelayanan staff juga sangat ramah.Lokasi hotel yg berdekatan dgn mall membuat suasana semakin menyenangkan.Bravo Hotel Harris Bekasi&thank u for excellent service!', 'Hotel ini merupakan salah satu hotel yang sangat nyaman dan bersih di bekasi', '+element.review+"] ['Ke 2 kalinya saya bermalam di RED Planet bekasi, dan emang smuanya it dr pelayan dan kamar sudah baik. Intinya pas dan nyaman bwt kt yg jenuh dan mau beristirahat dg tenang di RED Planet aj', 'Ramah bersih tenang itu yg sy rasakan selama nginep di Red Planet sehingga sy dan psngan sy merasa nyaman honeymoon disitu...', 'Clean and nice places wt affordable prices on pegipegi and strategic hotel next to carefour and siloam hospital, many food&#039;s cafe around hotel area, recommended', 'overall memuaska n', 'Akses sangat deket dengan akses tol, sangat mudah di jangkau.', 'Kamar bersih, staff semua ramah. next holiday nginep disini lagi', 'Rekomend deh mantap fasilitas dan pelanannya Top bwt saya pribadi krn saya baru pertama x ke Red Planet bekasi', 'Dengan harga yg cukup murah saya mendapatkan kamar yg bagus dan nyaman.', 'Hotel yg bersih dan sangat nyaman untuk menginap serta di dukung staff hotel yg terbaik,juga dekat deng an lokasi saya berkerja pada saat itu.serta di dukung koneksi wifi yang cepat memudah kan saya untuk berkirim email dan jejaring sosial', 'sangat nyaman menginap di hotel red planet bekasi', "+element.review+"] ['hote l yg cukup nyaman baik kamar maupun kmr mandi nya, lokasi sgt strategis krn bersebelahan dgn mall summarecon jd mudah utk cari variasi kuliner & hiburan..swimming poll jg sgt nyaman utk balita..karyawannya jg ramah..breakfastnya lumayan enak & byk sekali variasi menu tradisionalnya...mantap pokoknya lah...', 'saya dan keluarga puas dg harga, pelayanan dan fasilitas hotel', 'Kamar luas, bersih, dan fasilitas lengkap, lokasi strategis, pelayanan staff ramah cuma sedikit lambat..', 'Pengalaman pertama menginap di hotel harris sumarecon bekasi sangat mengesankan. Kami memutuskan untuk menginap di harris krn disana terdapat kolam renang anak, kids station dan program2 fun untuk anak2. Kamar hotel juga sangat nyaman,bersih&ckp besar.Variasi sarapannya jg sangat byk. Pelayanan staff juga sangat ramah.Lokasi hotel yg berdekatan dgn mall membuat suasana semakin menyenangkan.Bravo Hotel Harris Bekasi&thank u for excellent service!', 'Hotel ini merupakan salah satu hotel yang sangat nyaman dan bersih di bekasi', '+element.review+"] ['Ke 2 kalinya saya bermalam di RED Planet bekasi, dan emang smuanya it dr pelayan dan kamar sudah baik. Intinya pas dan nyaman bwt kt yg jenuh dan mau beristirahat dg tenang di RED Planet aj', 'Ramah bersih tenang itu yg sy rasakan selama nginep di Red Planet sehingga sy dan psngan sy merasa nyaman honeymoon disitu...', 'Clean and nice places wt affordable prices on pegipegi and strategic hotel next to carefour and siloam hospital, many food&#039;s cafe around hotel area, recommended', 'overall memuaska n', 'Akses sangat deket dengan akses tol, sangat mudah di jangkau.', 'Kamar bersih, staff semua ramah. next holiday nginep disini lagi', 'Rekomend deh mantap fasilitas dan pelanannya Top bwt saya pribadi krn saya baru pertama x ke Red Planet bekasi', 'Dengan harga yg cukup murah saya mendapatkan kamar yg bagus dan nyaman.', 'Hotel yg bersih dan sangat nyaman untuk menginap serta di dukung staff hotel yg terbaik,juga dekat deng an lokasi saya berkerja pada saat itu.serta di dukung koneksi wifi yang cepat memudah kan saya untuk berkirim email dan jejaring sosial', 'sangat nyaman menginap di hotel red planet bekasi', "+element.review+"] ['hote l yg cukup nyaman baik kamar maupun kmr mandi nya, lokasi sgt strategis krn bersebelahan dgn mall summarecon jd mudah utk cari variasi kuliner & hiburan..swimming poll jg sgt nyaman utk balita..karyawannya jg ramah..breakfastnya lumayan enak & byk sekali variasi menu tradisionalnya...mantap pokoknya lah...', 'saya dan keluarga puas dg harga, pelayanan dan fasilitas hotel', 'Kamar luas, bersih, dan fasilitas lengkap, lokasi strategis, pelayanan staff ramah cuma sedikit lambat..', 'Pengalaman pertama menginap di hotel harris sumarecon bekasi sangat mengesankan. Kami memutuskan untuk menginap di harris krn disana terdapat kolam renang anak, kids station dan program2 fun untuk anak2. Kamar hotel juga sangat nyaman,bersih&ckp besar.Variasi sarapannya jg sangat byk. Pelayanan staff juga sangat ramah.Lokasi hotel yg berdekatan dgn mall membuat suasana semakin menyenangkan.Bravo Hotel Harris Bekasi&thank u for excellent service!', 'Hotel ini merupakan salah satu hotel yang sangat nyaman dan bersih di bekasi', '+element.review+"] ['Ke 2 kalinya saya bermalam di RED Planet bekasi, dan emang smuanya it dr pelayan dan kamar sudah baik. Intinya pas dan nyaman bwt kt yg jenuh dan mau beristirahat dg tenang di RED Planet aj', 'Ramah bersih tenang itu yg sy rasakan selama nginep di Red Planet sehingga sy dan psngan sy merasa nyaman honeymoon disitu...', 'Clean and nice places wt affordable prices on pegipegi and strategic hotel next to carefour and siloam hospital, many food&#039;s cafe around hotel area, recommended', 'overall memuaska n', 'Akses sangat deket dengan akses tol, sangat mudah di jangkau.', 'Kamar bersih, staff semua ramah. next holiday nginep disini lagi', 'Rekomend deh mantap fasilitas dan pelanannya Top bwt saya pribadi krn saya baru pertama x ke Red Planet bekasi', 'Dengan harga yg cukup murah saya mendapatkan kamar yg bagus dan nyaman.', 'Hotel yg bersih dan sangat nyaman untuk menginap serta di dukung staff hotel yg terbaik,juga dekat deng an lokasi saya berkerja pada saat itu.serta di dukung koneksi wifi yang cepat memudah kan saya untuk berkirim email dan jejaring sosial', 'sangat nyaman menginap di hotel red planet bekasi', "+element.review+"] ['hote l yg cukup nyaman baik kamar maupun kmr mandi nya, lokasi sgt strategis krn bersebelahan dgn mall summarecon jd mudah utk cari variasi kuliner & hiburan..swimming poll jg sgt nyaman utk balita..karyawannya jg ramah..breakfastnya lumayan enak & byk sekali variasi menu tradisionalnya...mantap pokoknya lah...', 'saya dan keluarga puas dg harga, pelayanan dan fasilitas hotel', 'Kamar luas, bersih, dan fasilitas lengkap, lokasi strategis, pelayanan staff ramah cuma sedikit lambat..', 'Pengalaman pertama menginap di hotel harris sumarecon bekasi sangat mengesankan. Kami memutuskan untuk menginap di harris krn disana terdapat kolam renang anak, kids station dan program2 fun untuk anak2. Kamar hotel juga sangat nyaman,bersih&ckp besar.Variasi sarapannya jg sangat byk. Pelayanan staff juga sangat ramah.Lokasi hotel yg berdekatan dgn mall membuat suasana semakin menyenangkan.Bravo Hotel Harris Bekasi&thank u for excellent service!', 'Hotel ini merupakan salah satu hotel yang sangat nyaman dan bersih di bekasi', '+element.review+"] ['Ke 2 kalinya saya bermalam di RED Planet bekasi, dan emang smuanya it dr pelayan dan kamar sudah baik. Intinya pas dan nyaman bwt kt yg jenuh dan mau beristirahat dg tenang di RED Planet aj', 'Ramah bersih tenang itu yg sy rasakan selama nginep di Red Planet sehingga sy dan psngan sy merasa nyaman honeymoon disitu...', 'Clean and nice places wt affordable prices on pegipegi and strategic hotel next to carefour and siloam hospital, many food&#039;s cafe around hotel area, recommended', 'overall memuaska n', 'Akses sangat deket dengan akses tol, sangat mudah di jangkau.', 'Kamar bersih, staff semua ramah. next holiday nginep disini lagi', 'Rekomend deh mantap fasilitas dan pelanannya Top bwt saya pribadi krn saya baru pertama x ke Red Planet bekasi', 'Dengan harga yg cukup murah saya mendapatkan kamar yg bagus dan nyaman.', 'Hotel yg bersih dan sangat nyaman untuk menginap serta di dukung staff hotel yg terbaik,juga dekat deng an lokasi saya berkerja pada saat itu.serta di dukung koneksi wifi yang cepat memudah kan saya untuk berkirim email dan jejaring sosial', 'sangat nyaman menginap di hotel red planet bekasi', "+element.review+"] ['hote l yg cukup nyaman baik kamar maupun kmr mandi nya, lokasi sgt strategis krn bersebelahan dgn mall summarecon jd mudah utk cari variasi kuliner & hiburan..swimming poll jg sgt nyaman utk balita..karyawannya jg ramah..breakfastnya lumayan enak & byk sekali variasi menu tradisionalnya...mantap pokoknya lah...', 'saya dan keluarga puas dg harga, pelayanan dan fasilitas hotel', 'Kamar luas, bersih, dan fasilitas lengkap, lokasi strategis, pelayanan staff ramah cuma sedikit lambat..', 'Pengalaman pertama menginap di hotel harris sumarecon bekasi sangat mengesankan. Kami memutuskan untuk menginap di harris krn disana terdapat kolam renang anak, kids station dan program2 fun untuk anak2. Kamar hotel juga sangat nyaman,bersih&ckp besar.Variasi sarapannya jg sangat byk. Pelayanan staff juga sangat ramah.Lokasi hotel yg berdekatan dgn mall membuat suasana semakin menyenangkan.Bravo Hotel Harris Bekasi&thank u for excellent service!', 'Hotel ini merupakan salah satu hotel yang sangat nyaman dan bersih di bekasi', '+element.review']]
```

Tampilan Data Hasil Scraping



```
File Edit Shell Debug Options Window Help
[Tempat', 'nya', 'oke', 'banget', 'strategis', 'kebersihan', 'ok', 'Over', 'all', 'baik', 'Hotel', 'Horison', 'Ultima', 'bekasi', 'sangat', 'nyaman', 'selain', 'letaknya', 'yang', 'strategis', 'karena', 'letaknya', 'berada', 'di', 'jantung', 'kota', 'bekasi', 'pelayanannya', 'pun', 'sangat', 'baik', 'dan', 'hotel', 'ini', 'sangat', 'recomended', 'agai', 'pelancong', 'pelancong', 'yang', 'ingin', 'mendapatkan', 'kenyamanan', 'yang', 'lebih', 'Sangat', 'baik', 'dari', 'semua', 'segit', 'baik', 'pelayanan', 'kebersihan', 'dll', 'sehingga', 'perjalanan', 'dengan', 'keluarga', ' bisa', 'sangat', 'berkesan', 'Horison', 'Hotel', 'Bekasi', 'memberikan', 'kenyamanan', 'utk', 'berlibur', 'bersama', 'keluarga', 'selain', 'kamarnya', 'yg', 'cukup', 'nyaman', 'dan', 'bersih', 'pelayanannya', 'pun', 'ramah', 'So', 'ti', 'idak', 'akan', 'mengecewakan', 'deh', 'Happy', 'to', 'be', 'back', 'again', 'and', 'again', 'to', 'this', 'place', 'I', '039', 'm', 'stayed', 'near', 'on', 'Bekasi', 'But', 'if', 'it', 'needed', 'to', 'quality', 'time', 'no', 'need', 'to', 'go', 'to', 'far', 'Just', 'stay', 'and', 'try', 'the', 'facilities', 'I', 'will', 'be', 'back', 'as', 'soon', 'Nyaman', 'banget', 'fasilitas', 'mantap', 'makanannya', 'top', 'bnget', 'Intinya', 'betah', 'dah', 'Mantap', 'untuk', 'bisnis', 'dan', 'Cocok', 'Untuk', 'keluar', 'Anda', 'di', 'Weekend', 'atau', 'liburan', 'element', 'review', 'Ke', '12']
```

Tampilan Data Hasil Tokenizing

tempat nya oke banget strategis bersih ok over all baik hotel horison ultima bekas sangat nyaman selain letak yang strategis karena letak ada di jantung kota bekas layan pun sangat baik dan hotel ini sangat recomended bagi lancong lancong yang ingin dapat nyaman yang lebih sangat baik dari semua segi baik layan bersih dll sehi ngga jalan dengan keluarga bisa sangat kesan horison hotel bekas beri nyaman utk libur sama keluarga selain k amar yg cukup nyaman dan bersih pelayanannya pun ramah so tidak akan kecewa deh happy to be back again an d again to this place i 039 m stayed near on bekas but if i needed to quality time no need to go to far away ju st stay and try the facilities i will be back as soon nyaman banget fasilitas mantap makananya top bngt inti betah dah mantap untuk bisnis dan cocok untuk keluar anda di weekend atau libur element review ke 2 kali sa ya malam di red planet bekas dan emang smuanya it dr layan dan kamar sudah baik inti pas dan nyaman bwt kt yg jenuh dan mau istirahat dg tenang di red planet aj ramah bersih tenang itu yg sy rasa lama nginep di red pl anet sehingga sy dan psngan sy rasa nyaman honeymoon situ clean and nice places wt affordable prices on pegi pegi and strategic hotel next to carefour and siloam hospital many food 039 s cafe around hotel area recomm ended overall mudah akses sangat deket dengan akses tol sangat mudah di jangkau kamar bersih staff semua ra mah next holiday nginep sini lagi rekomened deh mantap fasilitas dan pelan top bwt saya pribadi krn saya baru

Tampilan Data Hasil Stemming

```
File Edit Shell Debug Options Window Help ===== RESTART: C:\Users\USER\Documents\PI\PI DDM.py =====
['tempat', 'oke', 'banget', 'strategis', 'bersih', 'oven', 'baik', 'hotel', 'ultima', 'bekas', 'sangat', 'nyaman', 'se lain', 'letak', 'yang', 'strategis', 'karena', 'letak', 'ada', 'di', 'jantung', 'kota', 'bekas', 'layan', 'pun', 'sangat', 'baik', 'dan', 'hotel', 'ini', 'sangat', 'bagi', 'lancong', 'lancong', 'yang', 'ingin', 'dapat', 'nyaman', 'yang', 'lebih', 'sangat', 'baik', 'dari', 'semua', 'segi', 'baik', 'layan', 'bersih', 'sehingga', 'jalan', 'dengan', 'keluarga', 'bisa', 'sangat', 'kesan', 'hotel', 'bekas', 'beri', 'nyaman', 'libur', 'sama', 'keluarga', 'selain', 'kamar', 'cukup', 'nyaman', 'dan', 'bersih', 'pun', 'ramah', 'tidak', 'akan', 'kecewa', 'deh', 'bekas', 'nyaman', 'banget', 'fasilitas', 'mantap', 'top', 'inti', 'betah', 'dah', 'mantap', 'untuk', 'bisnis', 'dan', 'cocok', 'untuk', 'keluar', 'di', 'atau', 'libur', 'ke', 'kali', 'saya', 'malam', 'di', 'planet', 'bekas', 'dan', 'emang', 'layan', 'dan', 'kamar', 'sudah', 'bai k', 'inti', 'pas', 'dan', 'nyaman', 'jenuh', 'dan', 'mau', 'istirahat', 'tenang', 'di', 'planet', 'ramah', 'bersih', 'ten
```

Tampilan Data Hasil Filter(Stopword) dan Topic Model #Stopword

Counter({'hotel': 30, 'idan': 24, 'nyaman': 20, 'sangat': 17, 'di': 17, 'untuk': 16, 'bersih': 15, 'kamar': 14, 'yang': 13, 'baik': 11, 'saya': 10, 'dengan': 10, 'bekas': 9, 'anak': 9, 'kolam': 8, 'inap': 8, 'ada': 8, 'ramah': 8, 'fasilitas': 7, 'strategis': 7, 'renang': 7, 'libur': 7, 'banget': 6, 'layan': 6, 'dapat': 6, 'juga': 6, 'ini': 6, 'lengkap': 5, 'lokasi': 5, 'harga': 5, 'pun': 5, 'tidak': 5, 'luas': 5, 'karena': 5, 'planet': 5, 'variasi': 4, 'cukup': 4, 'keluar': 4, 'gal': 4, 'mandi': 4, 'kami': 4, 'mantap': 4, 'biasa': 4, 'enak': 4, 'bagus': 4, 'besar': 4, 'muas': 4, 'kali': 3, 'putus': 3, 'selain': 3, 'rasa': 3, 'darit': 3, 'harap': 3, 'buat': 3, 'jadi': 3, 'murah': 3, 'paling': 3, 'semua': 3, 'dekat': 3, 'itu': 3, 'sehingga': 3, 'mudah': 3, 'tenang': 3, 'keluar': 3, 'caril': 3, 'letak': 3, 'pokok': 3, 'top': 3, 'senang': 3, 'kota': 3, 'puas': 3, 'indah': 3, 'sertal': 3, 'banyak': 2, 'dukung': 2, 'tempat': 2, 'sana': 2, 'bawa': 2, 'deh': 2, 'lain': 2, 'pakai': 2, 'sini': 2, 'akes': 2, 'lancang': 2, 'milik': 2, 'viat': 2, 'inti': 2, 'masalah': 2, 'makann': 2, 'cocok': 2, 'hibur': 2, 'sarap': 2, 'lebih': 2, 'beri': 2, 'sajat': 2, 'sedikit': 2, 'sama': 2, 'mau': 2, 'atau': 2, 'kesan': 2, 'kei': 2, 'pertama': 2, 'deposit': 2, 'over': 1, 'orang': 1, 'beli': 1, 'segit': 1, 'rupa': 1, 'akhiri': 1, 'menu': 1, 'bisnis': 1, 'perlu': 1, 'airt': 1, 'tapi': 1, 'jenak': 1, 'main': 1, 'kloset': 1, 'pelan': 1, 'relaks': 1, 'ingin': 1, 'dah': 1, 'oke': 1, 'darit': 1, 'cepat': 1, 'sahur': 1, 'baham': 1, 'cumam': 1, 'pada': 1, 'baru': 1, 'jantung': 1, 'terus': 1, 'kirim': 1, 'cumat': 1, 'lah': 1, 'sosial': 1, 'satit': 1, 'sudah': 1, 'guna': 1, 'hingga': 1, 'email': 1, 'sayang': 1, 'usah': 1, 'asal': 1, 'umayam': 1, 'tradisional': 1, 'kurang': 1, 'emang': 1, 'hosipital': 1, 'habis': 1, 'jangkau': 1, 'konkesi': 1, 'makin': 1, 'kecil': 1, 'ragam': 1, 'tinggal': 1, 'areal': 1, 'pribadi': 1, 'seluncut': 1, 'suara': 1, 'kerja': 1, 'lagi': 1, 'itol': 1, 'kalau': 1, 'akan': 1, 'alam': 1, 'khusus': 1, 'semangat': 1, 'ultima': 1, 'tanpa': 1, 'keren': 1, 'malam': 1, 'hanyal': 1, 'satut': 1, 'betah': 1, 'macet': 1, 'jenuh': 1, 'alir': 1, 'turup': 1, 'lambat': 1, 'olangkah': 1, 'utama': 1, 'situ': 1, 'pas': 1, 'banding': 1, 'maupun': 1, 'langsung': 1, 'kaya': 1, 'ya': 1, 'dewasa': 1, 'kedap': 1, 'selalu': 1, 'begit': 1, 'kuliner': 1, 'kan': 1, 'saunal': 1, 'staf': 1, 'lama': 1, 'susah': 1, 'aneka': 1, 'favorit': 1, 'jalan': 1, 'suka': 1, 'listirahat': 1, 'tidur': 1, 'kecewa': 1})

Topic Model

accuracy: 64.46% +/- 16.22% (mikro: 64.79%)			
	true positif	true negatif	class precision
pred. positif	37	15	71.15%
pred. negatif	10	9	47.37%
class recall	78.72%	37.50%	

Tampilan Hasil Total Akurasi yang di Dapat

Row No.	prediction(S...confidence(...confidence(...	Kata	Pos Tag	Freq	Sent		
1	positif	1	0	sangat	adv	17	positif
2	positif	1	0	nyaman	adj	17	positif
3	positif	1	0	bersih	adj	16	positif
4	positif	1.000	0.000	baik	adj	10	positif
5	positif	0.852	0.148	bekas	n	10	negatif
6	positif	1.000	0.000	ada	v	9	positif
7	positif	1.000	0.000	layan	v	9	positif
8	positif	1.000	0.000	fasilitas	n	9	positif
9	positif	1.000	0.000	ramah	adj	9	positif
10	positif	0.999	0.001	lengkap	adj	7	positif
11	positif	0.999	0.001	strategis	adj	7	positif
12	positif	0.999	0.001	bagus	adj	7	positif
13	positif	0.990	0.010	dapat	adv	6	positif
14	positif	0.996	0.004	dekat	adj	6	positif
15	positif	0.990	0.010	juga	adv	6	positif
16	positif	0.980	0.020	banget	adv	5	positif
17	negatif	0.010	0.990	tidak	adv	5	negatif
18	positif	0.992	0.008	luas	adj	5	positif
19	positif	0.992	0.008	senang	adj	5	positif
20	negatif	0.000	1.000	selain	i	4	negatif
21	positif	0.987	0.013	muas	adj	4	positif
22	positif	0.987	0.013	puas	adj	4	positif
23	negatif	0.005	0.995	putus	v	3	negatif

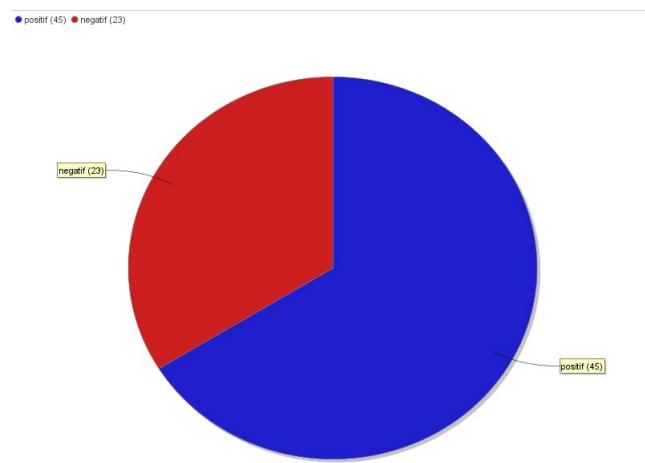
Tampilan Statistik Penjabaran Atribut

PerformanceVector

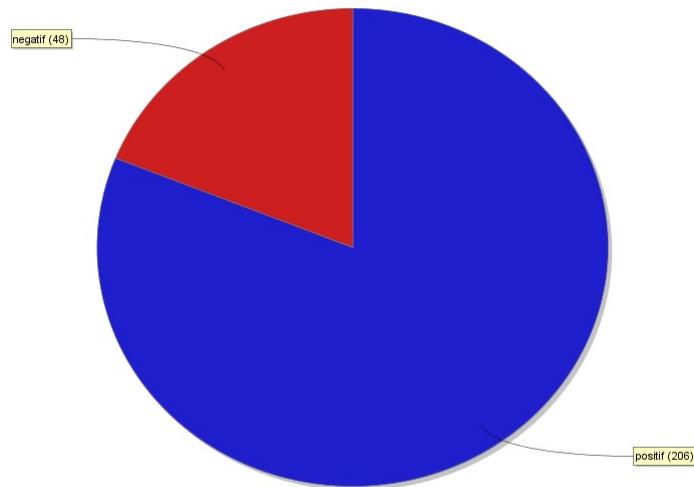
```

PerformanceVector:
accuracy: 64.46% +/- 16.22% (mikro: 64.79%)
ConfusionMatrix:
True: positif negatif
positif:    37    15
negatif:    10     9
  
```

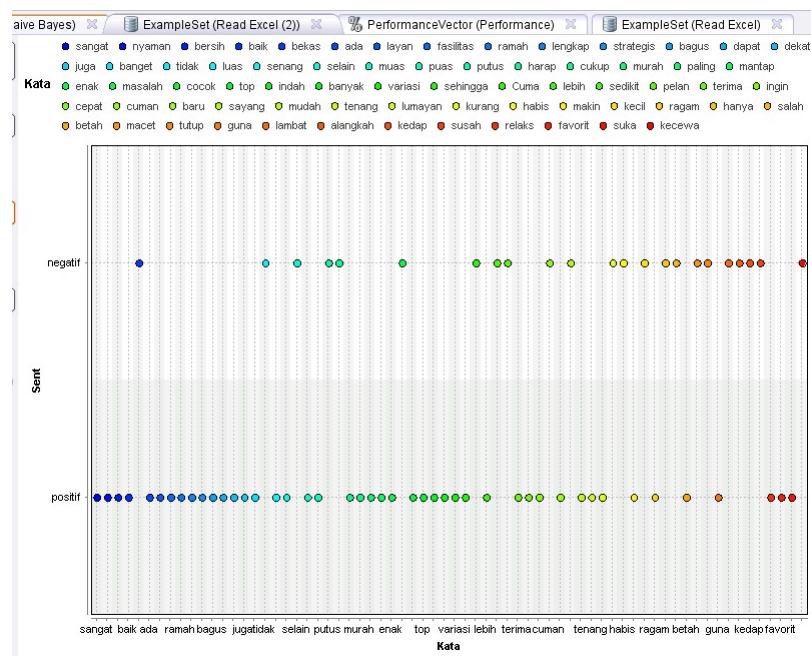
Tampilan Hasil Performance Vector (Confusion Matrix)



Tampilan Diagram Pie Sebelum Dikalikan dengan Frekuensi



Tampilan Diagram Pie Sebelum Dikalikan dengan Frekuensi



Tampilan Diagram Scatter analisis sentimen



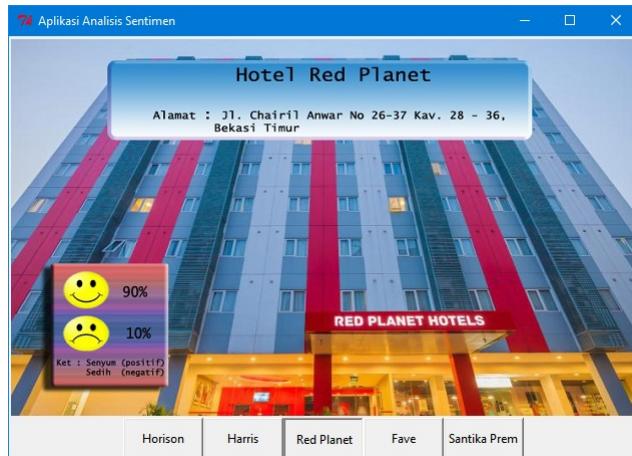
Tampilan Awal Aplikasi Sederhana



Tampilan Output Aplikasi Hotel Horison



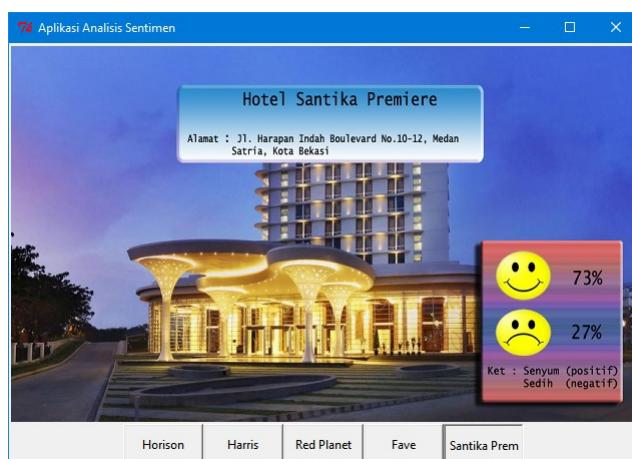
Tampilan Output Aplikasi Hotel Harris



Tampilan Output Aplikasi Hotel Red Planet



Tampilan Output Aplikasi Hotel Fave



Tampilan Output Aplikasi Hotel Santika Premiere