

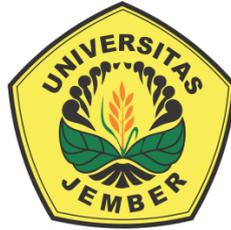


**ANALISIS SENTIMEN TERHADAP PRODUK TANAMAN
TRANSGENIK MENGGUNAKAN *LEXICON-BASED* DAN *NAIVE BAYES***

SKRIPSI

Oleh
Mahrita Arifah
NIM 151810101045

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2019**



**ANALISIS SENTIMEN TERHADAP PRODUK TANAMAN
TRANSGENIK MENGGUNAKAN *LEXICON BASED* DAN *NAIVE BAYES***

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat
untuk menyelesaikan studi pada Program Studi Matematika (S-1)
dan mencapai gelar Sarjana Sains

Oleh
Mahrifa Arifah
NIM 151810101045

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2019**

PERSEMBAHAN

Dengan nama Allah SWT Yang Maha Pengasih dan Maha Penyayang, skripsi ini penulis persembahkan kepada:

1. Ayahanda M. Khoirul Anam dan Ibunda Basiroh yang selalu mendoakan serta memberikan dukungan, semangat, dan motivasi yang tidak pernah terputus.
2. Kakak – kakak tersayang M. Miftachul Munir dan Arina Hidayati yang selalu mendengarkan keluh kesah penulis
3. Keluarga Besar di Jember yang selalu mendukung dan menguatkan melalui doa maupun nasihat.
4. Dr. Alfian Futuhul Hadi,S.Si., M.Si. selaku Dosen Pembimbing Utama dan Dian Anggrani,S.Si., M.Si selaku Dosen Pembimbing Anggota yang telah membimbing secara intensif dalam menyempurnakan tugas akhir ini.
5. Seluruh jajaran guru TK. Kemala Bhayangkari 99, SDN Kedungturi I Taman, SMPN 2 Taman, MBI Amanatul Ummah Pacet-Mojokerto yang telah memberikan banyak ilmu kepada penulis.
6. Sahabat “Songong” Bayu, Dwi, Ega, Hadi, Nazar, Rency, Salsa, dan Yanti yang selalu menemani dalam proses pengerjaan.
7. Teman-teman seperjuangan SIGMA’15 yang selalu memberikan semangat selama perkuliahan.
8. Semua pihak yang membantu penulis dalam menyelesaikan tugas akhir.

MOTTO

“Jangan anggap keberhasilan yang kita raih sebagai usaha kita sendiri, anggaplah semua itu sebagai anugerah dari Allah semata agar timbul rasa bersyukur dalam hati”

(KH Moh Zuhri Zaini)



PERNYATAAN

Saya yang bertanda tangan di bawah ini

Nama : Mahrita Arifah

NIM : 151810101045

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “ Analisis Sentimen Terhadap Produk Tanaman Transgenik Menggunakan *Lexicon Based* dan *Naïve Bayes* ” adalah benar-benar hasil karya ilmiah sendiri, kecuali jika dalam pengutipan substansi disebutkan sumbernya dan belum pernah diajukan pada institusi manapun, serta bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, Juli 2019

Yang menyatakan,

Mahrita Arifah

NIM 151810101045

SKRIPSI

**ANALISIS SENTIMEN TERHADAP PRODUK TANAMAN
TRANSGENIK MENGGUNAKAN *LEXICON BASED* DAN *NAIVE BAYES***

Oleh

**Mahrifa Arifah
NIM 151810101045**

Pembimbing

Dosen Pembimbing Utama : Dr. Alfian Futuhul Hadi, S.Si., M.Si.

Dosen Pembimbing Anggota : Dian Anggraeni, S.Si.,M.Si.

PENGESAHAN

Skripsi berjudul “Analisis Sentimen Terhadap Produk Tanaman Transgenik Menggunakan *Lexicon Based* dan *Naïve Bayes*”, telah diuji dan disahkan pada:
hari, tanggal :

tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas
Jember.

Tim Penguji:

Ketua,

Anggota I,

Dr. Alfian Futuhul Hadi, S.Si., M.Si
NIP. 197407192000121001

Dian Anggraeni, S.Si., M.Si.
NIP. 198202162006042002

Anggota II,

Anggota III,

Prof. Drs. I Made Tirta, M.Sc., Ph.D.
NIP. 195912201985031002

Bagus Juliyanto, S.Si., M.Si
NIP.198007022003121001

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Jember

Drs. Sujito, Ph.D.
NIP. 196102041987111001

RINGKASAN

Analisis Sentimen Terhadap Produk Tanaman Transgenik Menggunakan *lexicon Based* dan *Naïve Bayes*; Mahrita Arifah, 151810101045; 2019: 29 halaman; Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jember.

Tanaman transgenik merupakan hal baru sebagai salah satu solusi permasalahan pada pangan. Disamping keuntungannya, transgenik juga membawa dampak buruk bagi kesehatan dan lingkungan. Sehingga, penggunaan tanaman transgenik masih menuai pro dan kontra dalam masyarakat. Oleh sebab itu, diperlukan opini untuk pengembangan produk pangan tanaman transgenik ini. Untuk mengidentifikasi dan mengekstraksi informasi dari sebuah opini, dibutuhkan suatu metode analisis sentimen. Pada penelitian ini, metode analisis sentimen yang digunakan adalah metode *lexicon based* dan *naïve bayes*, serta dilakukan pengelompokan data dengan *topic modelling*.

Data yang digunakan penelitian ini berupa data teks yang diperoleh dari opini mahasiswa Universitas Jember sebanyak 268 opini. Analisis diawali dengan *text preprocessing*. Tahap ini bertujuan untuk membersihkan dan mengurangi data-data yang tidak diperlukan seperti simbol, tanda baca, dan kata-kata yang dianggap tidak memiliki makna (*stopwords*). Selanjutnya data akan diolah untuk memahami sentimen sosial dari produk tanaman transgenik. Penentuan sentimen pada metode *Lexicon Based* dilakukan dengan pencocokkan data pada kamus. Apabila dalam satu opini memiliki jumlah kata positif lebih banyak, maka opini tersebut bersentimen positif. Begitu pula sebaliknya apabila dalam satu opini memiliki jumlah kata negatif, maka opini tersebut bersentimen negatif. Dan apabila jumlah kata positif sama dengan kata negatif, maka opini tersebut bersentimen netral. Sedangkan penentuan sentimen pada metode klasifikasi *Naïve Bayes* dilakukan secara manual oleh penulis. Kemudian dari data yang telah terlabeli, akan dilakukan pelatihan untuk membentuk sebuah model *classifier*. Model tersebut digunakan untuk melakukan prediksi terhadap data baru. Kedua metode tersebut akan diuji

kinerjanya dalam melakukan prediksi dengan tepat. Model yang paling baik akan digunakan untuk memprediksi data set. Hasil sentimen diklasifikasikan menjadi 3 bagian, yaitu sentimen positif, negatif, dan netral. Sentimen positif menunjukkan, mahasiswa beropini bahwa tanaman transgenik membawa keuntungan sehingga setuju dengan adanya penggunaan tanaman transgenik sebagai bahan pangan dan mendukung perkembangannya. Begitu pula sebaliknya, sentimen negatif menunjukkan, mahasiswa beropini bahwa tanaman transgenik membawa dampak buruk sehingga tidak setuju dengan penggunaan tanaman transgenik sebagai bahan pangan. Sedangkan pada sentimen netral terjadi dua kemungkinan. Pertama, mahasiswa menganggap bahwa tanaman transgenik membawa dampak buruk tetapi tetap setuju dengan penggunaan tanaman transgenik sebagai bahan pangan dan mendukung pengembangannya. Dan yang kedua, mahasiswa menganggap bahwa tanaman transgenik membawa keuntungan, tetapi tidak setuju dengan penggunaan tanaman transgenik sebagai bahan pangan. Sedangkan pada pengelompokkan, metode *topic Modelling* dilakukan untuk mendapatkan topik yang dominan dibahas oleh mahasiswa tentang tanaman transgenik.

Metode yang digunakan untuk memprediksi data set adalah klasifikasi *naïve bayes* dikarenakan nilai akurasinya lebih tinggi dari *lexicon based* yaitu 70%. Analisis sentimen yang dilakukan menghasilkan 175 sentimen positif, 29 sentimen negatif, dan 64 sentimen netral. Sedangkan hasil dari proses pengelompokkan dengan *Topic Modelling* adalah menghasilkan 10 topik. Berdasarkan hasil tersebut dapat disimpulkan bahwa mayoritas dari mahasiswa beropini bahwa tanaman transgenik bermanfaat dan membantu dalam hal pangan. Selain itu tanaman transgenik juga menguntungkan kepada pihak petani dan masyarakat karena dapat meningkatkan hasil pangan. Tetapi adanya bahaya yang ditimbulkan, mtanaman transgenik masih perlu dilakukan pengembangan agar lebih bermutu dan aman untuk di konsumsi.

PRAKATA

Puji Syukur kehadiran Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi yang berjudul “Analisis Sentimen Terhadap Produk Tanaman Transgenik Menggunakan *Lexicon Based* dan *Naïve Bayes*”. Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan strata satu (S1) pada Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak, oleh karena itu penulis ingin menyampaikan ucapan terima kasih kepada :

1. Dr. Alfian Futuhul Hadi, S.Si., M.Si. selaku Dosen Pembimbing Utama dan Dian Anggraeni, S.Si., M.Si. selaku Dosen Pembimbing Anggota yang telah meluangkan waktu, pikiran, serta perhatian dalam penulisan skripsi ini;
2. Prof. Drs. I MadeTirta, M.Sc., Ph.D. dan Bagus Julianto, S.Si., M.Si. selaku dosen penguji yang telah memberikan kritik dan saran demi perbaikan tugas akhir ini;
3. Seluruh dosen dan karyawan Fakultas MIPA Universitas Jember yang telah memberikan ilmu serta bimbingannya sehingga penulis dapat menyelesaikan tugas akhir ini;
4. Seluruh teman seperjuangan SIGMA'15 Jurusan Matematika Universitas Jember;
5. Semua pihak yang telah memberikan sumbangan tenaga, semangat, dan pikiran yang tidak dapat disebutkan satu persatu oleh penulis.

Penulis juga menerima kritik dan saran dari semua pihak demi kesempurnaan skripsi ini. Akhirnya penulis berharap, semoga skripsi ini dapat bermanfaat.

Jember, Juli 2019

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PERSEMBAHAN	ii
HALAMAN MOTTO	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN PEMBIMBING	v
HALAMAN PENGESAHAN.....	vi
RINGKASAN	vii
PRAKATA	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.4 Tujuan.....	3
1.5 Manfaat.....	3
BAB 2. TINJAUAN PUSTAKA.....	4
2.1 Tanaman Transgenik	4
2.2 <i>Text Mining</i>	4
2.3 <i>Lexicon Based</i>	6
2.4 Klasifikasi <i>Naive Bayes</i>	6
2.5 Evaluasi Model.....	7
2.6 <i>Topic Modeling</i>	8
2.7 Visualisasi Data	9

BAB 3. METODE PENELITIAN	10
3.1 Data Penelitian	10
3.2 Langkah Penelitian	10
BAB 4. HASIL DAN PEMBAHASAN	15
4.1 Preprocessing	15
4.2 Analisis Sentimen dengan pendekatan <i>Lexicon Based</i>	19
4.3 Klasifikasi <i>Naïve Bayes</i>	21
4.4 Perbandingan Model	23
4.5 Visualisasi dan Ulasan sentimen	24
4.4 <i>Topic Modelling</i>	26
BAB 5. PENUTUP	29
5.1 Kesimpulan	27
5.2 Saran	27
DAFTAR PUSTAKA	29
LAMPIRAN	30

DAFTAR TABEL

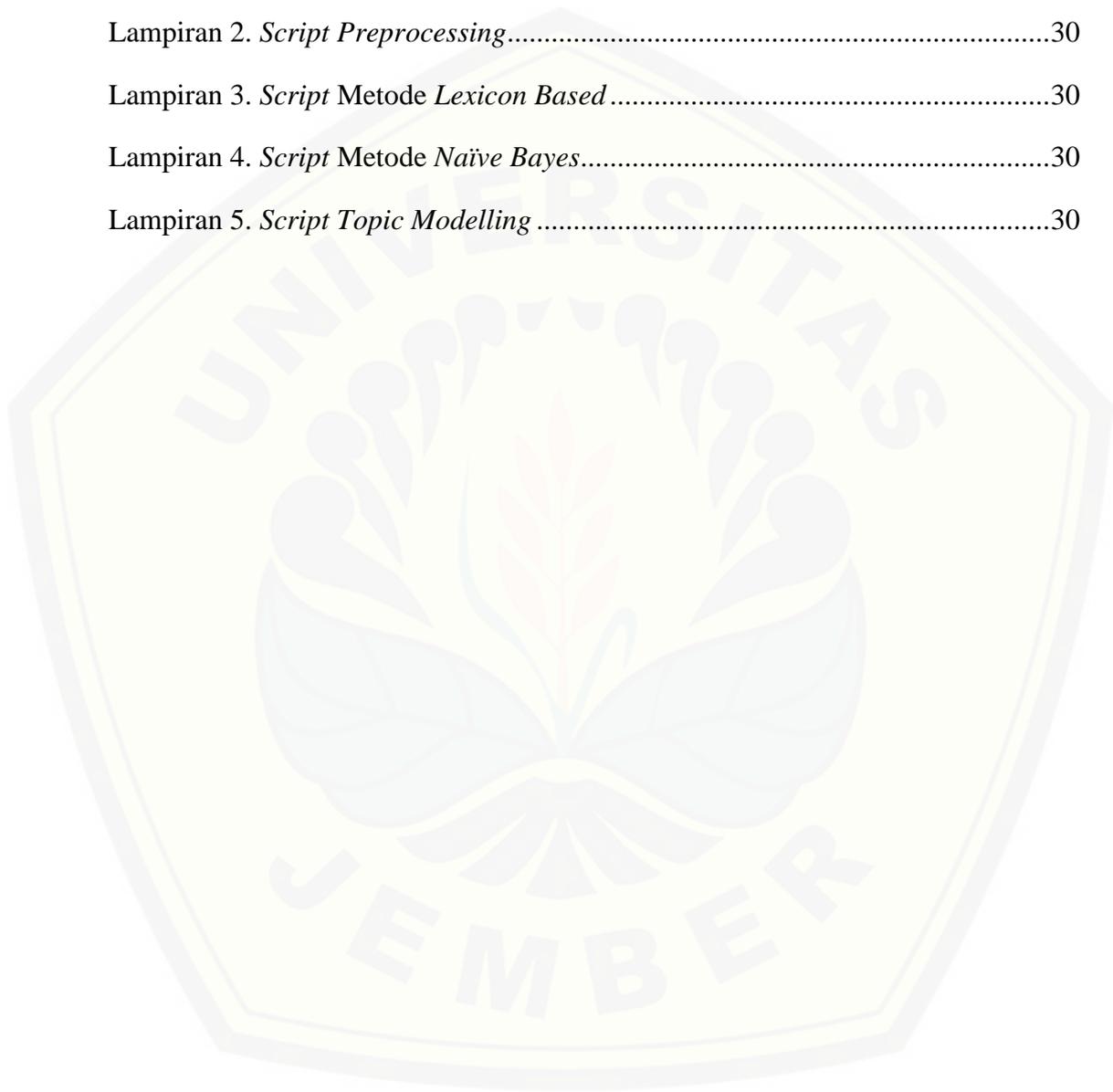
	Halaman
Tabel 2.1 <i>Confusion Matrix</i>	8
Tabel 3.1 Contoh Data	10
Tabel 4.1 Hasil proses <i>case folding</i>	15
Tabel 4.2 Hasil proses normalisasi.....	16
Tabel 4.3 Hasil proses <i>stopword removal</i>	17
Tabel 4.4 Hasil proses <i>stemming</i>	18
Tabel 4.5 Hasil proses <i>Tokenizing</i>	18
Tabel 4.6 Contoh hasil prediksi sistem	19
Tabel 4.7 Contoh Hasil Pengujian sistem <i>lexicon</i>	20
Tabel 4.8 Hasil <i>confusion matrix</i> metode <i>lexicon</i>	20
Tabel 4.9 Hasil performa <i>lexicon based</i>	21
Tabel 4.10 Hasil pengujian model klasifikasi <i>naïve bayes</i>	22
Tabel 4.11 Hasil <i>confusion matrix</i> metode <i>naïve bayes</i>	22
Tabel 4.12 Hasil Performa Klasifikasi <i>Naïve Bayes</i>	23
Tabel 4.13 Hasil rata-rata performa <i>Naïve Bayes</i>	23
Tabel 4.14 Hasil akurasi model.....	23
Tabel 4.15 Pemodelan topik dengan teknik LDA	28
Tabel 4.16 Pemodelan topik dengan teknik LDA	28

DAFTAR GAMBAR

	Halaman
Gambar 2.1 <i>Word Cloud</i>	10
Gambar 3.1 Diagram alir langkah-langkah penelitian	13
Gambar 3.2 Diagram alir proses klasifikasi <i>naïve bayes</i>	14
Gambar 4.1 Grafik nilai akurasi.....	23
Gambar 4.2 Jumlah hasil prediksi tiap kelas	24
Gambar 4.3 Frekuensi kata	25
Gambar 4.2 Visualisasi data dengan worldcloud.....	26
Gambar 4.2 Nilai <i>perplexity</i> kandidat topik	27

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. <i>Script</i> Pengambilan Data	30
Lampiran 2. <i>Script Preprocessing</i>	30
Lampiran 3. <i>Script Metode Lexicon Based</i>	30
Lampiran 4. <i>Script Metode Naïve Bayes</i>	30
Lampiran 5. <i>Script Topic Modelling</i>	30



BAB 1. PENDAHULUAN

1.1 Latar Belakang

Kebutuhan masyarakat akan pangan semakin meningkat, mengingat pangan menjadi salah satu kebutuhan pokok yang harus terpenuhi setiap hari. Seiring perkembangan jaman, banyak faktor yang mempengaruhi ketidakstabilan hasil pangan, seperti perubahan iklim hingga semakin minimnya lahan pertanian. Kehadiran tanaman transgenik menjadi salah satu alternatif untuk permasalahan pangan tersebut. Tanaman transgenik dihasilkan dengan cara mengintroduksi gen tertentu ke dalam tubuh tanaman, sehingga diperoleh sifat yang diinginkan. Sifat tersebut diantaranya tahan hama, toleran herbisida, tahan antibiotik, memiliki kualitas nutrisi lebih baik, serta produktivitas yang lebih tinggi (Karmana, 2009). Hal tersebut membuat produk tanaman transgenik dianggap mampu meningkatkan produksi pangan yang lebih tinggi dan efisien.

Universitas Jember menjadi salah satu perguruan tinggi yang ikut andil terhadap perkembangan tanaman transgenik. Penelitian yang telah dilakukan oleh Prof Bambang Sugiharto selama tahun 1994 hingga tahun 2003 menghasilkan produk tebu transgenik pertama di dunia yang mampu menghasilkan produksi mencapai 20-30 % lebih tinggi dari tebu varietas lain pada saat ditanam di lahan kering. Hingga saat ini para peneliti Universitas Jember terus mengembangkan produk pangan transgenik tersebut. Pemanfaatan tanaman transgenik sebagai pangan terlihat menguntungkan, namun di sisi lain masyarakat masih khawatir terhadap penggunaannya, terutama mengenai dampak pada lingkungan dan kesehatan jika dikonsumsi dalam jangka panjang. Sehingga penggunaan tanaman transgenik masih menuai pro dan kontra di kalangan masyarakat. Oleh sebab itu, untuk pengembangan produk pangan tanaman transgenik ini diperlukan sebuah opini mengenai penggunaan tanaman transgenik sebagai bahan pangan.

Dalam mengidentifikasi dan mengekstraksi informasi dari sebuah opini, dibutuhkan suatu metode analisis sentimen. Analisis sentimen merupakan suatu metode yang digunakan untuk mengekstrak dan mengolah data tekstual untuk mendapatkan informasi sentimen yang terkandung dalam data teks tersebut. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah. Terdapat beberapa metode untuk analisis sentimen, yaitu Metode *Lexicon Based*, *text mining*, dan *hybrid approach*. Metode *lexicon based* merupakan metode yang berdasarkan pendekatan kamus yang berisi kata-kata opini atau biasa disebut kamus opini (*opinion lexicon*). Metode ini bergantung pada kata penyusunnya, yang mana kata tersebut mengekspresikan sentimen positif atau negatif. (Khusboo et al., 2012). Penelitian yang terkait dengan metode *lexicon based* yaitu (Nurfalah, 2017). Hasil yang diperoleh yaitu metode *lexicon based* cukup baik dalam menganalisis sentimen dengan tingkat akurasi 66%. Metode *text mining* adalah sebuah alat untuk menganalisis suatu kumpulan dokumen. *Text mining* merupakan bagian dari keilmuan *data mining* yang berguna untuk mengkstraksi informasi penting yang tersembunyi dari data. Selain itu *text mining* juga mengidentifikasi dan mengeksplor pola-pola yang menarik pada data (Feldman et al., 2007). Terdapat beberapa metode *text mining* yaitu *Naïve Bayes*, *Neural Networks*, *K-means*, *Topic Modelling* dan lain sebagainya. Beberapa peniltian yang terkait dengan metode *text mining* yaitu dilakukan (Rasyadi, 2018) ,metode yang digunakan adalah *naïve bayes classifier*. Hasil penelitian tersebut menyatakan bahwa *naïve bayes* mampu mengklasifikasikan sentimen dengan baik.. Berdasarkan hal tersebut, maka penulis tertarik untuk melakukan penelitian terkait analisis sentimen terhadap tanaman transgenik dengan menggunakan metode *lexicon based* dan *naïve bayes*.

1.2 Rumusan Masalah

Adapun rumusan masalah yang akan dibahas pada penelitian kali ini adalah bagaimana analisis sentimen dari opini atau respon mahasiswa terhadap produk tanaman transgenik menggunakan pendekatan *Lexicon-Based* dan *machine learning*

1.3 Tujuan Penelitian

Tujuan yang ingin didapatkan dari penelitian ini adalah mengetahui analisis sentimen dari opini atau respon mahasiswa terhadap produk pangan transgenik.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini diharapkan memberikan pengetahuan kepada penulis dan pembaca, serta memberikan informasi mengenai opini mahasiswa terhadap produk tanaman transgenik, sehingga dapat dijadikan sebagai referensi untuk pengembangan dan peningkatannya.



BAB 2. TINJAUAN PUSTAKA

2.1 Tanaman Transgenik

Tanaman transgenik merupakan organisme yang gennya diubah dengan menggunakan teknik rekayasa genetika. Secara sederhana tanaman transgenik dibuat dengan cara mengambil gen-gen tertentu yang baik pada makhluk hidup lain untuk disisipkan melalui suatu perantara untuk menghasilkan tanaman yang dikehendaki. Tanaman transgenik memegang peran penting dalam ketahanan pangan nasional. Hal ini disebabkan adanya pertambahan jumlah penduduk dunia yang tidak berbanding lurus dengan ketersediaan lahan pertanian (Kementrian Pertahanan Republik Indonesia, 2015). Kehadiran tanaman transgenik membawa beberapa keuntungan, diantaranya yaitu mampu meningkatkan hasil yang tinggi dan efisien tanpa membutuhkan banyak lahan. Tanaman ini juga memiliki sifat yang unggul, seperti tahan terhadap herbisida, hama, serangga, dan virus, serta dapat menghambat pelunakan buah (pada tomat), meningkatkan nilai gizi tanaman, dan meningkatkan kemampuan tanaman untuk hidup pada lahan yang ekstrim seperti lahan kering, lahan dengan keasaman dan kadar garam yang tinggi (Karmana, 2018)

Terlepas dari kelebihanannya, penggunaan tanaman transgenik untuk bahan pangan juga masih dikhawatirkan, karena dapat menimbulkan alergi, karsinogenik, resistensi antibiotik, dan perpindahan gen ke lingkungan. Transfer gen resisten hama dapat mengganggu keseimbangan ekosistem. Kasus yang pernah terjadi yaitu pada jagung transgenik yang mengandung gen Bt menyebabkan kematian larva monarch dan efek buruk terhadap perkembangan, berat badan, serta perilaku kupu-kupu (Losey *et al.*, 1999).

2.2 Text Mining

Text mining merupakan suatu metode analisis yang digunakan untuk mencari dan mengekstraksi informasi yang berguna pada suatu kumpulan data. *Text mining* merupakan bagian dari keilmuan *data mining* yang berfokus pada data teks. Sehingga

text mining juga mampu menginterpretasikan pola-pola menarik, seperti keterkaitan antar kata, dan lain sebagainya. Secara umum *text mining* digunakan untuk menganalisa kuantitas dengan jumlah besar. Hal tersebut dikarenakan penggunaan teknik *machine learning* pada *text mining*, sehingga hanya perlu melakukan pembelajaran data pada sistem yang selanjutnya akan dilakukan pengolahan data oleh sistem itu sendiri. Disamping menggunakan teknik *data mining* dan *machine learning*, *text mining* juga mengimplementasikan teknik *natural language processing* (NLP), *information retrieval* (IR), dan *knowledge management* (Feldman et al, 2007). *Text mining* sering digunakan untuk melakukan pengklasifikasian dan pengelompokkan data teks, ekstraksi pola atau keterkaitan, peringkasan dokumen hingga analisis sentimen. Terdapat beberapa tahap dalam melakukan analisis *text mining*, yaitu (Jo,2019):

- a. Data teks, adalah kumpulan data dokumen yang menjadi objek penelitian
- b. Teks *Preprocessing*, merupakan langkah awal agar data bisa digunakan untuk tahap selanjutnya. Proses ini dibutuhkan karena data teks merupakan data yang tidak terstruktur, sehingga dibutuhkan penyeragaman pada data. Selain itu, tahap ini berguna untuk menghapus data yang tidak diperlukan sehingga proses analisis lebih efektif dan efisien. Terdapat 3 tahap dasar pada tahapan ini yaitu:
 - 1) *Tokenization*, tahap ini melakukan pemotongan data kedalam bentuk per kata yang menyusunnya, serta menghilangkan tanda baca pada data
 - 2) *Stemming*, tahap ini mengubah setiap kata ke bentuk kata dasarnya.
 - 3) *Stopword removal*, tahap ini melakukan penghapusan pada kata-kata yang tidak dibutuhkan, seperti kata hubung, keterangan, dan lain sebagainya.
- c. Transformasi teks, langkah ini mentransformasi teks kedalam bentuk dokumen ruang vektor notasi model atau disebut dengan istilah *bag of word* yang berguna untuk memudahkan analisis selanjutnya.
- d. Teknik *Text Mining*
- e. Evaluasi

Namun dalam beberapa literasi, menurut Mathiak dan Eckstein pada tahun 2004 serta Gaikwad tahun 2014 dalam (Dimas,2017) beberapa peneliti menambahkan beberapa tahapan-tahapan pada *text preprocessing* seperti *case folding*, *cleansing*, dan *normalization*.

2.3 *Lexicon Based*

Metode berbasis *lexicon* yaitu suatu metode analisis sentimen yang menggunakan pendekatan kamus opini (*opinion lexicon*). Pada metode ini masing-masing kata akan di periksa oleh sistem. Apabila kata tersebut masuk kedalam kamus positif maka kata tersebut adalah positif. Begitu juga apabila kata tersebut masuk kedalam kamus negatif maka kata tersebut adalah kata negatif. Menurut Khusboo et al., (2012) pelabelan data dapat dilakukan dengan cara sebagai berikut:

1. Menentukan label pada masing-masing kata, kemudian hitung jumlah kata positif dan negatifnya
2. Jika jumlah kata positif $>$ jumlah kata negatif, maka label sentimennya adalah positif (skor 1)
3. Jika jumlah kata positif $<$ jumlah kata negatif, maka label sentimennya adalah negatif (skor -1)
4. Jika jumlah kata positif $=$ jumlah kata negatif, maka label sentimennya adalah netral (skor 0)

2.4 *Klasifikasi Naive Bayes*

Klasifikasi *naive bayes* merupakan implementasi teknik *machine learning* yaitu *supervised learning*. Klasifikasi ini bekerja dengan cara mempelajari probabilitas dari setiap anggota kelasnya pada proses pelatihan data. Setelah tahap pelatihan, model klasifikasi yang telah dibuat, diaplikasikan pada data baru. Setiap anggota kelas pada data baru akan diklasifikasikan pada kategori yang dianggap paling memungkinkan berdasarkan probabilitas gabungan yang ditentukan oleh kombinasi prediktor-

prediktor (Han et al, 2012). Aturan Bayesien yang akan digunakan adalah sebagai berikut :

$$p(C_i|X) = \frac{P(X|C_i)*P(C_i)}{P(X)} \quad (2.1)$$

Kemudian diasumsikan masing-masing petunjuk saling bebas satu sama lain sehingga berlaku

$$p(C_i|X_1, \dots, X_n) = P(C_i) \prod_{i=1}^n P(X|C_i) \quad (2.2)$$

Sehingga persamaan asumsi diatas menghasilkan

$$p(C_i|X) = P(x_1|C_i)P(x_2|C_i) \dots P(x_n|C_i)P(C_i) \quad (2.3)$$

2.5 Evaluasi Model

Model klasifikasi yang telah dibuat, selanjutnya akan digunakan untuk memprediksi data baru. Oleh karena itu, diperlukan sebuah pengujian untuk mengetahui seberapa akurat klasifikasi tersebut untuk memprediksi data. Salah satu metode untuk menguji kinerja model yaitu dengan menggunakan *confusion matrix* (Han et al.2010). Bagian baris pada matriks merupakan kelas aktual sedangkan untuk bagian kolom merupakan kelas hasil prediksi.

Tabel 2.1 *Confusion Matrix*

<i>Actual class</i>	<i>Predicted class</i>		<i>Total</i>
	<i>Yes</i>	<i>No</i>	
<i>Yes</i>	<i>TP</i>	FN	<i>P</i>
<i>No</i>	FP	TN	N
<i>Total</i>	P'	N'	P+N

TP (*True Positive*) adalah kelas positif yang di beri prediksi label dengan benar oleh model klasifikasi. TN (*True Negatif*), kelas negatif yang di beri prediksi label dengan benar oleh model klasifikasi. FP (*False Positive*), kelas positif yang di prediksi salah oleh mesin (diprediksi negatif). FN (*False Negatif*), kelas negatif yang di

prediksi salah oleh mesin (diprediksi positif). P' merupakan jumlah kelas yang di prediksi positif ($TP+FP$) dan N' adalah jumlah kelas yang diprediksi negatif ($TN+FN$). Sehingga jumlah kelas adalah $TP+TN+FP+FN$, atau $P+N$, atau $P'+N'$. Nilai akurasi dari klasifikasi diperoleh dengan menghitung kelas yang diprediksi oleh sistem dengan benar, sehingga rumus akurasi adalah sebagai berikut:

$$Akurasi = \frac{TP+TN}{TP+FN+FP+TN} \quad (2.4)$$

Kinerja model klasifikasi juga dapat dihitung dengan nilai *recall* dan *precision*. *Recall* merupakan nilai persentase dari model untuk mengingat pengelompokan teks, rumus *recall* sebagai berikut :

$$Recall = \frac{TP}{TP+FN} \quad (2.5)$$

Precision merupakan nilai persentase dari model untuk melakukan prediksi dengan tepat. Rumus *precision* sebagai berikut :

$$Precision = \frac{TP}{TP+FP} \quad (2.6)$$

2.5 Topic Modeling

Topic modeling merupakan salah satu metode *clustering* yang termasuk dalam bidang *unsupervised learning*. Terdapat 3 metode *clustering* yaitu *hard*, *hierarchical*, dan *soft/fuzzy clustering*. *Topic modelling* termasuk dalam *soft/fuzzy clustering* dimana setiap objek dapat dimiliki lebih dari satu *cluster* dengan tingkat tertentu. Pemodelan topik digunakan untuk menggali tema atau topic yang menjadi dasar dalam kumpulan dokumen yang tidak terstruktur (Kumar et al, 2016).

Latent Dirichlet Allocation merupakan salah satu metode *topic modelling*. Kumar (2016:124) mengatakan bahwa LDA mampu mengelompokkan dokumen dengan topik dan tema mendasar secara efektif. Disamping itu LDA juga mampu memecahkan masalah dalam mencari, mengatur, dan maerangkum data besar yang tidak terstruktur dengan mudah. Pemodelan topik mengadopsi strategi tiga cabang untuk mengekstraksi tema dari kumpulan dokumen , yaitu sebagai berikut:

1. Setiap kata dalam dokumen diberi topik

2. Proporsi setiap kata diperkirakan untuk setiap dokumen
3. Untuk setiap korpus, topik distribusi dieksplorasi

2.6 Visualisasi Data

Visualisasi merupakan rekayasa dalam pembuatan gambar, diagram, atau animasi untuk penampilan suatu informasi. Visualisasi mempermudah dalam menyimpulkan hingga menggambarkan karakteristik hingga korelasi data. Selain menggunakan line dan bar plot, visualisasi dapat dilakukan dengan menggunakan *word cloud*. *Word cloud* adalah salah satu metode visualisasi yang digunakan untuk mengidentifikasi fokus utama dari materi penulisan. *Word cloud* bekerja dengan memunculkan kata-kata berdasarkan jumlah kata. Semakin sering kata muncul, maka akan muncul dengan ukuran *font* yang lebih besar. Berikut ini adalah salah satu contoh dari *word cloud*.



Gambar 2.1 *Word Cloud*

BAB 3. METODOLOGI PENELITIAN

3.1 Data Penelitian

Data set yang digunakan pada penelitian ini diperoleh dari opini mahasiswa angkatan 2018 Universitas Jember. Pemilihan mahasiswa angkatan 2018, dikarenakan pada saat Pengenalan Kehidupan Kampus (PK2) mahasiswa diberi pembekalan informasi tentang tanaman transgenik. Sehingga dengan pembekalan yang didapat, diharapkan mahasiswa mampu memberikan opini mengenai penggunaan produk tanaman transgenik sebagai bahan pangan. Data yang terkumpul merupakan data mentah yang berjumlah 268 data yang diambil dari laman web Jurusan Matematika Fakultas MIPA Universitas Jember. Selain data set, pada penelitian ini juga menggunakan data kamus kata negatif dan positif untuk proses analisis sentimen dengan metode *lexicon-based*. Contoh data yang diperoleh dari web ditunjukkan oleh Tabel 3.1.

Tabel 3.1 Contoh Data

Dok	Contoh Data
1	Sangat membantu bagi masyarakat untuk memenuhi kebutuhan pangan. Apalagi jika ekosistem di Indonesia mengalami keadaan iklim yang tidak menentu.
2	Tanaman transgenik dapat mengatasi krisis pangan, akan tetapi dari artikel di atas ada efek samping yang ditimbulkan. Jika efek yang ditimbulkan oleh tanaman tersebut diminimalkan maka dapat menjadi solusi untuk menghadapi krisis pangan yang akan dihadapi masyarakat di masa depan

3.2 Langkah-Langkah Penelitian

Langkah-langkah yang akan dilakukan dalam penelitian ini, secara skematik dapat dilihat pada Gambar 3.1.

1. Pengambilan Data dari Laman Web

Data diperoleh melalui teknik *web scrapping* dengan menggunakan bahasa pemrograman R. Bagian yang diambil adalah identitas mahasiswa beserta komentarnya, kemudian disimpan dalam format “.csv”

2. *Text Preprocessing*

Text Preprocessing merupakan tahap awal sebelum data diproses secara *text mining*. Pada tahap ini dilakukan proses untuk mempersiapkan data teks agar dapat di proses pada tahapan berikutnya. Proses yang dilakukan pada tahapan ini antara lain :

a. *Case Folding*

Proses ini mengubah semua huruf pada teks menjadi huruf kecil, menghilangkan karakter seperti angka, tanda baca, dan *Uniform Resources Locator* (URL).

b. Normalisasi

Pada proses ini bertujuan menyeragamkan kata yang penulisannya berbeda tetapi bermakna sama.

c. *Stopword Removal*

Tahap ini bertujuan untuk menghapus dan memilih kata penting saja yang akan digunakan.

d. *Stemming*

Pada tahap ini akan dilakukan pemotongan kata ke bentuk dasar.

e. *Tokenization*

Pada tahap ini dilakukan pemotongan dokumen teks ke bentuk kata tunggal berdasarkan kata-kata yang menyusunnya.

f. Data Olah

Data Olah merupakan data yang diperoleh setelah semua tahap tersebut dilakukan data siap untuk di olah dengan metode-metode *text mining*.

3. Analisis Sentimen

Metode analisis sentimen yang digunakan adalah metode *lexicon based* dan *text mining*.

a. *Lexicon based*

Sentimen pada metode ini akan bergantung pada kamus. Pada tahap ini terdiri dari beberapa tahap :

1) Penentuan Sentimen

b. Klasifikasi *Naïve Bayes*

Terdapat beberapa langkah pada klasifikasi *Naïve Bayes*, seperti yang ditunjukkan pada Gambar 3.2

1) Pembagian data

Data olah yang telah didapatkan dari *pre processing* dibagi menjadi dua, yaitu data latih dan data uji. Pembagian data uji dan data latih adalah 80 % dan 20%

2) Pembuatan Model Klasifikasi *Naïve Bayes*

Data yang telah diketahui kelas datanya (data latih) digunakan untuk membentuk model

3) Model Klasifikasi *Naïve Bayes*

Model yang telah didapatkan dari sebelumnya, akan diuji dengan menggunakan data uji untuk mengetahui akurasi model tersebut

c. Evaluasi Model

d. *Topic Modelling*

Pada tahap ini akan dilakukan pengelompokkan kedalam beberapa topik yang bertujuan untuk merepresentasikan informasi yang ada dalam data teks.

Terdapat beberapa tahap, diantaranya:

1) Penentuan kandidat jumlah topik menggunakan nilai perplexity

2) Pemodelan topik

3) Analisis topik

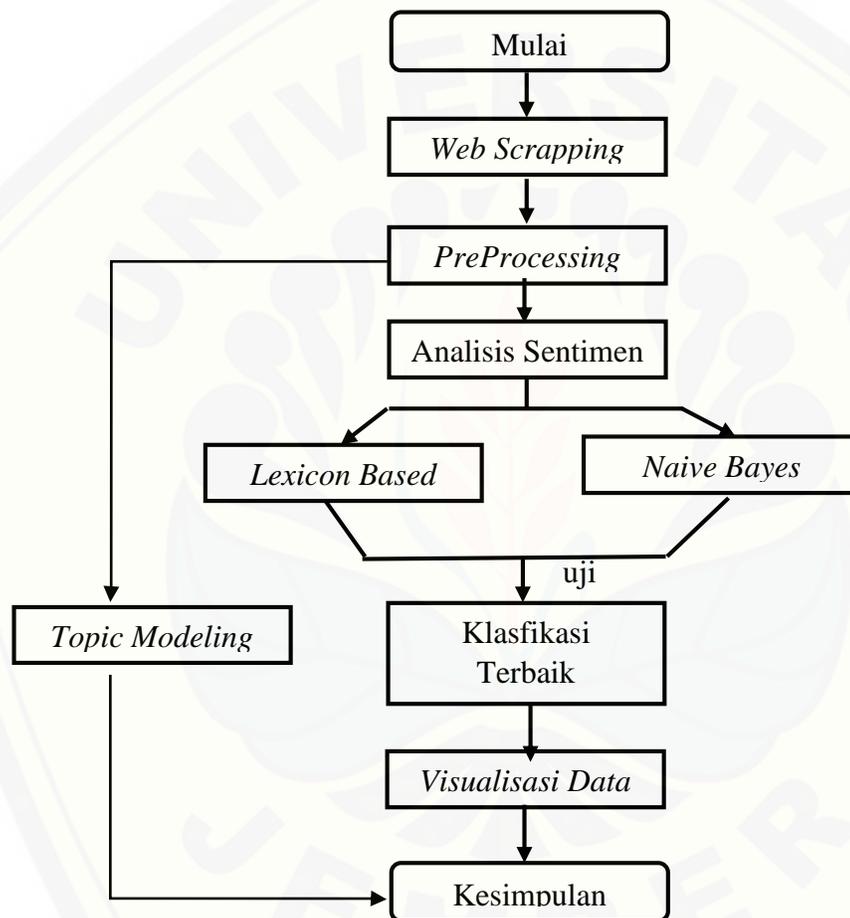
4. Visualisasi Data

Visualisasi dilakukan untuk mempermudah dalam menyimpulkan hingga menggambarkan karakteristik hingga korelasi data. Visualisasi dilakukan dengan

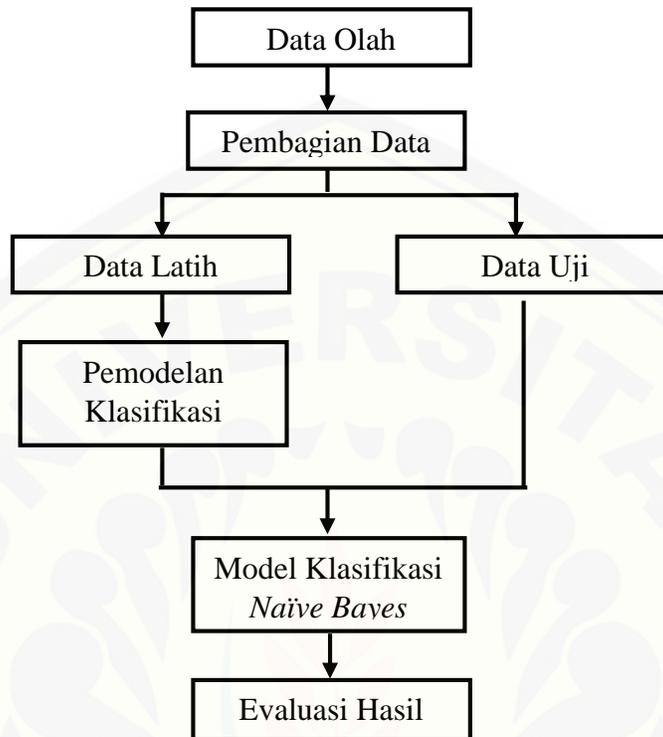
menggunakan metode *Word cloud* atau salah satu metode visualisasi yang bekerja dengan memunculkan kata-kata berdasarkan jumlah kata.

5. Kesimpulan

Langkah terakhir yaitu menarik kesimpulan dari hasil yang dibahas.



Gambar 3.1 Diagram Alir Langkah-langkah Penelitian



Gambar 3.2 Diagram alir proses klasifikasi *naïve bayes*

BAB 5. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa:

1. Sentimen yang dominan adalah sentimen positif. Artinya sebagian besar dari mahasiswa setuju dengan penggunaan tanaman transgenik sebagai bahan pangan, serta mendukung perkembangan tanaman transgenik
2. Hasil dari metode *Topic Modelling* adalah menghasilkan 10 topik. Berdasarkan 10 topik tersebut dapat ditarik kesimpulan bahwa tanaman transgenik menguntungkan banyak pihak. Akan tetapi perlu adanya peningkatan dan pengembangan pada tanaman transgenik terutama dampak dalam hal kesehatan dan lingkungan.

5.2 Saran

Saran untuk penelitian selanjutnya diharapkan dapat menggunakan metode analisis sentimen dengan metode *text mining* yang lainnya, atau dengan *hybrid* (penggabungan analisis *lexicon* dan *text mining*).

DAFTAR PUSTAKA

- Bagus, D. 2011. *Text Mining* pada Media Sosial Teitter Studi kasus : Pilkada DKI 2017 Putaran 2. *Disertasi*. Jember : Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Jember.
- Feldman, R. dan J. Sanger. 2007. *The Text Mining Handbook*. New York : Cambrige University Press.
- Han J., K. Micheline, dan Pei.J. 2011. *Data Mining : Concepts and Techniques. Third Edition*. San Francisco (US): Morgan Kaufmann.
- Jo, T. 2019. *Text Mining: Concepts, Implementation, and Big Data Challenge*. Switzerland :Springer Nature.
- Karmana, W.I. 2009. *Adopsi Tanaman Transgenik Dan Beberapa Aspek Pertimbangannya. GaneC Swara*. 3(2): 12
- Kementrian Pertahanan Republik Indonesia. 2015. *Buku Putih Pertahanan Indinesia*. Jakarta, Indonesia:Kementrian Pertahanan Republik Indonesia.
- Kumar, A., dan A. Paul. 2007. *Mastering Text Mining with R*. Brimingham. Packt Publishing Ltd.
- Losey, J., Raynor, L., and Carter, M.E. 1999. *Itransgenic pollen harms Monarach Iarvae*. Nature, 399, 214.
- Nurfalah A., Adiwijya., A. Suryani. 2017. *Analisis Sentimen Opini Berbahasa Dengan Pendekatan Lexicon-Based Pada Media Sosial. Jurnal Masyarakat Informatika Indonesia*. 2(1) : 1.
- Rasyadi, M. H. 2018. Analisis Sentimen Pada Twitter Menggunakan Metode Naïve Bayes (Studi Kasus Pemilihan Gubernur DKI Jakarta 2017). *Skripsi*. Bogor : Fakultas Matematika Dan Ilmu Pengetahuan Alam Institut Pertanian Bogor.

Lampiran 1. *Script* Pengambilan Data

```
> library(xml2)
> library("rvest")
> library("NLP")
> library("tm")
> library("stringr")
> library("katadasaR")
> library("tokenizers")
> setwd("D:/programku")

#scraping data dari web
> web.transgenik = read_html("http://matematika.fmipa.unej.ac.id/?p=2069")
> nama = web.transgenik %>%
  html_nodes("b")%>%
  html_text()
> nama = str_squish(nama)
> komen = web.transgenik %>%
  html_nodes(".comment-content")%>%
  html_text()
> komen= str_squish(komen)
> datatransgenik = data.frame(Identitas=nama, Komentar=komen)
> # mengubah data ke bentuk korpus
> ambildata = iconv(datatransgenik$Komentar, from = "UTF-8", to = "ASCII", s
ub = "")
> df= data.frame(doc_id=1:268,text=ambildata)
> korpus = Corpus(DataframeSource(df))
```

Lampiran 2. *Script Preprocessing*

```
#Preprocessing(case folding)
#merubah ke huruf kecil
> csfldg = tm_map(korpus,content_transformer(tolower))

#merubah tanda ("/") ke bentuk spasi
> tospace = content_transformer(function(x,pattern)gsub(pattern," ",x))
> csfldg = tm_map(csfldg,tospace,"/")

#menghilangkan tanda baca
> csfldg = tm_map(csfldg, tospace,"[[:punct:]]")
```

```
#menghilangkan angka
> csfldg = tm_map(csfldg, tospace, "[[:digit:]]")
#menghilangkan spasi yang berlebihan
> csfldg = tm_map(csfldg, stripWhitespace)

#preprocessing (normalisasi)
> change = tm_map(csfldg, gsub,pattern="akurat",replacement="tepat")
> change = tm_map(change, gsub,pattern="new",replacement="baru")
> change = tm_map(change, gsub,pattern="menimbulkan",replacement="dampak")
> change = tm_map(change, gsub,pattern="efek",replacement="dampak")
> change = tm_map(change, gsub,pattern="dg",replacement="dengan")
> change = tm_map(change, gsub,pattern="dengn",replacement="dengan")
> change = tm_map(change, gsub,pattern="mudarat",replacement="rugi")
> change = tm_map(change, gsub,pattern="mudorot",replacement="rugi")
> change = tm_map(change, gsub,pattern="risiko",replacement="akibat")
> change = tm_map(change, gsub,pattern="problematika",replacement="masalah")
> change = tm_map(change, gsub,pattern="yabg",replacement="yang")
> change = tm_map(change, gsub,pattern="alternative",replacement="alternatif")
> change = tm_map(change, gsub,pattern="goals",replacement="tujuan")
> change = tm_map(change, gsub,pattern="problem",replacement="masalah")
> change = tm_map(change, gsub,pattern="teknologi",replacement="teknologi")

#preprocessing (stopword removal)
> daftar.sword = file("sword.csv", open="r")
> id_stopwords = readLines(daftar.sword)
> close(daftar.sword)
> id_stopwords=c(id_stopwords,"amp")
> stpwr = tm_map(change, removeWords, id_stopwords)

#stopword tambahan
> swordku = c("tanaman","transgenik","kak","sih","serba","nya")
> stpwr = tm_map(stpwr, removeWords,swordku)
> dataframe = data.frame(doc_id =268, text=unlist(sapply(stpwr, `[`)),
  stringsAsFactors=F)
> data = dataframe$text %>%
  as.character()

#preprocessing (Stemming)
> stemming = function(x){ paste(lapply(x,katadasar),collapse = " ") }
> datastem = lapply(tokenize_words(data[]), stemming)
> df.stem = data.frame(doc_id = 1:268,text=unlist(sapply(datastem, `[`)),
  stringsAsFactors=F)
> korpusstem = Corpus(DataframeSource(df.stem))
```

Lampiran 3. *Script R Metode Lexicon Based*

```
> library(plyr)
> library(stringr)
> library(dplyr)
> library(tm)
#Data pada Lexicon Based
> setwd("D:/programku")
> pos= scan("haipos.csv", what = "character", comment.char = ";")
> neg= scan("haineg.csv", what = "character", comment.char = ";")
> databersih = read.csv("databersih.csv")
> data = databersih$text

#Pelabelan Sentimen
> pelabelan.sentimen = function(teks, pos, neg, .progress = "none")
{
  scores = laply(teks, function(teks, pos, neg) {
    kata = unlist(str_split(teks, "\\s+"))
    nilai.pos = !is.na(match(kata, pos))
    nilai.neg = !is.na(match(kata, neg))
    nilai = sum(nilai.pos) - sum(nilai.neg)
    return(nilai)
  }, pos, neg, .progress=.progress)
  x= scores
  label="Netral"
  for (l in 1:length(x)) {
    if (x[l] < 0){
      label[l] = "Negatif"
    } else if (x[l]> 0){
      label[l] = "Positif"
    } else {
      label[l] = "Netral"
    }
  }
  return(data.frame(text = teks, nilai = scores, sentimen = label))
}
> hasil = pelabelan.sentimen(data,pos,neg)

> #Evaluasi Confusion Matrix
> conf.mat = confusionMatrix(datauji, prediksi)
```

Lampiran 4. *Script Metode Naïve Bayes*

```
> library(tm)
> library(e1071)
> library(dplyr)
> library("caret")

> setwd("D:/PROGRAM")
> View(dataku)
> #DATA
> dataku= read.csv("data nb.csv", stringsAsFactors = FALSE)
> glimpse(dataku)
> #Pengacakan Data
> set.seed(10)
> data = dataku[sample(nrow(dataku)), ]
> data = dataku[sample(nrow(dataku)), ]
> glimpse(data)

> data$Kelas = as.factor(data$Kelas)
> korpus = Corpus(VectorSource(data$Teks))
> dtm = DocumentTermMatrix(korpus)

> #Pembagian Data Latih dan Data Uji
> DataLatih = data[1:80,]
> DataUji = data[81:100,]
> dtm.Latih = dtm[1:80,]
> dtm.Uji = dtm[81:100,]

> KorpusLatih = korpus[1:80]
> KorpusUji = korpus[81:100]

> frek = findFreqTerms(dtm.Latih,1)
> #length((frek))
> dtm.Latih.NB = DocumentTermMatrix(KorpusLatih, control=list(dictionary = frek))
> dtm.Uji.NB = DocumentTermMatrix(KorpusUji, control=list(dictionary = frek))
>
> #mengubah label
> convert_count = function(x) {
+ y = ifelse(x > 0, 1,0)
+ y = factor(y, levels=c(0,1), labels=c("No", "Yes"))
+ y
+ }
```

```
> latihNB = apply(dtm.Latih.NB, 2, convert_count)
> ujiNB = apply(dtm.Uji.NB, 2, convert_count)
> # latih the classifier
> Latih = naiveBayes(latihNB, DataLatih$Kelas, laplace = 1)
> Latih2 = naiveBayes(latihNB, DataLatih$Kelas, laplace = 1)

# Model klasifikasi
> prediksi = predict(Latih, newdata=ujiNB)
# Pconfusion matrix
> conf.mat = confusionMatrix(DataUji$Kelas, prediksi )
> perbandingan = data.frame("aktual"=DataUji$Kelas, "predik"=prediksi)

#DaTA BARU
> databaru= read.csv("Pelabelan Manual.csv", stringsAsFactors = FALSE)
> korpusbaru = Corpus(VectorSource(databaru$Teks))
> dtmbaru = DocumentTermMatrix(korpusbaru)
> BaruNB = apply(dtmbaru, 2, convert_count)
> prediksibaru = predict(Latih, newdata =BaruNB)

> library(RColorBrewer)
> library(wordcloud)

> #evaluasi hasil
> tdmbaru = TermDocumentMatrix(korpusbaru)
> m = as.matrix(tdmbaru)
> v = sort(rowSums(m),decreasing=TRUE)
> rat = data.frame(word = names(v),freq=v)
> head(rat, 20)

> set.seed(1234)
> wordcloud(words = rat$word, freq = rat$freq, min.freq = 4,
+           max.words=200, random.order=FALSE, rot.per=0.35,
+           colors=brewer.pal(8, "Dark2"))
> #barplot
> matriks=as.matrix(tdmbaru)

> w = rowSums(matriks)
> fh=colSums(matriks)
> w = subset(w, w>=15)
> barplot(w,
+         las = 2,
+         col = brewer.pal(8, "RdBu"))
```

Lampiran 5. *Script Topic Modelling*

```
> library("tm")
> library(topicmodels)
> library(ggplot2)

#Data
> setwd("D:/programku")
> datatopik = read.csv("databersih.csv")
> datatopik = iconv(datatopik$text, from = "UTF-8", to = "ASCII", sub = "")
> datatopik = data.frame(doc_id=1:268, text=datatopik)
> korpus.datatopik = Corpus(DataframeSource(datatopik))
> panjang = rowSums(as.matrix(DocumentTermMatrix(korpus.datatopik)))
> dtm = DocumentTermMatrix(korpus.datatopik[panjang > 0])

#Pemilihan K terbaik
> perplexity_df = data.frame(train=numeric())
> topics = c(2:30)
> burnin = 100
> iter = 100
> keep = 50
>
> set.seed(12345)
> for (i in topics){
  fitted = LDA(dtm, k = i, method = "Gibbs",
               control = list(burnin = burnin, iter = iter, keep = keep) )
  perplexity_df[i,1] = perplexity(fitted, newdata = dtm)
}
>
> g = ggplot(data=perplexity_df, aes(x= as.numeric(row.names(perplexity_df)))) + labs(y="Perplexity of hold out and training data") + ggtitle("Perplexity of hold out and training data")
> g = g + geom_line(aes(y=train), colour="red")

> # Now for some topics
> acak = sample(1:1000, 1) # Pick a random seed for replication
> k = 10 # Let's start with 10 topics
> models = list(
  VEM = LDA(dtm, k = k, control = list(seed = acak)),
  Gibbs = LDA(dtm, k = k, method = "Gibbs", control = list(seed = acak, burnin = 100, iter = 1000)) )
```

Lampiran 6. Hasil Proses Klasifikasi *Naïve Baye*

Iterasi 1

Aktual	Prediksi			Recall (%)
	Negatif	Netral	Positif	
Negatif	0	0	0	-
Netral	3	1	1	0,167
Positif	2	0	35	0,74
<i>Precision (%)</i>	0	0,09	0,94	
Akurasi = 67,9%				

Iterasi 2

Aktual	Prediksi			Recall (%)
	Negatif	Netral	Positif	
Negatif	0	0	0	-
Netral	2	3	1	0,5
Positif	5	13	29	0,62
<i>Precision (%)</i>	0	0,18	0,96	
Akurasi = 60,3%				

Iterasi 3

Aktual	Prediksi			Recall (%)
	Negatif	Netral	Positif	
Negatif	0	0	0	-
Netral	1	0	0	0
Positif	5	16	31	0,59
<i>Precision (%)</i>	0	0	1	
Akurasi = 58,4%				

Iterasi 4

Aktual	Prediksi			Recall (%)
	Negatif	Netral	Positif	
Negatif	0	0	0	-
Netral	1	1	0	0,5
Positif	4	9	38	0,74
<i>Precision (%)</i>	0	0,1	1	
Akurasi = 73,5%				

Iterasi 5

Aktual	Prediksi			Recall (%)
	Negatif	Netral	Positif	
Negatif	0	0	0	-
Netral	0	2	4	0,33
Positif	4	13	30	0,64
<i>Precision (%)</i>	0	0,13	0,88	
Akurasi = 60,3%				

Iterasi 6

Aktual	Prediksi			Recall (%)
	Negatif	Netral	Positif	
Negatif	0	0	0	-
Netral	0	0	0	-
Positif	2	11	40	0,75
<i>Precision (%)</i>	0	0	1	
Akurasi = 75,4%				

Iterasi 7

Aktual	Prediksi			Recall (%)
	Negatif	Netral	Positif	
Negatif	1	0	0	1
Netral	2	3	1	0,5
Positif	4	8	34	0,739
<i>Precision (%)</i>	0,14	0,27	0,97	
Akurasi = 71,6%				

Iterasi 8

Aktual	Prediksi			Recall (%)
	Negatif	Netral	Positif	
Negatif	0	0	0	-
Netral	1	1	1	0,3
Positif	4	17	29	0,58
<i>Precision (%)</i>	0	0,05	0,96	
Akurasi = 56,6%				

Iterasi 9

Aktual	Prediksi			Recall (%)
	Negatif	Netral	Positif	
Negatif	0	0	0	-
Netral	2	2	1	0,4
Positif	2	10	36	0,75
<i>Precision (%)</i>	0	0,167	0,97	
Akurasi = 71,7 %				

Iterasi 10

Aktual	Prediksi			Recall (%)
	Negatif	Netral	Positif	
Negatif	0	0	0	-
Netral	2	1	0	0,3
Positif	5	9	36	0,72
<i>Precision (%)</i>	0	0,1	1	
Akurasi = 69,8%				