



**OPTIMASI *SUPPORT VECTOR MACHINE* (SVM)
MENGUNAKAN *GRID SEARCH* DAN *UNIGRAM* GUNA
MENINGKATKAN AKURASI *REVIEW* PADA SITUS *E-
COMMERCE***

Skripsi

disusun sebagai salah satu syarat
untuk memperoleh gelar Sarjana Komputer
Program Studi Teknik Informatika

Oleh

Sulistiana
4611416008

**JURUSAN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SEMARANG**

2020

PERNYATAAN

Saya menyatakan bahwa skripsi saya yang berjudul “Optimasi *Support Vector Machine* (SVM) Menggunakan *Grid Search* dan *Unigram* Guna Meningkatkan Akurasi *Review* pada Situs *E-Commerce*” disusun atas dasar penelitian saya dengan arahan dosen pembimbing. Sumber informasi atau kutipan yang berasal dari karya yang diterbitkan telah disebutkan dalam teks dan dicantumkan dalam daftar pustaka di bagian akhir skripsi ini. Dan saya menyatakan bahwa skripsi ini bebas plagiat dan apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya bersedia menerima sanksi sesuai ketentuan perundang-undangan.

Semarang, 12 Mei 2020



Sulistiana
4611416008

PERSETUJUAN PEMBIMBING

Nama : Sulistiana
NIM : 4611416008
Program Studi : Teknik Informatika S1
Judul Skripsi : Optimasi *Support Vector Machine* (SVM) Menggunakan
Grid Search dan *Unigram* Guna Meningkatkan Akurasi
Review pada Situs *E-Commerce*

Skripsi ini telah disetujui oleh pembimbing untuk diajukan ke sidang panitia ujian skripsi Program Studi Teknik Informatika FMIPA UNNES.

Semarang, 12 Mei 2020

Pembimbing



Much Aziz Muslim, S.Kom., M.Kom.
NIP 197404202008121001

PENGESAHAN

Skripsi yang berjudul

Optimasi *Support Vector Machine* (SVM) Menggunakan *Grid Search* dan
Unigram Guna Meningkatkan Akurasi *Review* pada Situs *E-Commerce*

disusun oleh

Sulistiana

4611416008

Telah dipertahankan di hadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada
tanggal 12 Mei 2020.



Sekretaris

Dr. Alamsyah, S.Si., M.Kom.
NIP 197405172006041001

Penguji 1

Dr. Alamsyah, S.Si., M.Kom.
NIP 197405172006041001

Penguji 2

Budi Prasetyo, S.Si., M.Kom.
NIP 198805012014041001

Anggota Penguji/
Pembimbing

Much Aziz Muslim, S.Kom., M.Kom.
NIP 197404202008121001

MOTTO DAN PERSEMBAHAN

MOTTO

- Hiduplah seperti pohon pisang, yang tidak akan mati sebelum berbuah.
- Melakukan sesuatu yang ada manfaatnya atau ada uangnya (Sulistiana).

PERSEMBAHAN

Skripsi ini saya persembahkan kepada:

- Kedua Orang Tua saya Bapak Sobirin dan Ibu Suburoh yang telah bekerja keras melakukan apapun demi kelancaran anaknya dalam menuntut ilmu, yang selalu memberikan kasih sayang dan do'a, menjadi pelabuhan disaat sulis *down* dan memberikan semangat yang membuat sulis bangkit kembali.
- Abah KH. Drs. M. Masroni dan Ibu Nyai Hj. Umi Chudloifah Al Khafidzah yang telah mendidik sulis dengan sabar selama di Semarang, memberikan semangat serta do'a.
- Adik-adik saya, Fitrohatun Nabila, Nazilus Sakinah, dan Zahra Izzatin Nahjua yang telah memberikan semangat, dukungan, serta do'a.
- Teman-teman saya di jurusan Ilmu Komputer, FMIPA, UNNES dan kamar An Najjahy PP. SGJB Semarang.
- Semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu hingga terselesaikannya penulisan skripsi ini.
- Almamater, Universitas Negeri Semarang.

PRAKATA

Puji syukur penulis panjatkan kepada Allah *Subhanahu wa ta'ala* atas berkat rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul **“Optimasi *Support Vector Machine* (SVM) Menggunakan *Grid Search* dan *Unigram* Guna Meningkatkan Akurasi *Review* pada Situs *E-Commerce*”**.

Penulis menyadari bahwa penulisan skripsi ini tidak akan selesai tanpa adanya dukungan serta bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada:

1. Prof. Dr. Fathur Rokhman, M.Hum., Rektor Universitas Negeri Semarang.
2. Dr. Sugianto, M.Si., Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
3. Dr. Alamsyah, S.Si., M.Kom., Ketua Jurusan Ilmu Komputer FMIPA Universitas Negeri Semarang yang telah meluangkan waktu, membantu, membimbing, mengarahkan dan memberikan saran sehingga penulis dapat menyelesaikan skripsi ini.
4. Much Aziz Muslim, S.Kom., M.Kom., Dosen Pembimbing yang telah meluangkan waktu, membantu, membimbing, mengarahkan dan memberikan saran sehingga penulis dapat menyelesaikan skripsi ini.
5. Bapak dan Ibu Dosen Jurusan Ilmu Komputer yang telah memberikan bekal kepada penulis dalam penyusunan skripsi ini.
6. Kedua Orang Tua saya Bapak Sobirin dan Ibu Suburoh yang telah bekerja keras melakukan apapun demi kelancaran anaknya dalam menuntut ilmu,

yang selalu memberikan kasih sayang dan do'a, menjadi pelabuhan disaat sulis *down* dan memberikan semangat yang membuat sulis bangkit kembali.

7. Abah KH. Drs. M. Masroni dan Ibu Nyai Hj. Umi Chudloifah Al Khafidzah yang telah mendidik sulis dengan sabar selama di Semarang, memberikan semangat serta do'a.
8. Adik-adik saya, Fitrohatun Nabila, Nazilus Sakinah, dan Zahra Izzatin Nahjua yang telah memberikan semangat, dukungan, serta do'a.
9. Segenap keluarga besar dari Bapak Sopari - Ibu Arifah Alm, serta Bapak As'ad Alm – Ibu Sriati, khususnya juga Ma'Eng – Pak De Imron Alm yang sudah mengurus sulis waktu kecil.
10. Kamu sebagai orang kiriman Allah atas terijabahnya do'aku untuk minta didatangkan orang *special* yang menemani dan menyemangatti proses mengerjakan skripsi hingga selesai, serta mendampingi kala wisuda bersama orang tua.
11. Teman-teman terdekat saya, Anind, Anisa, Esa, Tata, Rio, dan Wina sebagai teman diskusi yang banyak memberikan dukungan dan motivasi.
12. Teman-teman saya di jurusan Ilmu Komputer, khususnya teman-teman ilkom angkatan 2016 dan teman-teman Pondok Pesantren Sunan Gunung Jati Ba'alawy Semarang, khususnya kamar An-Najjahy khususnya Mba Irin, Kak lala, Dek Aliyah, Mba Isti, Dek Iin, Dek Bella, Dek Nada, Dek Itoh, Dek Ita, Deak Asih, Mba Mahda, Mb Ulya, Mb Rahma yang telah memberikan semangat dan dukungannya.

13. Teman-teman organisasi PKPT IPNU IPPNU UNNES, Anak Komisariat Hasyim Asy'ari IPNU IPPNU UNNES, PAC IPNU IPPNU Gunungpati, BPH ILKOM 2018, dan Al Husna 2018, yang telah memberikan pengalaman selama kuliah.
14. Semua pihak yang telah membantu terselesaikannya skripsi ini yang tidak dapat penulis sebutkan satu persatu, terimakasih atas bantuannya.

Semoga skripsi ini dapat memberikan manfaat bagi pembaca di masa yang akan datang.

Semarang, 12 Mei 2020

Penulis



Sulistiana
4611416008

ABSTRAK

Sulistiana. 2020. Optimasi *Support Vector Machine* (SVM) Menggunakan *Grid Search* dan *Unigram* Guna Meningkatkan Akurasi *Review* pada Situs *E-Commerce*. Skripsi, Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing Much Aziz Muslim, S.Kom., M.Kom.

Kata kunci: *Text Mining*, *Support Vector Machine*, *Unigram*, *Grid Search*

E-Commerce (*Electronic Commerce*) adalah proses distribusi, pembelian, penjualan, pemasaran barang, dan jasa melalui sistem elektronik seperti internet, televisi, web, atau jaringan komputer lainnya. Banyak situs *E-Commerce* seperti *amazon.com* dan *lazada.co.id* yang menyediakan berbagai macam produk dengan harga dan kualitas yang bervariasi. Analisis sentimen digunakan untuk memahami popularitas suatu produk diantara ulasan pembeli. Analisis sentimen memiliki beberapa pendekatan, salah satunya *machine learning*. Bagian dari *machine learning* yang berfungsi untuk memproses teks yaitu *text mining*. Salah satu teknik dalam *text mining* yaitu klasifikasi. *Support Vector Machine* (SVM) merupakan algoritma yang sering digunakan dalam klasifikasi. Pemilihan fitur dan penentuan parameter di SVM secara signifikan mempengaruhi hasil akurasi klasifikasi. *Unigram* digunakan sebagai *feature extraction* dan *Grid Search* sebagai optimasi parameter agar dapat meningkatkan akurasi pengklasifikasi SVM. Dalam penelitian ini menggunakan dua *dataset* yaitu *review Amazon* dan *review Lazada*. Pengukuran akurasi diukur dengan *confusion matrix* dengan 10 *cross validation*. Hasil akurasi penerapan *Unigram* dan *Grid Search* pada algoritma SVM terbukti cukup bagus dibandingkan dengan penelitian sebelumnya yaitu untuk data *review Amazon* dapat ditingkatkan sebesar 26,4%, sedangkan dalam penelitian sebelumnya hanya sebesar 19,2%. Untuk data *review Lazada* dapat ditingkatkan sebesar 4,26%, sedangkan pada penelitian sebelumnya hanya sebesar 1,06%.

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
PERNYATAAN.....	ii
PERSETUJUAN PEMBIMBING.....	iii
PENGESAHAN	iv
MOTTO DAN PERSEMBAHAN	v
PRAKATA.....	vi
ABSTRAK	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xv
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	4
1.3. Batasan Masalah	5
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	6
1.6. Sistematika Penulisan	6
1.6.1 Bagian Awal Skripsi	6
1.6.2 Bagian Isi Skripsi	7
1.6.3 Bagian Akhir Skripsi.....	7
BAB 2 TINJAUAN PUSTAKA	8
2.1. Tinjauan Pustaka.....	8

2.2.	Landasan Teori.....	9
2.2.1	<i>Text Mining</i>	9
2.2.2	Klasifikasi.....	12
2.2.3	<i>Support Vector Machine</i>	12
2.2.4	<i>Grid Sreach</i>	15
2.2.5	<i>Unigram</i>	16
2.2.6	<i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	17
2.2.7	Validasi dan Evaluasi.....	18
BAB 3 METODE PENELITIAN.....		20
3.1.	Studi Literatur	20
3.2.	Perancangan Sistem	20
3.3.	Pengambilan Data	21
3.4.	Pengolahan Data	21
3.5.	<i>Preprocessing Data</i>	24
3.5.1	<i>Transform Cases</i>	24
3.5.2	<i>Punctuation Removal</i>	25
3.5.3	<i>Tokenize</i>	25
3.5.4	<i>Stopword Removal</i>	25
3.5.5	<i>Stemming</i>	26
3.6.	Metode yang Digunakan.....	27
3.7.	Penarikan Kesimpulan	30
BAB 4 HASIL DAN PEMBAHASAN.....		31
4.1	Hasil Penelitian	31
4.1.1	Hasil <i>Preprocessing</i>	31
4.1.1.1	Hasil <i>Transform Cases</i>	31
4.1.1.2	Hasil <i>Punctuation Removal</i>	32

4.1.1.3	Hasil <i>Tokenize</i>	32
4.1.1.4	Hasil <i>Stopword Removal</i>	33
4.1.1.5	Hasil <i>Stemming</i>	34
4.1.2	Hasil <i>Text Mining</i>	35
4.1.2.1	Hasil Klasifikasi dengan Algoritma SVM.....	35
4.1.2.2	Hasil Klasifikasi Algoritma SVM dengan menerapkan <i>Unigram</i>	36
4.1.2.3	Hasil Klasifikasi Algoritma SVM dengan menerapkan <i>Unigram</i> dan <i>Grid Search</i>	37
4.1.3.	Hasil Perbandingan dengan Penelitian Sebelumnya.....	41
4.1.4	Hasil Implementasi Sistem	43
4.2.	Pembahasan.....	49
BAB 5 PENUTUP.....		51
5.1.	Kesimpulan	51
5.2.	Saran	52
DAFTAR PUSTAKA		53
LAMPIRAN.....		58

DAFTAR TABEL

Tabel	Halaman
Tabel 2.1. <i>Confussion Matrix</i> untuk Klasifikasi Dua Kelas.....	18
Tabel 3.1 <i>Dataset</i> dalam Bentuk <i>Excel</i>	23
Tabel 4.1 Hasil <i>Transform Cases</i>	31
Tabel 4.2 Hasil <i>Punctuation Removal</i>	32
Tabel 4.3 Hasil <i>Tokenize</i>	33
Tabel 4.4 Hasil <i>Stopword Removal</i>	33
Tabel 4.5 Hasil <i>Stemming</i>	34
Tabel 4.6 Hasil Pengujian SVM dengan <i>Unigram</i> dan <i>Grid Search</i> pada Data <i>Review Amazon</i>	38
Tabel 4.7 Hasil Pengujian SVM dengan <i>Unigram</i> dan <i>Grid Search</i> pada Data <i>Review Lazada</i>	39
Tabel 4.8 Peningkatan Akurasi SVM dengan <i>Unigram</i> dan <i>Grid Search</i> pada Data <i>Review Amazon</i>	40
Tabel 4.9 Peningkatan Akurasi SVM dengan <i>Unigram</i> dan <i>Grid Search</i> pada Data <i>Review Lazada</i>	41
Tabel 4.10 Perbandingan Hasil Penelitian Sekarang dengan Penelitian Sebelumnya.....	42

DAFTAR GAMBAR

Gambar	Halaman
Gambar 2.1. Proses <i>Text mining</i>	10
Gambar 2.2. Pemisahan Dua Kelas Data dengan <i>Margin</i> Maksimum	13
Gambar 3.1. Tampilan <i>Dataset</i> dengan Format .ft.....	22
Gambar 3.2. Tampilan <i>Dataset</i> dengan Format .csv.....	23
Gambar 3.3. <i>Flowchart</i> Algoritma SVM dengan <i>Grid Search</i> dan <i>Unigram</i>	28
Gambar 4.1. Tampilan <i>Home</i> Aplikasi	44
Gambar 4.2. Tampilan Pilihan <i>Dataset</i>	44
Gambar 4.3. Tampilan <i>Dataset</i>	45
Gambar 4.4. Tampilan Hasil <i>Preprocessing</i>	46
Gambar 4.5. Tampilan Hasil Akurasi Perhitungan Algoritma untuk <i>Dataset Review Amazon</i>	47
Gambar 4.6. Tampilan Hasil Akurasi Perhitungan Algoritma untuk <i>Dataset Review Lazada</i>	47
Gambar 4.7. Tampilan Menu <i>Profile</i>	48
Gambar 4.8. Tampilan Menu <i>About</i>	49

DAFTAR LAMPIRAN

Lampiran	Halaman
1 Kode Program Sistem GUI <i>Python</i>	56
2 <i>Dataset Review Amazon</i>	70
3 <i>Dataset Review Lazada</i>	73

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Situs komersial dunia hampir sepenuhnya dijalankan dalam *platform online*. Orang memperdagangkan produk melalui situs *e-commerce* yang berbeda (Haque *et al.*, 2018). Produk dapat dibeli di berbagai situs dengan harga yang bervariasi. Pelanggan biasanya menginginkan produk dengan kualitas terbaik dan harga yang rendah tetapi tidak bisa langsung memeriksanya. Ulasan dari pelanggan lain menjadi cara yang paling diandalkan untuk memutuskan apakah akan membeli produk atau tidak (Haque *et al.*, 2018). Ulasan ini memiliki informasi penting tentang masalah pelanggan dan pengalaman mereka yang dapat berguna untuk desain konseptual, personalisasi, rekomendasi produk, pemahaman pelanggan yang lebih baik, dan akuisisi pelanggan (Zhan *et al.*, 2009). Untuk memahami popularitas suatu produk di antara ulasan pembeli di seluruh dunia dapat dilakukan menggunakan analisis sentimen.

Analisis sentimen adalah suatu bidang studi yang perkembangannya melibatkan proses evaluasi dan membedakan pendapat atau emosi yang diungkapkan dalam teks (Ravi & Khettry, 2019). Konsep penting dalam analisis sentimen adalah mengidentifikasi polaritas teks secara keseluruhan untuk mengetahui apakah teks bernilai positif, negatif, atau netral (Ramanathan & Meyyappan, 2019). Secara umum, analisis sentimen diaplikasikan ke dalam tiga

pendekatan: (1) pendekatan berbasis aturan, dimana analisis perasaan didasarkan pada seperangkat aturan yang dibuat secara manual; (2) pendekatan otomatis, yang mengandalkan teknologi *machine learning* untuk belajar dari data; (3) pendekatan *hybrid*, yang menggabungkan pendekatan berbasis manual dan pendekatan otomatis (Saura & Bennett, 2019).

Text mining berbeda dengan *data mining*. *Data mining* difokuskan pada penemuan pola yang menarik dari basis data besar. Sedangkan *text mining* adalah bidang penting dalam konteks data mining untuk menemukan pola yang menarik dalam data tekstual (Sbou, 2018). Salah satu teknik dalam *text mining* yang dapat digunakan untuk memprediksi sebuah keputusan adalah klasifikasi.

Klasifikasi merupakan proses menemukan model dari sebuah data dengan tujuan mengambil suatu keputusan dengan memprediksi suatu kasus berdasarkan klasifikasi yang diperoleh (Zuhdi *et al.*, 2019). Klasifikasi dapat digunakan dalam proses pencarian sekumpulan model atau fungsi yang membedakan kelas-kelas data atau kategori, yang tujuannya agar model tersebut dapat digunakan untuk memprediksi kelas target secara akurat (Parul Sinha & Poonam Sinha, 2015). Terdapat beberapa cara untuk mendapatkan hasil yang akurat. Salah satunya dengan dilakukan tahap *preprocessing*. Teknik *preprocessing* dapat membantu meningkatkan kualitas data dan menghasilkan hasil yang lebih akurat karena kualitas data menentukan kinerja metode prediksi dan kegunaan dari pengetahuan yang diekstraksi (Asgarnezhad *et al.*, 2017).

Klasifikasi di dalam *machine learning* membutuhkan identifikasi ciri-ciri khusus atau fitur yang dimiliki oleh data untuk menentukan kategori kelas. Proses

identifikasi fitur disebut dengan *feature extraction*. Menurut Laoh *et al.*, (2019) penerapan *N-Gram* untuk *feature extraction* dapat meningkatkan akurasi dalam tinjauan klasifikasi sentimen.

Salah satu jenis algoritma untuk klasifikasi yaitu *Support Vector Machine* (SVM) (Muslim *et al.*, 2019). SVM adalah metode *Supervised Machine Learning* yang efektif dan memberikan kinerja generalisasi yang baik untuk menyelesaikan masalah klasifikasi dan regresi (Byun & Lee, 2002). Tujuan dari SVM untuk klasifikasi adalah menentukan *hyper plane* secara optimal untuk memisahkan dua kelas dengan menggunakan set *data train* dan kemampuan generalisasinya diverifikasi menggunakan set *data test* (Kavzoglu, 2009). Kinerja SVM sangat tergantung pada kernel (Amari & Wu, 1999). SVM memiliki beberapa kernel, seperti: linear, polinomial, dan fungsi dasar radial (RBF).

Kernel linear adalah kernel paling sederhana dan hanya memiliki satu parameter (Tharwat, 2019). Gutiérrez *et al.* (2018) dalam penelitiannya membandingkan 3 tipe kernel SVM yaitu linear, radial, dan polinomial pada data *review* siswa terhadap guru. Kernel linear terbukti memperoleh akurasi terbaik yaitu 80,38%, kemudian disusul dengan kernel radial dan polinomial yaitu 78,50% dan 67,79%. Masing-masing kernel memiliki parameter khusus untuk mencapai kinerja terbaik.

Pengaturan parameter memiliki dampak signifikan pada peningkatan akurasi (Zakrani *et al.*, 2018). Oleh karena itu, membangun model yang akurat membutuhkan pemilihan nilai optimal dari parameter pembelajarannya. *Grid Search* adalah metode optimasi parameter yang penggunaannya dikombinasikan

dengan *cross validation*. Keuntungan utama dari *Grid Search* adalah akurasi pembelajaran yang tinggi dan kemampuan pemrosesan paralel pada pelatihan setiap SVM, karena independen satu sama lain (Zakrani *et al.*, 2018).

Pada penelitian Ravi dan Khettry (2019) menghitung akurasi *review E-Commerce* yaitu *Amazon* menggunakan *feature extraction bigram* serta *preprocessing* meliputi *tokenize*, *punctuation removal*, *stopword removal*, dan *stemming* pada empat algoritma klasifikasi yaitu *Naive Bayes*, *SVM*, *Random Forest Classifier*, dan *K-Nearest Neighbor* menghasilkan akurasi terendah pada SVM yaitu 62,5%, sedangkan *Naive Bayes*, *Random Forest Classifier*, dan *K-Nearest Neighbor* secara urut yaitu 70%, 77,65%, dan 65%.

Berdasarkan uraian di atas, maka penelitian berfokus pada peningkatan akurasi *Support Vector Machine* menggunakan *Grid Search* yang dikombinasikan dengan *feature extraction Unigram* untuk *review E-Commerce* dengan mengambil judul “**OPTIMASI SUPPORT VECTOR MACHINE (SVM) MENGGUNAKAN GRID SEARCH DAN UNIGRAM GUNA MENINGKATKAN AKURASI REVIEW PADA SITUS E-COMMERCE**”.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana cara kerja metode *Grid Search* dan *Unigram* yang diaplikasikan pada algoritma SVM dalam Meningkatkan Akurasi *Review* pada Situs *E-Commerce*?

2. Bagaimana tingkat akurasi hasil yang diperoleh dari algoritma SVM dengan menerapkan metode *Grid Search* sebagai optimasi parameter dan *Unigram* sebagai *feature extraction* pada *review* situs *E-Commerce*?
3. Bagaimana hasil peningkatan akurasi algoritma SVM metode *Grid Search* dan *Unigram* yang dibandingkan dengan penelitian sebelumnya.

1.3. Batasan Masalah

Pada penelitian ini diperlukan batasan-batasan agar tujuan penelitian dapat tercapai. Adapun batasan masalah yang dibahas dalam penelitian ini adalah:

1. Algoritma klasifikasi yang digunakan dalam penelitian ini adalah SVM,
2. Data yang digunakan dalam penelitian ini adalah *review* situs *E-Commerce* dari *Amazon* dan *Lazada* dari web *kaggle.com* dengan jumlah 1500 data pada masing-masing *E-Commerce* dan terdiri dari ulasan positif dan negatif,
3. Optimasi parameter SVM yang digunakan adalah *Grid Search*,
4. *Feature extraction* yang digunakan adalah *Unigram*,
5. Validasi yang digunakan adalah *10 cross validation*,
6. Bahasa pemrograman yang digunakan dalam penelitian adalah Bahasa *Python*.

1.4. Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Untuk mengetahui cara kerja metode *Grid Search* sebagai optimasi parameter kernel SVM dan *Unigram* sebagai *feature extraction* pada *review* situs *E-Commerce*,
2. Untuk mengetahui tingkat akurasi pada algoritma SVM sebelum dan sesudah diterapkan *Grid Search* dan *Unigram* pada *review* situs *E-Commerce*,
3. Untuk mengetahui perbandingan peningkatan akurasi SVM metode *Grid Search* dan *Unigram* dengan penelitian sebelumnya.

1.5. Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

1. Dengan diterapkannya metode *Grid Search* dan *Unigram* pada algoritma SVM yang diimplementasikan dalam suatu sistem berbasis *web* dapat membantu penulis dan masyarakat dalam menentukan pilihan suatu produk *E-Commerce* sesuai yang diinginkan,
2. Dalam lingkungan akademis diperoleh pengetahuan terhadap peningkatan akurasi SVM metode *Grid Search* dan *Unigram* dari penelitian sebelumnya.

1.6. Sistematika Penulisan

1.6.1 Bagian Awal Skripsi

Bagian awal skripsi terdiri dari halaman judul, halaman pengesahan, halaman pernyataan, halaman motto dan persembahan, abstrak, kata pengantar, daftar isi, daftar gambar, daftar tabel dan daftar lampiran.

1.6.2 Bagian Isi Skripsi

Bagian isi skripsi terdiri dari lima bab, yaitu sebagai berikut.

1. **BAB 1: PENDAHULUAN**

Bab ini berisi tentang latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian serta sistematika penulisan skripsi.

2. **BAB 2: TINJAUAN PUSTAKA**

Bab ini berisi penjelasan mengenai definisi maupun pemikiran-pemikiran yang dijadikan kerangka teoritis yang mendasari pemecahan masalah dalam skripsi ini.

3. **BAB 3: METODE PENELITIAN**

Bab ini berisi penjelasan mengenai studi pendahuluan, tahap pengumpulan data, dan tahap pengembangan sistem.

4. **BAB 4: HASIL DAN PEMBAHASAN**

Bab ini berisi hasil penelitian beserta pembahasannya.

5. **BAB 5: PENUTUP**

Bab ini berisi simpulan dari penulisan skripsi dan saran yang diberikan penulis untuk mengembangkan skripsi ini.

1.6.3 Bagian Akhir Skripsi

Bagian akhir skripsi ini berisi daftar pustaka yang merupakan informasi mengenai buku-buku, sumber-sumber dan referensi yang digunakan penulis serta lampiran-lampiran yang mendukung dalam penulisan skripsi ini.

BAB 2

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Penelitian ini dikembangkan dari beberapa referensi penelitian terdahulu yang mempunyai keterkaitan dengan metode dan objek penelitian. Penggunaan referensi ini ditujukan untuk memberikan batasan-batasan terhadap metode yang nantinya akan dikembangkan lebih lanjut. Berikut adalah hasil dari penelitian sebelumnya.

Ravi & Khettry (2019) dalam jurnalnya yang berjudul “*Amazon Reviews as Corpus for Sentimen Analysis Using Machine Learning*” yang mengidentifikasi empat algoritma klasifikasi *text mining* di antaranya *Bayesen Network (Naive Bayes)*, *Support Vector Machine (SVM)*, *Random Forest Classifier* dan *K-Nearest Neighbor* untuk menentukan tingkat akurasi dan performa yang lebih baik dalam penggunaannya sebagai *review data Amazon*. Hasil akurasi yang diperoleh dari algoritma klasifikasi berturut-turut adalah 70%; 62,5%; 77,5%; dan 65%.

Lameski *et al.*, (2015) dalam penelitiannya yang berjudul “*SVM Parameter Tuning with Grid Search and Its Impact on Reduction of Model Over-fitting*” menunjukkan bahwa penggunaan *Grid Search* dalam berbagai jenis *dataset* yang diuji dapat menyempurnakan parameter *SVM* dan kinerjanya dapat ditingkatkan secara signifikan.

Laoh *et al.*, (2019) dalam penelitiannya yang berjudul “*Enhancing Hospitality Sentimen Reviews Analysis Performance using SVM N-Gram Method*” menghasilkan bahwa penerapan *N-Gram* dibanyak literatur dapat meningkatkan akurasi dalam tinjauan klasifikasi sentimen dengan akurasi rata-rata 94%.

Gutiérrez *et al.*, (2018) dalam penelitiannya yang berjudul “*Analyzing Students Reviews of Teacher Performance Using Support Vector Machines by a Proposed Model*” yang membandingkan tiga kernel *SVM* (Linear, Radial, dan Polinomial) yang mempengaruhi keakuratan dalam klasifikasi komentar. Dari perbandingan tiga kernel tersebut, kernel Linear memperoleh akurasi tertinggi yaitu 80,38% dibandingkan kernel Radial dan Polinomial yaitu 78,50% dan 67,79.

2.2. Landasan Teori

2.2.1 Text Mining

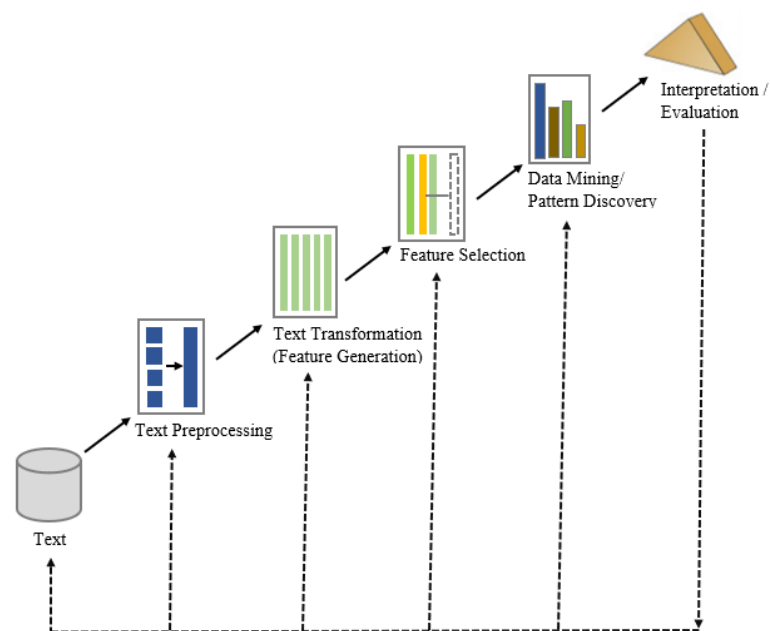
Text mining memiliki definisi menambang data yang berupa teks di mana sumber data biasanya didapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen (Nurzahputra & Muslim, 2016). Tujuan dari *text mining* adalah untuk mengubah teks menjadi atribut numerik yang kemudian dapat digunakan dalam algoritma penambangan data (Williams & Gong, 2014).

Gupta & Lehal (2009) dalam penelitiannya menyurvei tentang pentingnya penggunaan *text mining*. Pada penelitian tersebut disampaikan bahwa sebagian besar informasi (lebih dari 80%) disimpan sebagai teks, penambangan teks diyakini memiliki nilai potensial komersial yang tinggi. Pengetahuan dapat ditemukan dari

banyak sumber informasi, namun, teks-teks yang tidak terstruktur tetap menjadi sumber pengetahuan terbesar yang tersedia.

Allahyari *et al.* (2017) dalam penelitiannya juga membahas tentang *text mining*. Dalam penelitian tersebut disampaikan bahwa adanya jumlah teks yang dihasilkan setiap hari meningkat secara dramatis. Hal tersebut menyebabkan sebagian besar teks yang tidak terstruktur tidak dapat dengan mudah diproses dan dirasakan oleh komputer. Oleh karena itu, teknik dan algoritma yang efisien dan efektif diperlukan untuk menemukan pola yang bermanfaat. Dalam penelitian tersebut juga dijelaskan beberapa teknik dalam *text mining*, diantaranya yaitu klasifikasi, *clustering*, dan *information extraction*.

.Dalam *text mining* terdapat tahapan-tahapan proses. Proses *text mining* diilustrasikan pada Gambar 2.1. (Anadakumar, 2013).



Gambar 2.1. Proses *Text Mining*

Pada Gambar 2.1. terdapat 5 (lima) tahap yaitu sebagai berikut:

1. *Text preprocessing*

Pada tahap ini dilakukan upaya pengolahan data teks yang tidak terstruktur (*unstructured*) menjadi terstruktur dan menghilangkan bagian-bagian yang tidak diperlukan dalam proses *mining*.

2. *Text transformation*

Text transformation atau *feature generation* adalah tahapan untuk menghasilkan/membangkitkan *feature* dari data teks yang diolah. *Feature* adalah pola menarik yang dianggap dapat mewakili data sebenarnya. *Feature* merupakan bentuk representatif dari data. Salah satu bentuk *feature* dari teks adalah bentuk *sequence of word*.

3. *Feature selection*

Seluruh *feature* yang dihasilkan harus relevan dan bermanfaat untuk proses–proses selanjutnya, sehingga pada tahap ini dilakukan seleksi untuk menghasilkan *feature* yang benar–benar bermanfaat.

4. *Data Mining*

Tahap *data mining* adalah tahap *mining* yang sebenarnya. Pada tahap ini akan dilakukan pencarian pola menarik dari sekumpulan data yang telah dibersihkan dan diseleksi sebelumnya

5. *Interpretation / evaluation*

Hasil *mining* yang didapat pada tahap sebelumnya akan dievaluasi lebih lanjut.

2.2.2 Klasifikasi

Klasifikasi adalah suatu proses untuk mengelompokkan sejumlah data ke dalam kelas-kelas tertentu yang sudah diberikan berdasarkan kemiripan sifat dan pola yang terdapat dalam setiap data (Prasetyo *et al.*, 2019). Klasifikasi bertujuan untuk memprediksi kelas dari suatu objek yang labelnya tidak diketahui (Eka Ratnawati *et al.*, 2014).

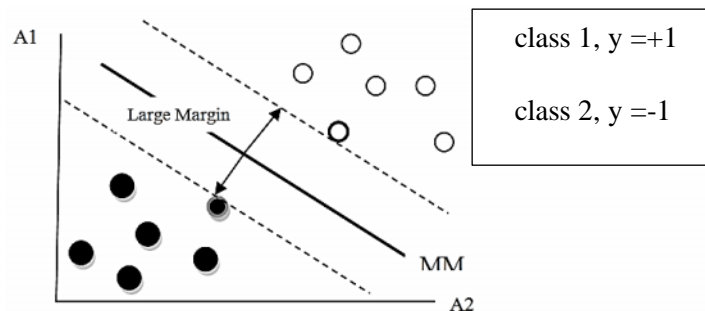
Sartika & Indra (2017) dalam penelitiannya menjelaskan dua proses klasifikasi yaitu tahap pembelajaran dan tahap pengklasifikasian. Tahap pembelajaran merupakan tahapan dalam pembentukan model klasifikasi, sedangkan tahap pengklasifikasian merupakan tahapan penggunaan model klasifikasi untuk memprediksi label kelas dari suatu data.

Klasifikasi terbagi atas 2 jenis (Kamavisdar *et al.*, 2013), yaitu: (1) *Supervised Classification* merupakan proses menggunakan data yang telah diketahui untuk mengklasifikasikan data asing. Contoh algoritma *Logistic Regression*, *Naïve Bayes*, *SVM (Support Vector Machine)*, *ANN (Artificial Neural Networks)*, dan *Random Forests*; (2) *Unsupervised Classification* merupakan proses *clustering*, yaitu pembagian data asing dalam jumlah tertentu menjadi beberapa kelompok berdasarkan fitur yang serupa. Contoh algoritma: *K-Means Clustering*, *PCA (Principal Component Analysis)*, dan *Autoencoder*.

2.2.3 Support Vector Machine

SVM adalah suatu metode *machine learning* yang bekerja atas dasar prinsip *Structural Risk Minimization* dengan tujuan menemukan *hyperplane* terbaik yang

memisahkan dua buah kelas pada *input space* (Feldman & Sanger, 2006). Grafik pemisahan dua kelas data dengan *margin* maksimum dapat dilihat pada Gambar 2.2.



Gambar 2.2. Pemisahan Dua Kelas Data dengan *Margin* Maksimum

Pada Gambar 2.2. SVM menemukan *hyperlane* pemisah maksimum, yaitu *hyperlane* yang mempunyai jarak maksimum antara tupel pelatihan terdekat. *Support vector* ditunjukkan dengan batasan tebal pada titik tupel. Inti dari proses pembelajaran pada SVM adalah mendapatkan lokasi *hyperplane* (Nugroho *et al.*, 2003).

Beberapa penelitian menunjukkan bahwa SVM mampu memberikan performa lebih tinggi dalam tingkat akurasi pada proses klasifikasi data dibandingkan dengan algoritma klasifikasi lainnya (Srivastava & Bhambhu, 2009). Pada penelitian Dubey & Jain (2019) SVM diterapkan untuk klasifikasi ulasan produk ponsel yang menganalisis kumpulan data yang digunakan untuk klasifikasi sentimen dan teks dan menghasilkan akurasi tinggi yaitu 94,63%. Pada penelitian Tubishat *et al.* (2019) klasifikasi SVM menjadi solusi terbaik untuk mengurangi

ruang pencarian yang dieksplorasi oleh *Whale Optimization Algorithm* dalam analisis sentimen bahasa arab. Pada penelitian Lu & Wu (2019) SVM dapat meningkatkan tingkat analisis sentimen yang tepat dari ulasan film, dapat lebih baik diterapkan pada penilaian dan rekomendasi preferensi film, dan memiliki peran penuntun dalam promosi film platform media.

SVM sangat umum diterapkan dalam *machine learning* yang memungkinkan analisis data dan identifikasi pola tertentu (Aljawarneh *et al.*, 2017). Namun SVM juga memiliki kelemahan diantaranya adalah pemilihan kernel yang tepat sesuai dengan studi kasus. Setiap proses pada SVM membutuhkan parameter sesuai dengan kernelnya untuk mendapatkan akurasi yang terbaik. SVM memiliki beberapa kernel, seperti: linear, polinomial, dan fungsi dasar radial (RBF).

Pada penelitian Lidya *et al.* (2015) SVM Kernel Linear memiliki akurasi yang terbaik untuk menganalisis teks, karena kernel ini adalah yang paling sederhana dari semua fungsi kernel. Kernel Linear ditulis pada persamaan 1.

$$K(X_i, X_j) = X_i X_j \quad (1)$$

Parameter kernel linear SVM yaitu parameter C yang merupakan parameter penalti atau regularisasi (Guyon *et al.*, 1993). Parameter C memiliki pengaruh besar pada kinerja klasifikasi SVM karena menentukan *trade-off* antara minimalisasi kesalahan dan memaksimalkan *margin* klasifikasi (Alex J. Smola, 2004). Secara praktis, mengubah nilai C mengontrol kesalahan pelatihan, kesalahan pengujian, jumlah *support vector*, dan *margin* SVM (Tharwat, 2019).

2.2.4 *Grid Search*

Grid Search adalah pencarian lengkap berdasarkan subset ruang *hyperparameter* yang ditentukan menggunakan nilai minimal (*lower bound*/batas bawah), nilai maksimal (*upper bound*/batas atas), dan jumlah angka (Syarif *et al.*, 2016). *Grid Search* membagi jangkauan parameter yang akan dioptimalkan ke dalam *grid* dan melintasi semua titik untuk mendapatkan parameter yang optimal. *Grid Search* mengoptimalkan parameter SVM menggunakan teknik *cross validation* sebagai metrik kinerja. Tujuannya adalah untuk mengidentifikasi kombinasi *hyperparameter* yang baik sehingga *classifier* dapat memprediksi data yang tidak diketahui secara akurat. Menurut Lin *et al.*, (2008). Teknik *cross validation* dapat mencegah masalah *overfitting*. Berikut adalah algoritma *Grid Search* untuk parameter C pada SVM.

```

ALGORITMA: Grid Search untuk parameter C pada SVM
Initialize list of C candidates
FOR every c in list of C candidates
    Train SVM with c on TrainingSet
    Evaluate SVM classification on ValidationSet
    IF accuracy > MaxAccuracy
        THEN save MaxC = c
    ENDIF
ENDFOR
RETURN MaxC

```

Pada algoritma tersebut dijelaskan bahwa langkah *Grid Search* dimulai dari menginisialisasi kandidat C yang akan dicari parameter optimalnya, kemudian dilakukan pelatihan data SVM dengan *cross validation* sehingga menemukan parameter C yang optimal.

Pada penelitian Syarif *et al.* (2016) *Grid Sreach* sebagai optimasi parameter terbukti dapat meningkatkan akurasi pada SVM dibanding dengan *Genetic Algorithm*. Pada penelitian Deshwal & Sharma (2019) diterapkannya *Grid Sreach* pada *dataset* kanker payudara dengan model SVM memberikan hasil yang jauh lebih baik dari pada tidak diterapkan *Grid Sreach*. Pada penelitian Eliana *et al.* (2019) *Finding Anomalies Around the Mean* (FAM) yang dikombinasikan dengan *Grid Search* mampu memberikan kinerja yang baik, menunjukkan ketelitian yang tinggi, dan mampu memberikan hasil akurasi yang besar yaitu sebesar 92,63%.

2.2.5 Unigram

N-Gram didefinisikan sebagai irisan karakter bersebelahan dari *string* ke dalam *substring* yang lebih kecil, masing-masing berukuran n (Agyemang *et al.*, 2005). *N-Gram* juga merupakan salah satu metrik kesamaan berbasis *string*, tetapi menggunakan strategi kesamaan berdasarkan karakter (Diana & Hanana Ulfa, 2019). *N-Gram* dari *string* dengan panjang k diperoleh dengan menggeser jendela berukuran n di atas *string*, mulai dari posisi pertama dan menggerakkan jendela satu posisi pada satu waktu hingga mencapai ujung *string*. Himpunan karakter yang muncul di jendela pada posisi apa pun membentuk *N-Gram string* tersebut.

Pada penelitian Khreisat (2009) *N-Gram* bekerja dengan baik dalam kasus teks berisik (teks yang mengandung kesalahan ketik). Pada penelitian Diana & Hanana Ulfa (2019) Metode *N-Gram* memiliki waktu yang efektif dalam memproses data dibandingkan dengan *Jaccard*.

2.2.6 *Term Frequency Inverse Document Frequency (TF-IDF)*

Metode TF-IDF merupakan metode untuk menghitung bobot setiap kata yang paling umum digunakan pada *information retrieval*. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Metode TF-IDF menggunakan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan *inverse* frekuensi dokumen yang mengandung kata tersebut. TF-IDF dapat ditulis dengan persamaan 2:

$$tfidf_t = f_{t,d} \times \log \frac{N}{df_t} \quad (2)$$

Keterangan:

$tfidf_t$: bobot dari *term t*

$f_{t,d}$: frekuensi munculnya *term t* pada dokumen d

N : jumlah kumpulan dokumen

df_t : jumlah dokumen yang mengandung *term t*

Pada penelitian Kadhim (2019) TF-IDF terbukti efisien, mudah, memiliki hasil yang akurat, dan dapat meningkatkan evaluasi kinerja ekstraksi fitur sesuai dengan nilai maksimum adalah 89,77. Pada penelitian Hossein *et al.* (2016) TF-IDF yang diterapkan pada dua data dengan klasifikasi SVM memiliki hasil lebih tinggi yaitu 97,84% dan 94,93% dibandingkan dengan metode klasifikasi lainnya.

2.2.7 Validasi dan Evaluasi

Validasi dan evaluasi dilakukan untuk menentukan hasil akurasi percobaan yang telah dilakukan sebelumnya. Pada tahap ini *dataset* dibagi menjadi *data training* dan *data testing*. *Data training* diproses dengan algoritma klasifikasi untuk menghasilkan model klasifikasi. Kemudian model klasifikasi tersebut diuji dengan *data testing*. Hasil dari proses tersebut digunakan untuk mengukur kinerja algoritma dari metode yang diusulkan. Evaluasi model klasifikasi didasarkan pada pengujian untuk memprediksi objek yang benar dan salah, pengukuran kinerja klasifikasi dilakukan dengan *confusion matrix*.

Confusion matrix merupakan tabel pencatat hasil kerja klasifikasi, seperti yang ditunjukkan pada tabel 2.1.

Tabel 2.1. *Confusion Matrix* untuk Klasifikasi Dua Kelas

f_{ij}		Kelas hasil prediksi (j)	
		Kelas =1	Kelas =0
Kelas asli (i)	Kelas = 1	f_{11}	f_{10}
	Kelas = 0	f_{01}	f_{00}

Tabel 2.1. merupakan contoh *confusion matrix* yang melakukan klasifikasi dalam dua kelas, di mana nilai dua kelas tersebut yaitu 0 dan 1. Jika f_{ij} menotasikan jumlah *record* dari kelas i yang hasil prediksinya masuk ke kelas j pada saat pengujian. Misalnya, sel f_{11} adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1, dan f_{10} adalah data dalam kelas 1 yang dipetakan secara salah ke kelas 0.

Berdasarkan isi *confussion matrix*, dapat diketahui jumlah data dari masing-masing kelas yang diprediksi secara benar, yaitu $(f_{11}+f_{00})$, dan data yang diklasifikasikan secara salah, yaitu $((f_{10} + f_{01})$. Kuantitas *confussion matrix* dapat diringkas menjadi dua nilai, yaitu akurasi dan laju eror. Dengan mengetahui jumlah data yang diklasifikasikan secara benar, dapat diketahui akurasi hasil prediksi, dan dengan mengetahui jumlah data yang diklasifikasikan secara salah dapat mengetahui laju eror dari prediksi yang dilakukan. Oleh karena itu, pengukuran akurasi (*accuracy*) dapat dituliskan dengan persamaan 3:

$$Accuracy = \frac{\text{Jumlah prediksi yang benar}}{\text{Jumlah prediksi keseluruhan}} = \frac{f_{11}+f_{00}}{f_{11}+f_{10}+f_{01}+f_{00}} \quad (3)$$

BAB 3

METODE PENELITIAN

Metode penelitian yang digunakan sebagai acuan dalam melakukan penelitian adalah sebagai berikut.

3.1. Studi Literatur

Studi literatur dilakukan untuk mendapatkan gambaran secara lengkap mengenai permasalahan yang ada dalam penerapan *Grid Search* dan *Unigram* pada algoritma SVM. Pada tahap ini, dilakukan pengumpulan data, informasi, serta rujukan dari buku, artikel, jurnal, dan karya ilmiah lainnya yang berhubungan dengan objek penelitian. Bahan referensi yang digunakan mencakup penemuan atau modifikasi terhadap SVM sebagai algoritma klasifikasi dan implementasi *Grid Search* dan *Unigram*.

3.2. Perancangan Sistem

Pada penelitian ini, perancangan sistem dilakukan sebagai alat uji penerapan *Grid Search* dan *Unigram* pada algoritma SVM. Sistem dirancang menggunakan bahasa pemrograman *Python 3.8*. Bahasa pemrograman *Python* merupakan salah satu bahasa pemrograman yang paling populer untuk komputasi ilmiah. *Python* memiliki banyak *library* untuk kegiatan ilmiah, sehingga mendukung untuk pengembangan algoritma dan analisis data (Pedregosa *et al.*, 2011). *Library* yang

digunakan dalam penelitian ini yaitu *Django* 3.0.3, *nltk* 3.4.5, *numpy* 1.17.4, *openpyxl* 3.0.3, *pandas* 0.25.3, *sastrawi* 1.0.1, *scikit-learn* 0.22, dan *xlrd* 1.2.0. Aplikasi dibangun berbasis web dengan menggunakan *Django* yang merupakan web *framework Python* yang dikembangkan oleh *Django Software Foundation*.

3.3. Pengambilan Data

Dalam tahap ini, dilakukan pengambilan data sebagai objek dalam eksperimen penelitian. Data yang diambil berjumlah 1500 *review Amazon* dan 1500 *review Lazada*. Data tersebut merupakan *public dataset* yang tersedia di web *Kaggle.com* yang terdiri dari komentar positif dan negatif.

Data yang digunakan pada penelitian ini adalah *sentiment labelled dataset* yaitu *review Amazon* (berbahasa Inggris) dan *review Lazada* (berbahasa Indonesia). Pada data *review Amazon* pelabelan positif ditandai dengan angka 1 dan pelabelan negatif ditandai dengan angka 0. Sedangkan pada data *review Lazada* pelabelan dilakukan berdasarkan rating. Untuk pelabelan positif ditujukan untuk rating 1 sampai dengan 3 dan pelabelan negatif ditujukan untuk rating 4 sampai dengan 5.

3.4. Pengolahan Data

Tahap pengolahan data dimulai dengan mengonversikan ekstensi dari *dataset* ke ekstensi *excel*. *Dataset review Amazon* yang diperoleh dari web *Kaggle.com* memiliki format *.ft* (*full text index*). *Dataset review Amazon* dalam format *.ft* ditunjukkan pada Gambar 3.1.

__label__2 Stuning even for the non-gamer: This sound track was beautiful! It paints the senery in your mind so well I would recomend it even to people who hate vid. game music! I have played the game Chrono Cross but out of all of the games I have ever played it has the best music! It backs away from crude keyboarding and takes a fresher step with grate guitars and soulful orchestras. It would impress anyone who cares to listen! ^_^

__label__2 The best soundtrack ever to anything.: I'm reading a lot of reviews saying that this is the best 'game soundtrack' and I figured that I'd write a review to disagree a bit. This in my opinino is Yasunori Mitsuda's ultimate masterpiece. The music is timeless and I'm been listening to it for years now and its beauty simply refuses to fade.The price tag on this is pretty staggering I must say, but if you are going to buy any cd for this much money, this is the only one that I feel would be worth every penny.

__label__2 Amazing!: This soundtrack is my favorite music of all time, hands down. The intense sadness of "Prisoners of Fate" (which means all the more if you've played the game) and the hope in "A Distant Promise" and "Girl who Stole the Star" have been an important inspiration to me personally throughout my teen years. The higher energy tracks like "Chrono Cross ~ Time's Scar~", "Time of the Dreamwatch", and "Chronomantique" (indefinably remeniscent of Chrono Trigger) are all absolutely superb as well.This soundtrack is amazing music, probably the best of this composer's work (I haven't heard the Xenogears soundtrack, so I can't say for sure), and even if you've never played the game, it would be worth twice the price to buy it.I wish I could give it 6 stars.

__label__2 Excellent Soundtrack: I truly like this soundtrack and I enjoy video game music. I have played this game and most of the music on here I enjoy and it's truly relaxing and peaceful.On disk one. my favorites are Scars Of Time, Between Life and Death, Forest Of Illusion, Fortress of Ancient Dragons, Lost Fragment, and Drowned Valley.Disk Two: The Draggons, Galdorb - Home, Chronomantique, Prisoners of Fate, Gale, and my girlfriend likes ZebessDisk Three: The best of the three. Garden Of God, Chronopolis, Fates, Jellyfish sea, Burning Orphange, Dragon's Prayer, Tower Of Stars, Dragon God, and Radical Dreamers - Unstealable Jewel.Overall, this is a excellent soundtrack and should be brought by those that like video game music.Xander Cross

__label__2 Remember, Pull Your Jaw Off The Floor After Hearing it: If you've played the game, you know how divine the music is! Every single song tells a story of the game, it's that good! The greatest songs are without a doubt, Chrono Cross: Time's Scar, Magical Dreamers: The Wind, The Stars, and the Sea and Radical Dreamers: Unstolen Jewel. (Translation varies) This music is perfect if you ask me, the best it can be. Yasunori Mitsuda just poured his heart on and wrote it down on paper.

Gambar 3.1. Tampilan *Dataset* dengan Format .ft

Sedangkan *dataset review* Lazada yang diperoleh dari web *Kaggle.com* memiliki format *.csv (comma separated values)*. *Dataset* tersebut memiliki banyak atribut. Dalam penelitian ini, atribut yang digunakan adalah *review content* dan *rating*. *Rating* digunakan untuk pelabelan dengan dua kategori yaitu positif dan negatif. Positif digunakan untuk rating 3 sampai dengan 5, dan negatif untuk rating 1 sampai dengan 2. *Review* Lazada dalam format *.csv* yang sudah diproses ditunjukkan pada Gambar 3.2.

Review Content	Label
bagus mantap dah sesuai pesanan	Positive
Bagus, sesuai foto	Positive
okkkkk mantaaaaaaapppp ... goood	Positive
bagus sesuai	Positive
Barang udah sampai tapi pas dinyalain g bisa. Gmn pertanggung jawaban lazada?	Negative
Barang sampai dengan aman. Kemampuan laptop sesuai dengan yang di iklankan.	Positive
Hanya saja, laptop datang dengan DOS sehingga harus di install sendiri windows nya.	
Overall, it's worth the money.	
Pengirim barang tidak sesuai janji. Katanya express 1 hari. Dari hari jumat barang sampai hari rabu. SeMoga barangnya awet saja.	Negative
Kualitas ok	Positive
Bagaimana tv LED saya merek Sharp 24 hitam LC 24LE170i sekarang mati total baru 2 bln kemana saya harus menghubungi dan meminta bantuan mohon kebijakan dan solusinya ,, so,alnya masih garansi kepada pihak lazada saya mohon bantuannya terimakasih!!!	Negative

Gambar 3.2. Tampilan *Dataset* dengan Format .csv

Karena *dataset review Amazon* masih berbentuk .ft maka diperlukan konversi ekstensi menjadi .csv. *Dataset* hasil konversi ditunjukkan pada Tabel 3.1.

Tabel 3.1. *Dataset* dalam Bentuk *Excel*

Review	Label
Stuning even for the non-gamer: This sound track was beautiful! It paints the senery in your mind so well I would recomend it even to people who hate vid. game music! I have played the game Chrono Cross but out of all of the games I have ever played it has the best music! It backs away from crude keyboarding and takes a fresher step with grate guitars and soulful orchestras. It would impress anyone who cares to listen! ^ ^	positive
The best soundtrack ever to anything.: I'm reading a lot of reviews saying that this is the best 'game soundtrack' and I figured that I'd write a review to disagree a bit. This in my opinino is Yasunori Mitsuda's ultimate masterpiece. The music is timeless and I'm been listening to it for years now and its beauty simply refuses to fade.The price tag on this is pretty staggering I must say, but if you are going to buy any cd for this much money, this is the only one that I feel would be worth every penny.	positive
Oh please: I guess you have to be a romance novel lover for this one, and not a very discerning one. All others beware! It is absolute drivel. I figured I was in trouble when a typo is prominently featured on the back cover, but the first page of the book removed all doubt. Wait - maybe I'm missing the point. A quick re-read of the beginning now makes it clear. This has to be an intentional churning of over-heated prose for satiric purposes. Phew, so glad I didn't waste \$10.95 after all.	negative
Awful beyond belief!: I feel I have to write to keep others from wasting their money. This book seems to have been written by a 7th grader with poor grammatical skills for her age! As another reviewer points out, there is a misspelling on the cover, and I believe there is at least one per chapter. For example, it was mentioned twice that she had a "lean" on her house. I was so distracted by the poor writing and weak plot, that	negative

I decided to read with a pencil in hand to mark all of the horrible grammar and spelling. Please don't waste your money. I too, believe that the good reviews must have been written by the author's relatives. I will not put much faith in the reviews from now on!

3.5. Preprocessing Data

Pada *text mining*, sebelum melakukan proses klasifikasi, data perlu dipersiapkan terlebih dahulu untuk dapat diolah. Proses tersebut dikenal dengan *preprocessing*. Proses ini akan sangat berpengaruh terhadap performa dari algoritma *text mining* karena proses ini akan membersihkan data untuk proses *mining*. Pada penelitian ini terdiri dari 5 tahap *preprocessing*, yaitu sebagai berikut.

3.5.1 Transform Cases

Transform Cases dilakukan untuk mengubah semua huruf menjadi kecil (*lower cases*). Hal ini dilakukan untuk meminimalkan variasi penulisan data teks. Adapun tahapan proses *Transform Cases* yaitu sebagai berikut.

1. Input data yang digunakan berupa *review Amazon* dan *review Lazada*,
2. Data *review Amazon* dan *review Lazada* apabila terdapat karakter yang menggunakan huruf kapital (*uppercase*), maka karakter tersebut akan diubah menjadi huruf kecil (*lowercase*),
3. Data *review Amazon* dan *review Lazada* menjadi *lowercase* yang kemudian digunakan pada proses *Punctuation Removal*,
4. Proses selesai.

3.5.2 Punctuation Removal

Punctuation Removal dilakukan untuk menghilangkan seluruh tanda baca yang ada di dalam teks, seperti titik, koma, tanda tanya, tanda seru, dan lain-lain. Hal ini dilakukan untuk membuang kata-karakter yang tidak penting dan meminimalkan variasi penulisan. Adapun tahapan proses *Punctuation Removal* yaitu sebagai berikut.

1. Kata dari hasil *Transform Cases* akan dilakukan pengecekan tanda baca,
2. Apabila terdapat tanda baca, maka tanda baca tersebut akan dihapus,
3. Proses selesai.

3.5.3 Tokenize

Tokenize yaitu proses pembagian teks pada setiap dokumen menjadi sebuah rangkaian dari token/*term*. Hasilnya adalah masing-masing kata dalam teks direpresentasikan sebagai sebuah token/*term*. Adapun tahapan proses *Tokenize* yaitu sebagai berikut.

1. Kata dari hasil *Punctuation Removal* yang masih bentuk paragraf akan dipisahkan menjadi satu kata,
2. Setiap satu kata direpresentasikan sebagai sebuah token/*term*,
3. Proses selesai.

3.5.4 Stopword Removal

Stopword Removal bekerja dengan memfilter seluruh token kemudian menghilangkannya apabila token tersebut terdapat pada *stopword list*. Penelitian ini

menggunakan *english stopword list* dan *indonesian stopword list* sebagai *stopword filtering*-nya dengan hasil token dari tahap sebelumnya, kemudian dilakukan *stopword filtering* yang menghasilkan token. Adapun tahapan proses *Stopword Removal* yaitu sebagai berikut.

1. Kata dari hasil *Tokenize* akan dibandingkan dengan kata yang ada pada *stopword list*,
2. Kemudian lakukan pengecekan apakah kata sama dengan *stopword list* atau tidak,
3. Apabila kata sama dengan *stopword list*, maka kata itu akan di hapus,
4. Proses selesai.

3.5.5 Stemming

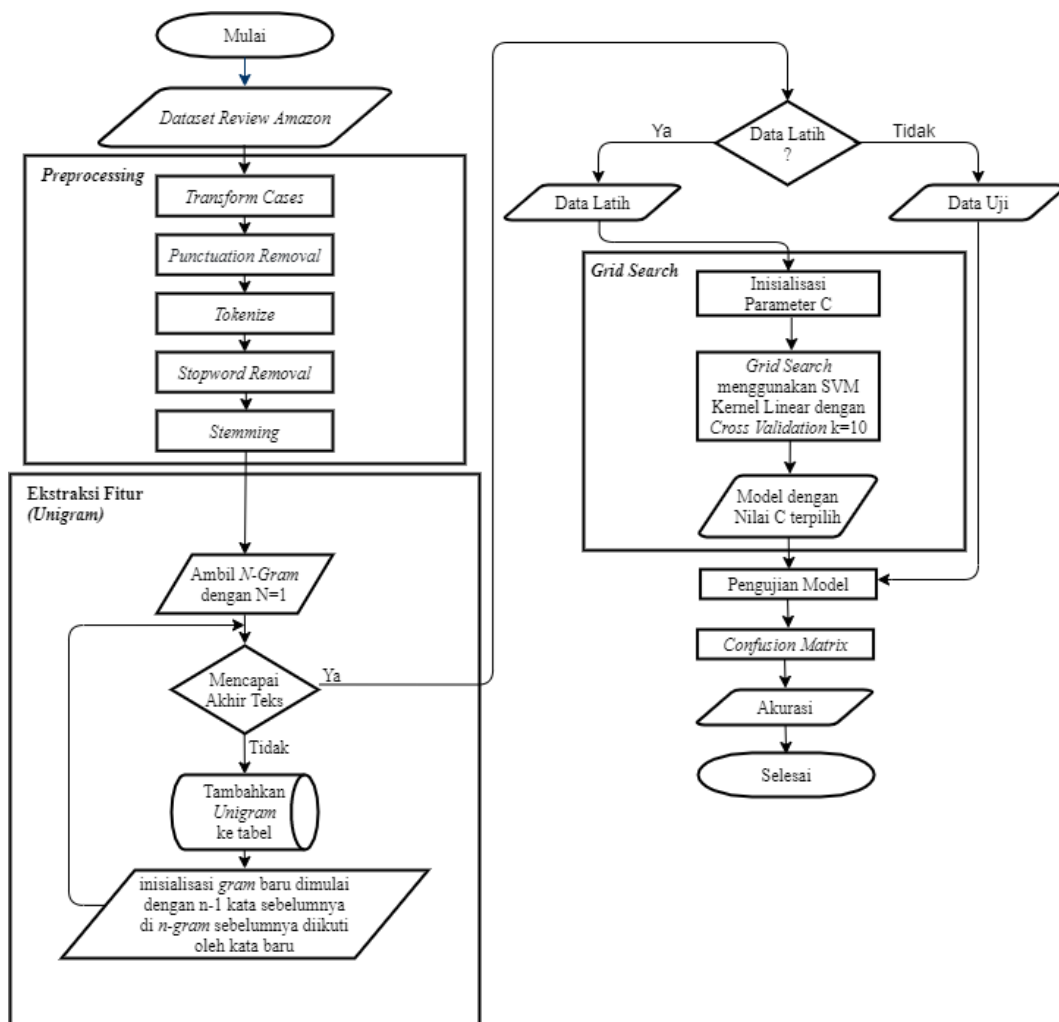
Stemming adalah proses pengubahan kata ke dalam bentuk dasarnya. Pada tahap ini setiap kata diidentifikasi untuk diubah ke dalam bentuk dasarnya. Apabila terdapat kata yang sama teridentifikasi lebih dari satu maka kata tersebut diidentifikasi sebagai satu kata. Adapun tahapan proses *Stemming* yaitu sebagai berikut.

1. Kata dari hasil *stopword removal* dilakukan pengecekan, apakah kata dari hasil *stopword removal* merupakan kata dasar atau tidak,
2. Apabila kata dasar maka proses telah berhenti atau selesai tetapi apabila bukan merupakan kata dasar maka hapus *sufiks* (imbuan yang terletak di akhir kata),

3. Kata hasil penghapusan *sufiks* apabila merupakan kata dasar maka proses selesai tetapi apabila bukan kata dasar pencarian kata tidak ditemukan berarti kata awal diasumsikan sebagai *root word* (kata dasar),
4. Proses selesai.

3.6. Metode yang Digunakan

Pada penelitian ini dirancang dengan 3 skenario. Pertama, melakukan klasifikasi menggunakan algoritma SVM pada *dataset review Amazon* dan *review Lazada*. Kedua, melakukan klasifikasi menggunakan algoritma SVM dengan menerapkan *Unigram* sebagai *feature extraction* pada *dataset review Amazon* dan *review Lazada*. Ketiga, melakukan klasifikasi menggunakan algoritma SVM dengan menerapkan *Grid Search* sebagai optimasi parameter pada *dataset review Amazon* dan *review Lazada*. Adapun *Flowchart* dari metode yang digunakan dalam penelitian ini ditunjukkan pada Gambar 3.3.



Gambar 3.3. Flowchart Algoritma SVM dengan Grid Search dan Unigram

Penelitian ini dimulai dengan mempersiapkan *dataset*. *Dataset* yang digunakan dalam penelitian ini adalah *dataset review Amazon* dan *review Lazada* yang diambil dari web *Kaggle.com*. *Dataset* yang diperoleh kemudian dilakukan tahap *preprocessing*. *Preprocessing* yang digunakan pada penelitian ini yaitu *transform cases* untuk mengubah huruf menjadi kecil secara keseluruhan, *punctuation removal* untuk menghilangkan tanda baca, *tokenize* untuk memisahkan

setiap kata pada dokumen, *stopword removal* untuk menghilangkan kata yang tidak berguna, *stemming* untuk mengubah kata imbuhan sehingga menjadi kata dasar.

Setelah dilakukan *preprocessing*, langkah selanjutnya yaitu melakukan ekstraksi fitur dengan *Unigram* (dengan $n=1$) untuk mengubah *string* atau kalimat menjadi satu bagian kata. Setelah didapatkan *dataset* dengan fitur yang terpilih pada tahap *feature extraction*, kemudian *dataset* tersebut dibagi menjadi data latih dan data uji menggunakan metode *splitter*. Data latih terdiri dari 75% dari keseluruhan *dataset* dan data uji terdiri dari 25% dari keseluruhan *dataset*.

Data latih digunakan untuk proses *learning* pada algoritma SVM sehingga terbentuk model klasifikasi SVM. Klasifikasi SVM dilakukan dengan *Grid Search* menggunakan kernel linear dengan *cross validation* $k=10$ sehingga terpilih nilai parameter C yang optimal. Selanjutnya adalah melakukan pengujian model klasifikasi SVM menggunakan data uji. Berdasarkan pengujian tersebut, analisis hasil dapat dilakukan dengan menghitung akurasi yang didapatkan menggunakan *confusion matrix* sehingga didapatkan akurasi dari algoritma SVM dan akurasi dari metode yang digunakan dalam penelitian ini. Akurasi dihitung berdasarkan persamaan 3 pada Bab 2. Dengan membandingkan hasil akurasi dari metode algoritma SVM tanpa proses *Grid Search* dan *Unigram* dengan algoritma SVM yang menerapkan *Grid Search* dan *Unigram* maka akan terlihat peningkatan akurasi dari kedua metode tersebut.

3.7. Penarikan Kesimpulan

Penarikan kesimpulan didasarkan pada studi pustaka, pengumpulan data dan pengembangan sistem serta hasil analisis dari penelitian. Simpulan dari penelitian ini adalah bagaimana tingkat akurasi yang dihasilkan algoritma SVM dalam menerapkan *Grid Search* dan *Unigram* pada *review Amazon* dan *review Lazada*.

BAB 4

HASIL DAN PEMBAHASAN

Pada bab empat ini berisikan hasil dan pembahasan dari langkah-langkah yang sudah ditulis pada bab, yaitu sebagai berikut.

4.1 Hasil Penelitian

Pada hasil penelitian akan dipaparkan hasil dari *preprocessing*, hasil *text mining*, hasil perbandingan dengan penelitian sebelumnya, dan hasil implementasi sistem.

4.1.1 Hasil *Preprocessing*

Berikut ini akan ditampilkan hasil dari 5 tahap *preprocessing*, yaitu sebagai berikut.

4.1.1.1 Hasil *Transform Cases*

Hasil dari *Transform Cases* ditunjukkan pada Tabel 4.1.

Tabel 4.1. Hasil *Transform Cases*

Teks	Token
Stuning even for the non-gamer: This sound track was beautiful! It paints the senery in your mind so well I would recomed it even to people who hate vid. game music! I have played the game Chrono Cross but out of all of the games I have ever played it has the best music! It backs away from crude keyboarding and takes a fresher step with grate guitars and soulful orchestras. It would impress anyone who cares to listen! ^_^	stuning even for the non-gamer: this sound track was beautiful! it paints the senery in your mind so well i would recomed it even to people who hate vid. game music! i have played the game chrono cross but out of all of the games i have ever played it has the best music! it backs away from crude keyboarding and takes a fresher step with grate guitars and soulful orchestras. it would impress anyone who cares to listen! ^_^

Pada Tabel 4.1. kolom sebelah kiri berisikan teks yang akan diproses dengan *Transform Cases*, sedangkan kolom sebelah kanan merupakan gambaran hasil dari proses *Transform Cases*. Terlihat jelas dikolom sebelah kanan bahwa semua huruf sudah berubah menjadi kecil tanpa ada huruf kapital.

4.1.1.2 Hasil *Punctuation Removal*

Hasil dari *Punctuation Removal* ditunjukkan pada Tabel 4.2.

Tabel 4.2. Hasil *Punctuation Removal*

Teks	Token
Stuning even for the non-gamer: This sound track was beautiful! It paints the senery in your mind so well I would recomend it even to people who hate vid. game music! I have played the game Chrono Cross but out of all of the games I have ever played it has the best music! It backs away from crude keyboarding and takes a fresher step with grate guitars and soulful orchestras. It would impress anyone who cares to listen! ^_^	stuning even for the nongamer this sound track was beautiful it paints the senery in your mind so well i would recomend it even to people who hate vid game music i have played the game chrono cross but out of all of the games i have ever played it has the best music it backs away from crude keyboarding and takes a fresher step with grate guitars and soulful orchestras it would impress anyone who cares to listen

Pada Tabel 4.2. kolom sebelah kiri berisikan teks yang akan diproses dengan *Punctuation Removal*, sedangkan kolom sebelah kanan merupakan gambaran hasil dari proses *Punctuation Removal*. Terlihat jelas dikolom sebelah kanan bahwa teks sudah bersih dari tanda baca.

4.1.1.3 Hasil *Tokenize*

Hasil dari *Tokenize* adalah masing-masing kata dalam teks direpresentasikan sebagai sebuah token/*term* sebagaimana ditunjukkan Tabel 4.3.

Tabel 4.3. Hasil *Tokenize*

Teks	Token
stuning even for the nongamer this sound track was beautiful it paints the senery in your mind so well i would recomend it even to people who hate vid game music i have played the game chrono cross but out of all of the games i have ever played it has the best music it backs away from crude keyboarding and takes a fresher step with grate guitars and soulful orchestras it would impress anyone who cares to listen	“stuning”; “even”; “for”; “the”; “nongamer”; “this”; “sound”; “track”; “was”; “beautiful”; “it”; “paints”; “the”; “senery”; “in”; “your”; “mind”; “so”; “well”; “I”; “would”; “recommend”; “it”; “even”; “to”; “people”; “who”; “hate”; “vid”; “game”; “music”; “I”; “have”; “played”; “the”; “game”; “chrono”; “cross”; “but”; “out”; “of”; “all”; “of”; “the”; “games”; “I”; “have”; “ever”; “played”; “it”; “has”; “the”; “best”; “music”; “it”; “backs”; “away”; “from”; “crude”; “keyboarding”; “and”; “takes”; “a”; “fresher”; “step”; “with”; “grate”; “guitars”; “and”; “soulful”; “orchestras”; “it”; “would”; “impress”; “anyone”; “who”; “cares”; “to”; “listen”

Pada Tabel 4.3. kolom sebelah kiri berisikan teks yang akan diproses dengan *Tokenize*, sedangkan kolom sebelah kanan merupakan gambaran hasil dari proses *Tokenize*. Terlihat jelas dikolom sebelah kanan bahwa teks sudah terpisah menjadi satuan kata.

4.1.1.4 Hasil *Stopword Removal*

Hasil dari *Stopword Removal* ditunjukkan Tabel 4.4.

Tabel 4.4. Hasil *Stopword Removal*

Teks	Token
“stuning”; “even”; “for”; “the”; “nongamer”; “this”; “sound”; “track”; “was”; “beautiful”; “it”; “paints”; “the”; “senery”; “in”; “your”; “mind”; “so”; “well”; “I”; “would”; “recommend”; “it”; “even”; “to”; “people”; “who”; “hate”; “vid”; “game”; “music”; “I”; “have”; “played”; “the”; “game”; “chrono”; “cross”; “but”; “out”; “of”; “all”; “of”; “the”; “games”; “I”; “have”; “ever”; “played”; “it”; “has”; “the”; “best”; “music”; “it”; “backs”; “away”; “from”; “crude”; “keyboarding”;	“stuning”; “even”; “nongamer”; “sound”; “track”; “beautiful”; “paints”; “senery”; “mind”; “well”; “would”; “recommend”; “even”; “people”; “hate”; “vid”; “game”; “music”; “played”; “game”; “chrono”; “cross”; “games”; “ever”; “played”; “best”; “music”; “backs”; “away”; “crude”; “keyboarding”; “takes”; “fresher”; “step”; “grate”; “guitars”; “soulful”; “orchestras”; “would”; “impress”; “anyone”; “cares”; “listen”

“and”; “takes”; “a”; “fresher”; “step”; “with”;
 “grate”; “guitars”; “and”; “soulful”;
 “orchestras”; “it”; “would”; “impress”;
 “anyone”; “who”; “cares”; “to”; “listen”

Pada Tabel 4.4. kolom sebelah kiri berisikan teks yang akan diproses dengan *Stopword Removal*, sedangkan kolom sebelah kanan merupakan gambaran hasil dari proses *Stopword Removal*. Terlihat jelas dikolom sebelah kanan bahwa teks sudah bersih dari kata imbuhan atau kata hubung.

4.1.1.5 Hasil *Stemming*

Hasil token setelah dilakukan *stemming* ditunjukkan pada Tabel 4.5.

Tabel 4.5. Hasil *Stemming*

Teks	Token
“stuning”; “even”; “nongamer”; “sound”;	“stune”; “even”; “nongam”; “sound”; “track”;
“track”; “beautiful”; “paints”; “senery”;	“beauti”; “paint”; “seneri”; “mind”; “well”;
“mind”; “well”; “would”; “recommend”;	“would”; “recomend”; “even”; “peopl”; “hate”;
“even”; “people”; “hate”; “vid”; “game”;	“vid”; “game”; “music”; “play”; “game”;
“music”; “played”; “game”; “chrono”; “cross”;	“chrono”; “cross”; “game”; “ever”; “play”;
“games”; “ever”; “played”; “best”; “music”;	“best”; “music”; “back”; “away”; “crude”;
“backs”; “away”; “crude”; “keyboarding”;	“keyboard”; “take”; “fresher”; “step”; “grate”;
“takes”; “fresher”; “step”; “grate”; “guitars”;	“guitar”; “soul”; “orchestra”; “would”;
“soulful”; “orchestras”; “would”; “impress”;	“impress”; “anyon”; “care”; “listen”
“anyone”; “cares”; “listen”	

Pada Tabel 4.5. kolom sebelah kiri berisikan teks yang akan diproses dengan *Stemming*, sedangkan kolom sebelah kanan merupakan gambaran hasil dari proses *Stemming*. Terlihat jelas dikolom sebelah kanan bahwa masing-masing kata sudah diubah kedalam bentuk dasar kata tersebut.

4.1.2 Hasil *Text Mining*

Hasil *text mining* akan dijelaskan kedalam 3 skenario yaitu hasil dari klasifikasi algoritma SVM, hasil klasifikasi algoritma SVM dengan menerapkan *Unigram*, dan hasil klasifikasi SVM dengan menerapkan *Unigram* dan *Grid Sreach*.

4.1.2.1 Hasil Klasifikasi dengan Algoritma SVM

Penerapan klasifikasi yang pertama adalah dengan menerapkan algoritma SVM dengan perbandingan pembagian data sebesar 75:25 dan teknik evaluasi *confussion matrix* pada dataset *review Amazon* dan *review Lazada*. Hasil tersebut akan digunakan sebagai pembanding setelah menerapkan *Grid Search* dan *Unigram*. Dari hasil *testing* penerapan algoritma SVM yang dilakukan pada data *review Amazon* diperoleh hasil seperti pada *matrix* berikut ini.

f_{ij}		Kelas hasil prediksi (j)	
		Kelas =1	Kelas =0
Kelas asli (i)	Kelas = 1	$f_{11} = 15$	$f_{10} = 0$
	Kelas = 0	$f_{01} = 171$	$f_{00} = 189$

Kemudian dilakukan perhitungan akurasi sesuai Persamaan 3 pada bab 2.

$$Accuracy = \frac{15 + 189}{15 + 0 + 171 + 189} = 0,544$$

Menghasilkan akurasi 0,544 atau dalam satuan persen yaitu sebesar 54,40%.

Sedangkan dari hasil *testing* penerapan algoritma SVM yang dilakukan pada data *review Lazada* diperoleh hasil seperti pada *matrix* berikut ini.

f_{ij}		Kelas hasil prediksi (j)	
		Kelas =1	Kelas =0
Kelas asli (i)	Kelas = 1	$f_{11} = 322$	$f_{10} = 0$
	Kelas = 0	$f_{01} = 53$	$f_{00} = 0$

Kemudian dilakukan perhitungan akurasi sesuai Persamaan 3 pada bab 2.

$$Accuracy = \frac{322 + 0}{322 + 0 + 53 + 0} = 0,85866667$$

Menghasilkan akurasi 0,85866667 atau dalam satuan persen yaitu sebesar 85,87%.

4.1.2.2 Hasil Klasifikasi Algoritma SVM dengan menerapkan *Unigram*

Penerapan klasifikasi ini dilakukan dengan menerapkan *Unigram* sebagai *feature extraction* pada algoritma klasifikasi SVM untuk dataset *review Amazon* dan *review Lazada*. *Unigram* akan mengekstraksi fitur dengan membagi suatu kalimat menjadi satu kata pada setiap bagian.

Dari hasil *testing* penerapan SVM dan *Unigram* yang dilakukan pada data *review Amazon* diperoleh hasil seperti pada *matrix* berikut ini.

f_{ij}		Kelas hasil prediksi (j)	
		Kelas =1	Kelas =0
Kelas asli (i)	Kelas = 1	$f_{11} = 146$	$f_{10} = 32$
	Kelas = 0	$f_{01} = 40$	$f_{00} = 157$

Kemudian dilakukan perhitungan akurasi sesuai Persamaan 3 pada bab 2.

$$Accuracy = \frac{146 + 157}{146 + 32 + 40 + 157} = 0,808$$

Menghasilkan akurasi 0,808 atau dalam satuan persen yaitu sebesar 80,80%.

Sedangkan dari hasil *testing* penerapan SVM dan *Unigram* yang dilakukan pada data *review* Lazada diperoleh hasil seperti pada *matrix* berikut ini.

f_{ij}		Kelas hasil prediksi (j)	
		Kelas =1	Kelas =0
Kelas asli (i)	Kelas = 1	$f_{11} = 317$	$f_{10} = 40$
	Kelas = 0	$f_{01} = 5$	$f_{00} = 13$

Kemudian dilakukan perhitungan akurasi sesuai Persamaan 3 pada bab 2.

$$Accuracy = \frac{317 + 13}{317 + 40 + 5 + 13} = 0,88$$

Menghasilkan akurasi 0,88 atau dalam satuan persen yaitu sebesar 88,00%.

Hasil akurasi pada data *review Amazon* dari penerapan model ini dapat meningkatkan akurasi dari algoritma SVM sebesar 26,40% dan pada data *review* Lazada sebesar 4,26%. Namun, hasil tersebut masih dapat ditingkatkan dengan mengoptimalkan parameter SVM menggunakan *Grid Search*.

4.1.2.3 Hasil Klasifikasi Algoritma SVM dengan menerapkan *Unigram* dan *Grid Search*

Penerapan klasifikasi ini dilakukan dengan menerapkan kombinasi antara *Unigram* dan *Grid Search* pada algoritma klasifikasi SVM untuk dataset *review*

Amazon dan *review* Lazada. Setelah melakukan fitur ekstraksi dengan $n=1$ yang telah dipilih oleh *Unigram*, akan dilakukan pengoptimalan parameter dengan menggunakan *Grid Search*. Pada penerapan model ini, dilakukan pengujian pada data *train* dengan 10 *cross validation* pada 11 kandidat parameter C. Hasil rata-rata pengujian dari 11 kandidat parameter C untuk data *review Amazon* dapat dilihat pada Tabel 4.6.

Tabel 4.6. Hasil Pengujian SVM dengan *Unigram* dan *Grid Search*
pada Data *Review Amazon*

Pengujian ke	Nilai Parameter C	Hasil Akurasi
1	0,10	76,81
2	0,18	75,11
3	0,26	75,02
4	0,34	74,76
5	0,42	74,85
6	0,50	74,76
7	0,58	74,49
8	0,66	74,49
9	0,74	74,58
10	0,82	74,50
11	0,90	74,49
Parameter Optimal	0,1	76,81

Sedangkan hasil rata-rata pengujian dari 11 kandidat parameter C untuk data *review* Lazada dapat dilihat pada Tabel 4.7.

Tabel 4.7. Hasil Pengujian SVM dengan *Unigram* dan *Grid Search* pada Data *Review* Lazada

Pengujian ke	Nilai Parameter C	Hasil Akurasi
1	0,10	84,71
2	0,18	85,15
3	0,26	84,71
4	0,34	84,80
5	0,42	85,15
6	0,50	85,24
7	0,58	85,24
8	0,66	84,80
9	0,74	84,80
10	0,82	84,62
11	0,90	84,36
Parameter Optimal	0,58	85,24

Dari hasil *testing* penerapan SVM, *Unigram*, dan *Grid Sreach* yang dilakukan pada data *review Amazon* diperoleh hasil seperti pada *matrix* berikut ini.

f_{ij}		Kelas hasil prediksi (j)	
		Kelas =1	Kelas =0
Kelas asli (i)	Kelas = 1	$f_{11} = 146$	$f_{10} = 32$
	Kelas = 0	$f_{01} = 40$	$f_{00} = 157$

Kemudian dilakukan perhitungan akurasi sesuai Persamaan 3 pada bab 2.

$$Accuracy = \frac{146 + 157}{146 + 32 + 40 + 157} = 0,808$$

Menghasilkan akurasi 0,808 atau dalam satuan persen yaitu sebesar 80,80%.

Sedangkan dari hasil *testing* penerapan SVM, *Unigram*, dan *Grid Sreach* yang dilakukan pada data *review* Lazada diperoleh hasil seperti pada *matrix* berikut ini.

		Kelas hasil prediksi (j)	
		Kelas =1	Kelas =0
Kelas asli (i)	Kelas = 1	$f_{11} = 312$	$f_{10} = 27$
	Kelas = 0	$f_{01} = 10$	$f_{00} = 26$

Kemudian dilakukan perhitungan akurasi sesuai Persamaan 3 pada bab 2.

$$Accuracy = \frac{312 + 26}{312 + 27 + 10 + 26} = 0,90133333$$

Menghasilkan akurasi 0,90133333 atau dalam satuan persen yaitu sebesar 90,13%.

Peningkatan akurasi pada data *train* dan data *test* dari penerapan model *Unigram* dan *Grid Search* untuk *dataset review Amazon* ditunjukkan pada tabel 4.8.

Tabel 4.8. Peningkatan Akurasi SVM dengan *Unigram* dan *Grid Search* pada Data *Review Amazon*

	<i>Train</i>	<i>Test</i>
SVM	52,35%	54,40%
SVM+ <i>Unigram</i>	76,81%	80,80%
SVM+ <i>Unigram</i> + <i>Grid Search</i>	76,81%	80,80%

Pada Tabel 4.9. hasil akurasi SVM+*Unigram* dan SVM+*Unigram*+*Grid Search* tidak ada peningkatan akurasi dikarenakan parameter C optimal yang diperoleh adalah 0,1 yang merupakan parameter C *default*. Sedangkan peningkatan akurasi pada data *train* dan data *test* dari penerapan model *Unigram* dan *Grid Search* untuk data *review Lazada* dapat dilihat pada Tabel 4.9.

Tabel 4.9. Peningkatan Akurasi SVM dengan *Unigram* dan *Grid Search* pada Data *Review* Lazada

	<i>Train</i>	<i>Test</i>
SVM	84,45%	85,87%
SVM+ <i>Unigram</i>	84,71%	88,00%
SVM+ <i>Unigram</i> + <i>Grid Search</i>	85,24%	90,13%

Pada Tabel 4.9. hasil akurasi SVM dengan *Unigram* dan *Grid Search* pada data *review* Lazada menghasilkan parameter C optimal 0,58 yang berbeda dari parameter C *default*, sehingga terlihat jelas peningkatan akurasinya.

4.1.3. Hasil Perbandingan dengan Penelitian Sebelumnya

Pada penelitian sebelumnya yaitu pada jurnal Ravi dan Khettry (2019) menghitung akurasi *review E-Commerce* yaitu *Amazon* menggunakan *feature extraction bigram* serta *preprocessing* meliputi *tokenize*, *punctuation removal*, *stopword removal*, dan *stemming*. Untuk data *review Amazon* menghasilkan *testing* seperti *matrix* berikut.

f_{ij}		Kelas hasil prediksi (j)	
		Kelas =1	Kelas =0
Kelas asli (i)	Kelas = 1	$f_{11} = 164$	$f_{10} = 77$
	Kelas = 0	$f_{01} = 22$	$f_{00} = 112$

Kemudian dilakukan perhitungan akurasi sesuai Persamaan 3 pada bab 2.

$$Accuracy = \frac{164 + 112}{164 + 77 + 22 + 112} = 0,736$$

Menghasilkan akurasi 0,736 atau dalam satuan persen yaitu sebesar 73,60%.

Sedangkan pada data *review* Lazada menghasilkan *testing* seperti *matrix* berikut.

f_{ij}		Kelas hasil prediksi (j)	
		Kelas =1	Kelas =0
Kelas asli (i)	Kelas = 1	$f_{11} = 164$	$f_{10} = 77$
	Kelas = 0	$f_{01} = 22$	$f_{00} = 112$

Kemudian dilakukan perhitungan akurasi sesuai Persamaan 3 pada bab 2.

$$Accuracy = \frac{320 + 6}{320 + 6 + 47 + 2} = 0,86933333$$

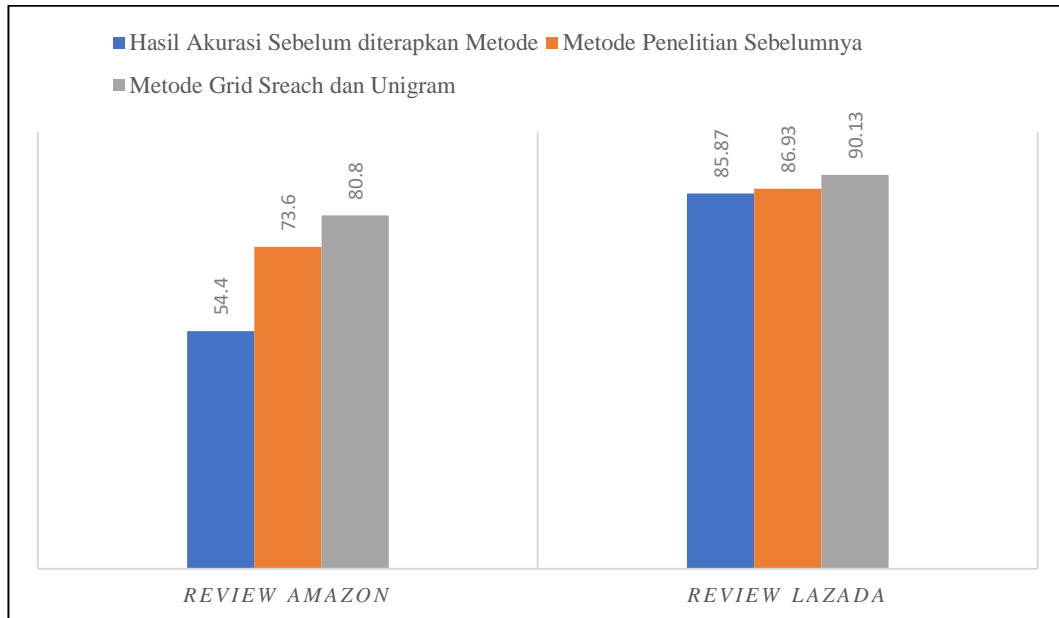
Menghasilkan akurasi 0,86933333 atau dalam satuan persen yaitu sebesar 86,93%.

Perbandingan akurasi pada penelitian sebelumnya dengan penelitian ini menggunakan metode *Grid Sreach* dan *Unigram* dapat dilihat pada tabel 4.10.

Tabel 4.10. Perbandingan Hasil Penelitian Sekarang dengan Penelitian Sebelumnya

Data	Hasil Sebelum diterapkan Metode	Metode Penelitian Sebelumnya	Metode <i>Grid Sreach</i> + <i>Unigram</i>
<i>Amazon</i>	54,40%	73,60%	80,80%
Lazada	85,87%	86,93%	90,13%

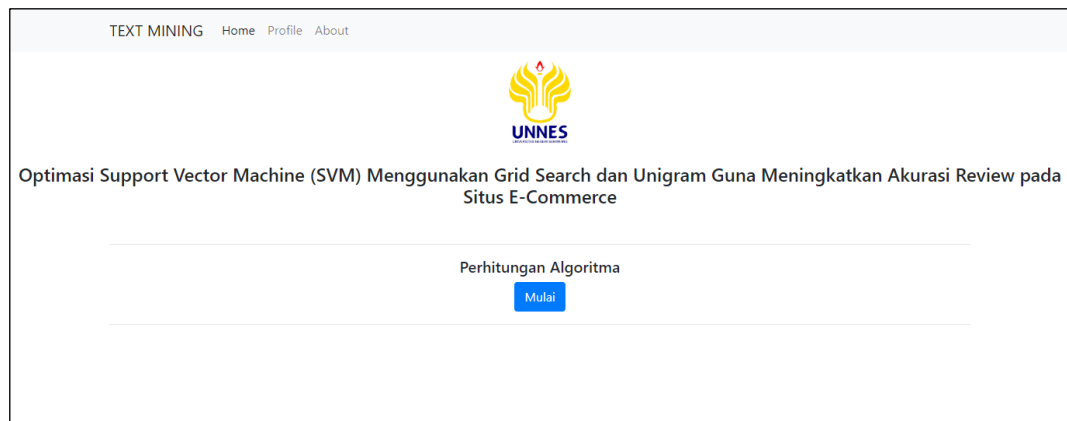
Adapun penyajian dalam bentuk grafik seperti berikut ini.



4.1.4 Hasil Implementasi Sistem

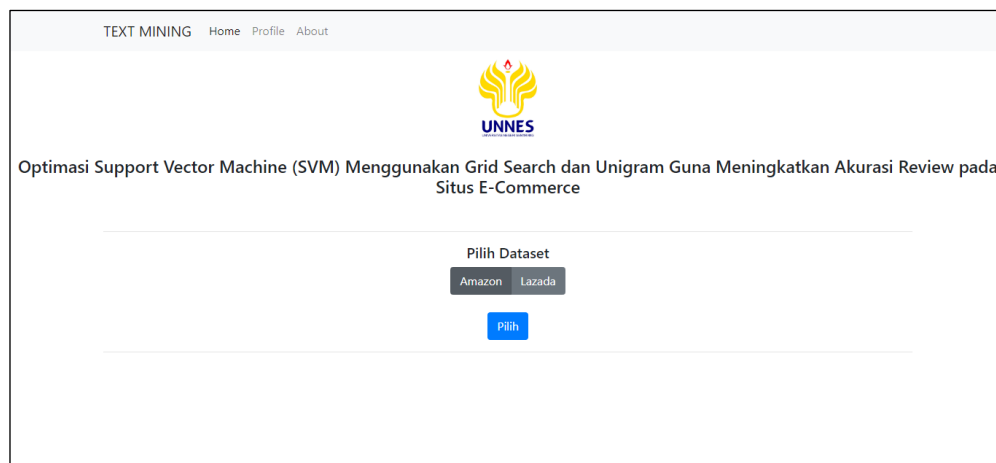
Pada penelitian ini sistem dirancang berbasis web. Aplikasi dibuat menggunakan bahasa pemrograman *Python* dan *Django* untuk membuat desain *user interface*. Sedangkan untuk melakukan pemrosesan juga menggunakan bahasa pemrograman *Python* dengan memanfaatkan *library Django 3.0.3*, *nlk 3.4.5*, *numpy 1.17.4*, *openpyxl 3.0.3*, *pandas 0.25.3*, *sastrawi 1.0.1*, *scikit-learn 0.22*, dan *xlrd 1.2.0*. Aplikasi ini memiliki 3 menu utama, yaitu *Home*, *Profil*, dan *About*.

Pada menu *Home* terdapat logo instansi dan keterangan judul penelitian. Desain tampilan dari menu *Home* dapat dilihat pada Gambar 4.1.



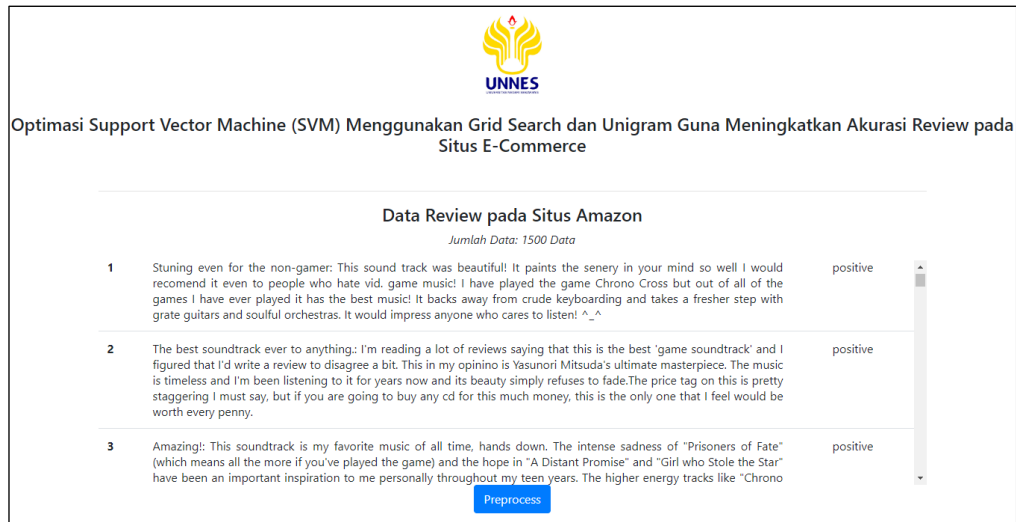
Gambar 4.1. Tampilan *Home* Aplikasi

Pada menu *Home*, terdapat tombol “Mulai” yang artinya untuk memulai perhitungan algoritma. Setelah klik tombol “Mulai” maka akan muncul pilihan *dataset* yang akan dihitung. Desain tampilan pilihan *dataset* dapat dilihat pada Gambar 4.2.



Gambar 4.2. Tampilan Pilihan *Dataset*

Setelah *dataset* dipilih pada tampilan pilihan *dataset*, maka akan muncul tampilan dari *dataset* dan terdapat tombol “Preprocess” untuk melakukan *preprocessing*. Desain tampilan *dataset* dapat dilihat pada Gambar 4.3.



Gambar 4.3. Tampilan *Dataset*

Pada tampilan *dataset* ketika tombol “Preprocess” diklik, maka akan ditampilkan hasil dari *preprocessing*. Desain tampilan hasil *preprocessing* dapat dilihat pada Gambar 4.4.

UNNES

Optimasi Support Vector Machine (SVM) Menggunakan Grid Search dan Unigram Guna Meningkatkan Akurasi Review pada Situs E-Commerce

Tahap Preprocessing
Transform Case - Punctuation Removal - Tokenize - Stopword Removal - Stemming


No.	Review	Setelah Preprocessing
1	Stuning even for the non-gamer: This sound track was beautiful! It paints the senery in your mind so well I would recomend it even to people who hate vid. game music! I have played the game Chrono Cross but out of all of the games I have ever played it has the best music! It backs away from crude keyboarding and takes a fresher step with grate guitars and souful orchestras. It would impress anyone who cares to listen! ^_^	stune even nongam sound track beauti paint seneri mind well would recomend even peopl hate vid game music play game chrono cross game ever play best music back away crude keyboard take fresher step grate guitar soul orchestra would impress anyon care listen
2	The best soundtrack ever to anything: I'm reading a lot of reviews saying that this is the best 'game soundtrack' and I figured that I'd write a review to disagree a bit. This in my opinino is Yasunori	best soundtrack ever anyth im read lot review say best game soundtrack figur id write review disagre bit opinino yasanori mitsuda ultim masterpiec music timeless im listen year beauti simpli refus

[Proses Data](#)

Gambar 4.4. Tampilan Hasil *Preprocessing*

Pada tampilan hasil *preprocessing* dapat dilihat perbedaan data sebelum dan sesudah dilakukan *preprocessing*. Kemudian ditampilan tersebut juga terdapat tombol “Proses Data” yang digunakan untuk memulai memproses data dengan *Unigram* dan *Grid Search* pada algoritma SVM. Setelah tombol “Proses Data” diklik maka akan muncul hasil akurasi perhitungan algoritma pada data *train* dan *test*. Desain tampilan hasil akurasi perhitungan algoritma untuk *dataset Amazon* dapat dilihat pada Gambar 4.5.

TEXT MINING Home Profile About


 UNNES

Optimasi Support Vector Machine (SVM) Menggunakan Grid Search dan Unigram Guna Meningkatkan Akurasi Review pada Situs E-Commerce

Hasil Eksperimen


No.	Skenario	Akurasi
1	SVM	54.40%
2	SVM + Unigram	80.80%
3	SVM + Unigram + Grid Search	C Optimal: 0.1 80.80%

[Ulangi Perhitungan](#)

Gambar 4.5. Tampilan Hasil Akurasi Perhitungan Algoritma untuk *Dataset Review Amazon*

Sedangkan desain tampilan hasil akurasi perhitungan algoritma untuk *dataset review Lazada* dapat dilihat pada Tabel 4.6.

TEXT MINING Home Profile About


 UNNES

Optimasi Support Vector Machine (SVM) Menggunakan Grid Search dan Unigram Guna Meningkatkan Akurasi Review pada Situs E-Commerce

Hasil Eksperimen

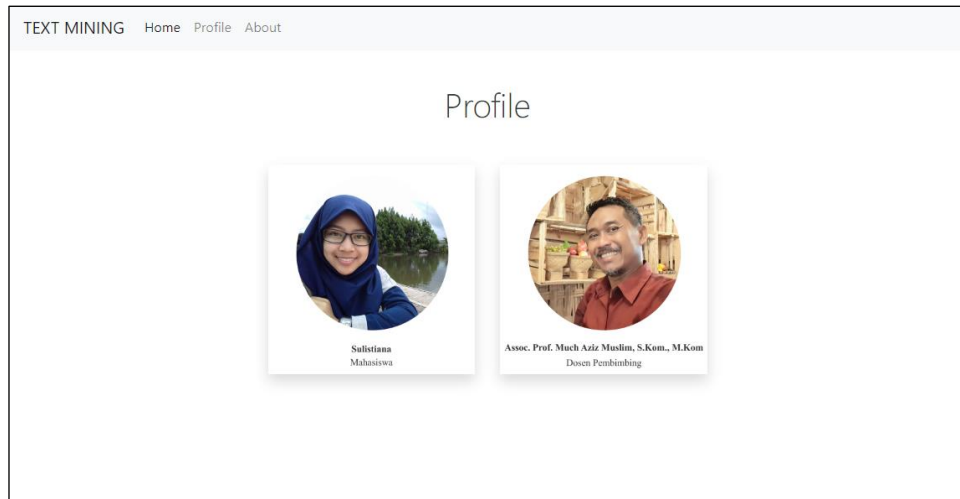
No.	Skenario	Akurasi
1	SVM	85.87%
2	SVM + Unigram	88.00%
3	SVM + Unigram + Grid Search	C Optimal: 0.58 90.13%

[Ulangi Perhitungan](#)

Gambar 4.6. Tampilan Hasil Akurasi Perhitungan Algoritma untuk *Dataset Review Lazada*

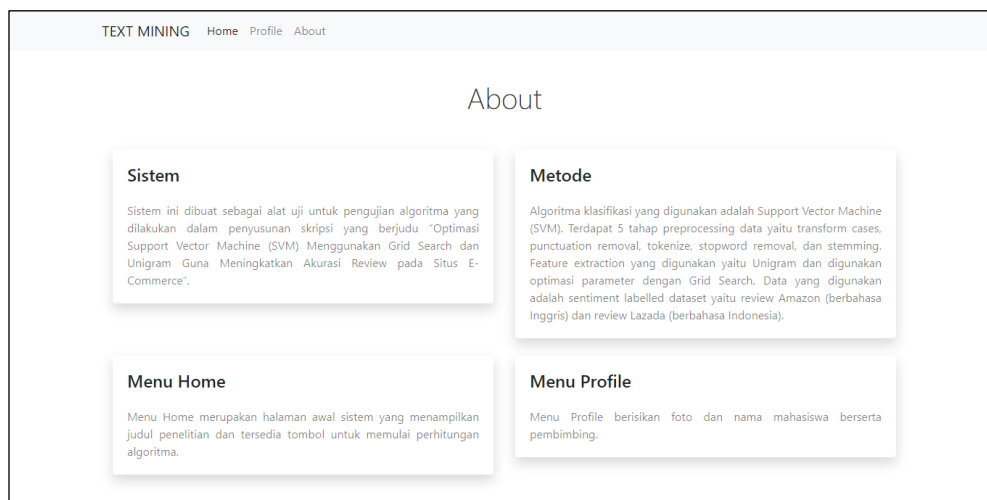
Pada menu *Profile*, berisi foto dan nama mahasiswa beserta pembimbing.

Desain tampilan dari menu *Profile* dapat dilihat pada Gambar 4.7.



Gambar 4.7. Tampilan Menu *Profile*

Pada menu *About*, berisi metode dan keterangan dari semua menu diantaranya *About Sistem*, *Menu Home*, dan *Menu Profile*. Desain tampilan dari menu *About* dapat dilihat pada Gambar 4.8.



Gambar 4.8. Tampilan Menu *About*

4.2. Pembahasan

Penelitian yang dilakukan adalah penerapan *Unigram* dan *Grid Search* untuk meningkatkan akurasi algoritma SVM pada klasifikasi analisis sentimen *review* situs *e-commerce* yaitu *Amazon* dan *Lazada*. Data yang digunakan adalah *sentiment labelled dataset* yaitu *review Amazon* (berbahasa Inggris) dan *review Lazada* (berbahasa Indonesia).

Hasil akurasi algoritma SVM pada klasifikasi analisis sentimen *review* situs *e-commerce* tanpa adanya *Unigram* dan *Grid Search* untuk *dataset review Amazon* adalah sebesar 54,40%. Setelah diberikan *Unigram* dan *Grid Search* algoritma SVM mencapai tingkat akurasi tertinggi dengan parameter C optimal=0,1 yaitu sebesar 80,80% dengan peningkatan akurasi sebesar 26,4%. Sedangkan untuk *dataset review Lazada* adalah sebesar 85,87%. Setelah diberikan *Unigram* dan *Grid Search* algoritma SVM mencapai tingkat akurasi tertinggi dengan parameter C optimal=0,58 yaitu sebesar 90,13% dengan peningkatan akurasi sebesar 4,26%.

Penerapan *Unigram* dan *Grid Search* pada algoritma SVM terbukti sebagai model yang cukup bagus dari penelitian sebelumnya untuk meningkatkan akurasi SVM pada analisis sentimen *review* situs *e-commerce* (*Amazon* dan *Lazada*) pada data *sentiment labelled dataset*. Pada penelitian sebelumnya untuk data *review Amazon* hanya dapat ditingkatkan sebesar 19,2% sedangkan dalam penelitian ini sebesar 26,4%. Untuk data *review Lazada* pada penelitian sebelumnya dapat ditingkatkan 1,06% sedangkan dalam penelitian ini sebesar 4,26%.

Dengan tingkat akurasi yang cukup baik, model ini diharapkan mampu menganalisis sentimen data *review* yang berbeda dengan tepat. Sehingga untuk penelitian selanjutnya, model ini dapat digunakan untuk klasifikasi analisis sentimen data *review* situs *e-commerce* lainnya.

Kelebihan dari penelitian ini yaitu dengan menerapkan *Unigram* dan *Grid Search* pada algoritma SVM dapat meningkatkan akurasi *review* situs *e-commerce* (*Amazon* dan *Lazada*) sehingga dapat digunakan sebagai referensi untuk penelitian berikutnya. Sedangkan kekurangan dari penelitian ini yaitu hasil parameter optimal *C* cenderung berubah-ubah karena penentuan kandidat nilai *C* ditentukan secara acak, penggunaan *dataset* hanya terbatas dua yaitu *review Amazon* dan *review Lazada*, dan metode *Grid Search* dan *Unigram* hanya diuji cobakan pada algoritma klasifikasi SVM.

BAB 5

PENUTUP

5.1. Kesimpulan

Dari hasil penelitian dan pembahasan terkait implementasi *Grid Search* dan *Unigram* untuk meningkatkan akurasi pada algoritma SVM menggunakan *dataset review* situs *e-commerce* yaitu *Amazon* dan *Lazada* yang diperoleh dari web *kaggle.com* dapat ditarik kesimpulan yaitu hasil akurasi dari penerapan algoritma SVM untuk *dataset review Amazon* sebesar 54,40% sedangkan setelah diterapkan *Unigram* didapatkan hasil akurasi sebesar 80,80% dan setelah dilakukan optimasi parameter menggunakan *Grid Search* didapat parameter C optimal yaitu 0,1 yang merupakan parameter inisial sehingga akurasi tetap 80,80%. Untuk *dataset review Lazada* sebesar 85,87% sedangkan setelah diterapkan *Unigram* didapatkan hasil akurasi sebesar 88,00% dan setelah dilakukan optimasi parameter menggunakan *Grid Search* didapat parameter C optimal yaitu 0,58 dengan peningkatan akurasi menjadi 90,13%. Dengan demikian, dapat disimpulkan bahwa penerapan *Grid Search* dan *Unigram* pada algoritma SVM dapat meningkatkan hasil akurasi *review Amazon* sebesar 26,40% dan dapat meningkatkan hasil akurasi *review Lazada* sebesar 4,26%. Hasil tersebut terbukti cukup bagus dibandingkan dengan penelitian sebelumnya yang hanya meningkatkan akurasi sebesar 19,2% pada data *review Amazon* dan 1,06% pada data *review Lazada*.

5.2. Saran

Untuk pengembangan lebih lanjut maka penulis memberikan saran sebagai berikut.

1. Untuk penelitian lebih lanjut apabila ingin mengembangkan metode penelitian ini dalam penggunaan parameter C SVM diharapkan sama, agar hasil akurasi tidak berubah-ubah.
2. Untuk penelitian lebih lanjut dapat melakukan uji coba menggunakan *dataset E-commerce* selain *Amazon* dan *Lazada*.
3. Untuk penelitian lebih lanjut dapat menerapkan kombinasi *Grid Search* dan *Unigram* untuk algoritma klasifikasi lainnya.

DAFTAR PUSTAKA

- Agyemang, M., Barker, K., & Alhadj, R. S. (2005). Mining web content outliers using structure oriented weighting techniques and N-grams. *Proceedings of the ACM Symposium on Applied Computing*, (pp. 482–487). Canada.
- Alex J. Smola, B. S. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3): 199–222.
- Aljawarneh, S., Yassein, M. B., & Aljundi, M. (2017). An enhanced J48 classification algorithm for the anomaly intrusion detection systems. *Cluster Computing*, 6(2): 1–17.
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint arXiv*, 17(07): 19-29.
- Amari, S., & Wu, S. (1999). Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, 12(6): 783–789.
- Anadakumar, V. P. (2013). A Survey on Preprocessing in Text Mining. *International Journal of Advanced Research in Computer Science*, 4(1): 79–91.
- Asgarnezhad, R., Shekofteh, M., & Boroujeni, F. Z. (2017). Improving diagnosis of diabetes mellitus using combination of preprocessing techniques. *Journal of Theoretical and Applied Information Technology*, 95(13): 2889–2895.
- Byun, H., & Lee, S. W. (2002). Applications of support vector machines for pattern recognition: A survey. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2388(5): 213–236.
- Chen, L., Mislove, A., & Wilson, C. (2016). An empirical analysis of algorithmic pricing on amazon marketplace. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 1339-1349). Canada.
- Chen, A., Lu, Y., & Wang, B. (2017). Customers' purchase decision-making process in social commerce: a social learning perspective. *International Journal of Information Management*, 37(6), 627-638.
- Chen, X., Xue, Y., Zhao, H., Lu, X., Hu, X., & Ma, Z. (2019). A novel feature extraction methodology for sentiment analysis of product reviews. *Neural Computing and Applications*, 31(10), 6625-6642.
- Diana, N. E., & Hanana Ulfa, I. (2019). Measuring performance of n-gram and jaccard-similarity metrics in document plagiarism application. *Journal of Physics: Conference Series*, 1196(1): 56-63.

- Eka Ratnawati, D., . M., & Muflikhah, L. (2014). Pengembangan Metode Klasifikasi Berdasarkan K-Means Dan LVQ. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 1(1): 1-14.
- Feldman, R., & Sanger, J. (2006). The Text Mining Handbook. In *The Text Mining Handbook*. New York: Cambridge University Press.
- Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, 1(1): 60-76.
- Gutiérrez, G., Ponce, J., Ochoa, A., & Álvarez, M. (2018). Analyzing students reviews of teacher performance using support vector machines by a proposed model. *Communications in Computer and Information Science*, 820(4): 113–122.
- Guyon, B. Boser, & V. Vapnik. (1993). Automatic Capacity Tuning of Very Large VC-Dimension Classifiers. *Advances in Neural Information Processing Systems*, 5(6): 147–155.
- Haque, T. U., Saber, N. N., & Shah, F. M. (2018). Sentiment analysis on large scale Amazon product reviews. *IEEE International Conference on Innovative Research and Development, ICIRD*, (pp. 1-6). Bangladesh.
- Hermawati, F. A. (2013). *Data mining*. Yogyakarta: CV. Andi Offset.
- Jena, L. & N. K. Kamila. (2015). Distributed Data Mining Classification Algorithms for Prediction of Chronic Kidney Disease. *International Journal of Emerging Research in Management and Technology*, 4(11): 110-118.
- Kamavisdar, P., Saluja, S., & Agrawal, S. (2013). A Survey on Image Classification Approaches and Techniques. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(1): 1005–1009.
- Kavzoglu, I. C. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation*, 11(4): 352–359.
- Khreisat, L. (2009). A machine learning approach for Arabic text classification using N-gram frequency statistics. *Journal of Informetrics*, 3(1): 72–77.
- Laoh, E., Surjandari, I., & Prabaningtyas, N. I. (2019). Enhancing hospitality sentiment reviews analysis performance using SVM N-grams method. *2019 16th International Conference on Service Systems and Service Management, ICSSSM 2019*, (pp. 1–5). Depok.
- Leidiyana, H. (2013). Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, 1(1): 65–76.

- Lidya, S. K., Sitompul, O. S., & Efendi, S. (2015). Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM). *Seminar Nasional Teknologi Dan Komunikasi 2015*, (pp. 1–8). Yogyakarta.
- Lin, S. W., Ying, K. C., Chen, S. C., & Lee, Z. J. (2008). Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications*, 35(4): 1817–1824.
- Liu, K., Luo, X., & Zhang, L. (2016). Evaluation of China's B2C e-commerce website: an analysis of factors that influence online buying decision. *International Journal of Multimedia and Ubiquitous Engineering*, 11(3), 143-156.
- Muslim, M. A., Prasetyo, B., Listiana, E., Mawarni, E. L. H., Juli, A., Mirqotussa'adah., Rukmana, S. H., & Nurzahputra, A. (2019). *Data Mining Algoritma C4.5*. Semarang: CV. Pilar Nusantara.
- Nema, A., Tiwari, B., & Tiwari, V. (2016). Improving accuracy for intrusion detection through layered approach using support vector machine with feature reduction. *ACM International Conference Proceeding Series*, (pp. 26–31). Bhopal.
- Nugroho, S., Witarto, B., & Handoko, D. (2003). Support Vector Machine Teori dan Aplikasinya dalam Bioinformatika1. *Proceeding of Indonesian Scientific Meeting in Central Japan*, (pp. 243–252). Japan.
- Nurzahputra, A., & Muslim, M. A. (2016). Analisis Sentimen pada Opini Mahasiswa Menggunakan Natural Language Processing. *Seminar Nasional Ilmu Komputer (SNIK)*, (pp. 114–118). Semarang.
- Paknejad, S. (2018). *Sentiment classification on Amazon reviews using machine learning approaches* ("Orjan Ekeberg, ed.). Stockholm: KTH Royal Institute Of Technology School.
- Parul Sinha, & Poonam Sinha. (2015). Comparative Study of Chronic Kidney Disease Prediction using KNN and SVM. *International Journal of Engineering Research And Technology*, 4(12): 56-64.
- Prasetyo, E. (2014). *Data mining: Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: CV. Andi Offset.
- Prasetyo, P. D., Suta Wijaya, I. G. P., & Yudo Husodo, A. (2019). Klasifikasi Genre Musik Menggunakan Metode Mel-Frequency Cepstrum Coefficients dan K-Nearest Neighbors Classifier. *Jurnal Teknologi Informasi, Komputer, Dan Aplikasinya (JTIKA)*, 1(2): 189–197.
- Ramanathan, V., & Meyyappan, T. (2019). Twitter text mining for sentiment analysis on people's feedback about Oman tourism. *2019 4th MEC International Conference on Big Data and Smart City, ICBDS 2019*, (pp. 1–5). Singapore.

- Ravi, A., Khettry, A. R., & Sethumadhavachar, S. Y. (2019). Amazon Reviews as Corpus for Sentiment Analysis Using Machine Learning. *International Conference on Advances in Computing and Data Sciences* (pp. 403-411). Singapore.
- Sartika, D., & Indra, D. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian. *Jurnal Teknik Informatika Dan Sistem Informasi*, 1(2): 151–161.
- Saura, J. R., & Bennett, D. R. (2019). A three-stage method for data text mining: Using UGC in business intelligence analysis. *Symmetry*, 11(4): 31-39.
- Sbou, A. M. F. (2018). A survey of Arabic text classification models. *International Journal of Electrical and Computer Engineering*, 8(6): 4352–4355.
- Srivastava, K., & Bhambhu, L. (2009). Data Classification Using Support Vector Machine. *Journal of Theoretical and Applied Information Technology*, 1(5): 1–7.
- Syarif, I., Prugel-Bennett, A., & Wills, G. (2016). SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *Telkomnika (Telecommunication Computing Electronics and Control)*, 14(4): 1502–1509.
- Tharwat, A. (2019). Parameter investigation of support vector machine classifier with kernel functions. *Knowledge and Information Systems*. 6(2): 24-31.
- Tuhin, R. A., Paul, B. K., Nawrine, F., Akter, M., & Das, A. K. (2019). An Automated System of Sentiment Analysis from Bangla Text using Supervised Learning Techniques. *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, (pp. 360–364). India.
- Williams, T. P., & Gong, J. (2014). Predicting construction cost overruns using text mining, numerical data and ensemble classifiers. *Automation in Construction*, 43(3): 23–29.
- Lameski, P., Zdravevski, E., Mingov, R., & Kulakov, A. (2015). SVM parameter tuning with grid search and its impact on reduction of model over-fitting. *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing* (pp. 464-474). Cham.
- Zakrani, A., Najm, A., & Marzak, A. (2018). Support Vector Regression Based on Grid-Search Method for Agile Software Effort Prediction. *Colloquium in Information Science and Technology*, 8(2): 26-32.
- Zhan, J., Tong, H., & Liu, Y. (2009). Gather customer concerns from online product reviews – A text summarization approach. *Expert Systems With Applications*, 36(2): 2107–2115.

Zuhdi, A. M., Utami, E., & Raharjo, S. (2019). Analisis Sentiment Twitter Terhadap Capres Indonesia 2019 Dengan Metode K-NN. *Jurnal Informa Politeknik Indonusa Surakarta*, 5(13): 1–7.

LAMPIRAN

*Lampiran 1.***Kode Program Sistem GUI Python***Source Code views.py difolder projects*

```

import json
import os
import pandas as pd
from django.shortcuts import render
from projects.services import load_data, preprocess, process_data

from django.conf import settings

def project_index(request):
    return render(request, "project_index.html", {})

def project_detail(request, dataset):
    df = load_data(dataset)

    reviews = df.to_json(orient="records")
    reviews = json.loads(reviews)

    context = {"dataset": dataset.title(), "reviews": reviews,
"total": len(df)}
    return render(request, "project_detail.html", context)

def project_preprocess(request, dataset):
    df = preprocess(dataset)

    reviews = df.to_json(orient="records")
    reviews = json.loads(reviews)

    context = {"dataset": dataset.title(), "reviews": reviews}
    return render(request, "project_preprocessing.html", context)

def project_result(request, dataset):
    context = process_data(dataset)
    return render(request, "project_result.html", context)

```

Source Code services.py difolder projects

```

import json
import os
import numpy as np
import pandas as pd
import string

from django.conf import settings
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize

```

```

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from sklearn.base import BaseEstimator, TransformerMixin
from sklearn.model_selection import train_test_split, cross_val_score,
GridSearchCV
from sklearn.linear_model import SGDClassifier
from sklearn.pipeline import Pipeline
from sklearn.svm import SVC
from sklearn.feature_extraction.text import TfidfVectorizer,
CountVectorizer
from sklearn.pipeline import FeatureUnion
from sklearn.metrics import classification_report

class TextSelector(BaseEstimator, TransformerMixin):
    """
    Transformer to select a single column from
    the data frame to perform additional transformations on
    Use on text columns in the data
    """
    def __init__(self, key):
        self.key = key

    def fit(self, X, y=None):
        return self

    def transform(self, X):
        return X[self.key]

def load_data(dataset):
    filename = "review_lazada.xlsx" if dataset.lower() == "lazada"
    else "review_amazon.xlsx"
    if settings.LOAD_FROM_FILE == True:
        filename = "lazada_preprocessed.xlsx" if dataset.lower() ==
        "lazada" else "amazon_preprocessed.xlsx"

    input_file = os.path.join(settings.DATASET_PATH, filename)

    # load data
    df = pd.read_excel(input_file)
    df.dropna(axis=0)
    return df

def load_stopwords(lang):
    stopwords_list = set(stopwords.words(lang))
    return stopwords_list

def build_stemmer(lang):
    if lang == 'english':
        # english stemmer
        stemmer_obj = PorterStemmer()
    elif lang == 'indonesian':
        # indonesian stemmer
        factory = StemmerFactory()
        stemmer_obj = factory.create_stemmer()
    print("build stemmer: done")
    return stemmer_obj

```

```

def preprocess_data(text, stopwords_list, stemmer_obj):
    print(text)
    # transform case
    text = text.lower().strip()

    # punctuation removal
    table = str.maketrans(dict.fromkeys(string.punctuation))
    text = text.translate(table)

    # stopwords removal
    words = word_tokenize(text)
    words_filtered = []

    for w in words:
        if w in stopwords_list:
            continue
        words_filtered.append(stemmer_obj.stem(w))

    if words_filtered == []:
        words_filtered = words

    return " ".join(words_filtered)

def preprocess(dataset):
    df = load_data(dataset)
    if settings.LOAD_FROM_FILE == True:
        return df

    lang = 'english' if dataset == 'amazon' else 'indonesian'
    stopwords_list = load_stopwords(lang)
    stemmer_obj = build_stemmer(lang)

    ## preprocessing steps ##
    df['processed'] = df['Review'].apply(lambda x: preprocess_data(x,
stopwords_list, stemmer_obj))

    # write results to file
    if dataset == 'amazon':
        df.to_excel(os.path.join(settings.DATASET_PATH,
'amazon_preprocessed.xlsx'), index=False)
    elif dataset == 'lazada':
        df.to_excel(os.path.join(settings.DATASET_PATH,
'lazada_preprocessed.xlsx'), index=False)

    return df

def enumerate_label(x):
    # convert label 'negative' as 0 and label 'positive' as 1
    return 0 if x == 'negative' else 1

def process_data(dataset):
    filename = 'result_{}.json'.format(dataset)
    file_path = os.path.join(settings.DATASET_PATH, filename)
    if settings.LOAD_FROM_FILE == True:
        file_obj = open(file_path, 'r')

```

```

        results = json.load(file_obj)
        return results

    filename = "lazada_preprocessed.xlsx" if dataset.lower() ==
"lazada" else "amazon_preprocessed.xlsx"
    input_file = os.path.join(settings.DATASET_PATH, filename)

    # load data
    df = pd.read_excel(input_file)
    df.dropna(axis=0)

    # convert label 'negative' as 0 and label 'positive' as 1
    df['Label'] = df['Label'].apply(lambda x: enumerate_label(x))

    # data splitting
    features = [c for c in df.columns.values if c not in []]

    # # define the label column
    target = 'Label'

    # use 25% of the dataset as testing data
    test_size = 0.25

    # just a random number
    random_state = 42

    print("split data")
    X_train, X_test, y_train, y_test = train_test_split(df[features],
df[target],
test_size=test_size,
random_state=random_state)

    k_cross_val = 10

    # # scenario 1
    print("scenario 1")
    text_clf1 = Pipeline([
        ('selector', TextSelector(key='processed')), # use
preprocessed text
        ('tfidf', TfidfVectorizer()), # tf-idf
        ('clf', SVC(kernel='linear', C=0.1)), # SVM linear
    ])

    # # train model
    text_clf1.fit(X_train, y_train)

    # calculating training data's accuracy using 10-fold cross
validation (without grid search)
    train_accuracies1 = cross_val_score(text_clf1, X_train, y_train,
cv=k_cross_val, scoring='accuracy', verbose=2)
    train_accuracy1 = train_accuracies1.mean()

    test_predict1 = text_clf1.predict(X_test)
    test_accuracy1 = np.mean(test_predict1 == y_test)

```

```

print ("result 1", train_accuracy1, test_accuracy1)

# # scenario 2
print("scenario 2")
text_clf2 = Pipeline([
    ('selector', TextSelector(key='processed')), # use
preprocessed text
    ('word_ngram', CountVectorizer(ngram_range=(1, 1),
analyzer='word')), # word unigram
    ('clf', SVC(kernel='linear', C=0.1)), # SVM linear
])

# # train model
text_clf2.fit(X_train, y_train)

# calculating training data's accuracy using 10-fold cross
validation (without grid search)
train_accuracies2 = cross_val_score(text_clf2, X_train, y_train,
cv=k_cross_val, scoring='accuracy', verbose=2)
train_accuracy2 = train_accuracies2.mean()

test_predict2 = text_clf2.predict(X_test)
test_accuracy2 = np.mean(test_predict2 == y_test)

print ("result 2", train_accuracy2, test_accuracy2)

# scenario 3
print("scenario 3")
c_candidates = [0.1, 0.18, 0.26, 0.34, 0.42, 0.5, 0.58, 0.66,
0.74, 0.82, 0.9]
parameters = {
    'clf__C': c_candidates,
}
gs_clf = GridSearchCV(text_clf2, parameters, cv=k_cross_val,
n_jobs=-1, verbose=2)
gs_clf = gs_clf.fit(X_train, y_train)

scores = gs_clf.cv_results_['mean_test_score']

gs_score = []
for i, c in enumerate(c_candidates):
    sc = {'c': c, 'accuracy': "{0:.2%}".format(scores[i])}
    sc['best_param'] = True if c == gs_clf.best_params_['clf__C']
else False
    gs_score.append(sc)

# test data
test_predict3 = gs_clf.predict(X_test)
test_accuracy3 = np.mean(test_predict3 == y_test)
print("3", test_accuracy3, "{0:.2%}".format(test_accuracy3))

results = {
    "scenario1": {"training_accuracy":
"{0:.2%}".format(train_accuracy1),

```

```

        "testing_accuracy":
"{0:.2%}".format(test_accuracy1)},
        "scenario2": {"training_accuracy":
"{0:.2%}".format(train_accuracy2),
        "testing_accuracy":
"{0:.2%}".format(test_accuracy2)},
        "scenario3": {"training_accuracy": gs_score,
        "testing_accuracy":
"{0:.2%}".format(test_accuracy3)},
    }

    print (results)

    # save results to file
    file_obj = open(file_path, 'w')
    json.dump(results, file_obj)

    return results

```

Source Code project_detail.html

```

{% extends "base.html" %}
{% load static %}

{% block page_content %}

<script type="text/javascript">

var dataset = "{{ dataset }}"&#x2D;toLowerCase();
function preprocess() {
    window.location = "http://127.0.0.1:8000/preprocess/" + dataset +
"";
}

</script>

<div class="header">
    <p align="center"></p>
    <h4 align="center" class="font-weight-light text-dark">Optimasi
Support Vector Machine (SVM) Menggunakan Grid Search dan Unigram Guna
Meningkatkan Akurasi Review pada Situs E-Commerce</h4>
</div>

<hr>

<div>
    <h4 align="center">Data Review pada Situs {{dataset}}</h4>
    <p align="center"><i>Jumlah Data: {{total}} Data</i></p>
</div>

<div id="dataset" style="height:300px;" class="table-responsive">
    <table class="table table-fixed">
        <thead class="thead-dark">

```

```

        <tr align="center">
            <th scope="col" width="5%">No.</th>
            <th scope="col" width="80%">Review</th>
            <th scope="col" width="15%">Label</th>
        </tr>
    </thead>
    <tbody>
        {% for review in reviews %}
        <tr>
            <th scope="row" align="center">{{review.No}}</th>
            <td align="justify">{{review.Review}}</td>
            <td align="center">{{review.Label}}</td>
        </tr>
        {% endfor%}
    </tbody>
</table>
</div>

<div id="preprocess-section">
    <p align="center"><button type="button" id="start"
onClick="preprocess()" class="btn btn-primary">Preprocess</button></p>
</div>

<hr>

{% endblock %}

```

Source Code project_index.html

```

{% extends "base.html" %}
{% load static %}
{% block page_content %}

<script src="{% static "jquery-3.4.1.min.js" %}"></script>
<script type="text/javascript">

$(document).ready(function(){
    $("#start-section").click(function(){
        $(this).hide();
        $("#select-data-section").show();
    });

    $("#select-data").click(function(){
        var dataset = "";
        if ($("#amazon").is(":checked")){
            dataset = $("#amazon").val();
        } else if ($("#lazada").is(":checked")) {
            dataset = $("#lazada").val();
        } else {
            alert("Silakan pilih dataset");
            return;
        }
    }
}

```



```

        window.location = "http://127.0.0.1:8000/detail/" + dataset +
    "";
    });
});
</script>

<div class="header">
    <p align="center"></p>
    <h4 align="center" class="font-weight-light text-dark">Optimasi
Support Vector Machine (SVM) Menggunakan Grid Search dan Unigram Guna
Meningkatkan Akurasi Review pada Situs E-Commerce</h4>
</div>

<hr>

<div id="start-section">
    <h5 align="center" class="font-weight-heavy text-dark">Perhitungan
Algoritma</h5>
    <p align="center"><button type="button" id="start" class="btn btn-
primary">Mulai</button></p>
</div>

<div id="select-data-section" style="display:none" align="center">
    <h5 align="center">Pilih Dataset</h5>
    <div class="btn-group btn-group-toggle" data-toggle="buttons"
align="center">
        <label class="btn btn-secondary active">
            <input type="radio" name="options" id="amazon"
value="amazon" autocomplete="off"> Amazon
        </label>
        <label class="btn btn-secondary">
            <input type="radio" name="options" id="lazada"
value="lazada" autocomplete="off"> Lazada
        </label>
    </div>
    <br/><br/>
    <button type="button" id="select-data" class="btn btn-
primary">Pilih</button>
</div>
<hr>

{% endblock %}

```

Source Code project_preprocessing.html

```

{% extends "base.html" %}
{% load static %}

{% block page_content %}

<script type="text/javascript">

```

```

var dataset = "{{ dataset }}" .toLowerCase();
function showResult() {
    window.location = "http://127.0.0.1:8000/result/" + dataset + "";
}
</script>

<div class="header">
    <p align="center"></p>
    <h4 align="center" class="font-weight-light text-dark">Optimasi
Support Vector Machine (SVM) Menggunakan Grid Search dan Unigram Guna
Meningkatkan Akurasi Review pada Situs E-Commerce</h4>
</div>

<hr>

<div>
    <h4 align="center">Tahap Preprocessing</h4>
    <p align="center"><i>Transform Case - Punctuation Removal -
Tokenize - Stopword Removal - Stemming</i></p>
</div>

<br/>

<div style="height:300px;" class="table-responsive">
    <table class="table table-fixed">
        <thead class="thead-dark">
            <tr align="center">
                <th scope="col" width="5%">No.</th>
                <th scope="col" width="47.5%">Review</th>
                <th scope="col" width="47.5%">Setelah Preprocessing</th>
            </tr>
        </thead>
        <tbody>
            {% for review in reviews %}
            <tr>
                <th scope="row">{{review.No}}</th>
                <td align="justify">{{review.Review}}</td>
                <td align="justify">{{review.processed}}</td>
            </tr>
            {% endfor%}
        </tbody>
    </table>
</div>

<div id="next-section">
    <p align="center"><button type="button" id="start"
onClick="showResult()" class="btn btn-primary">Proses
Data</button></p>
</div>

<hr>

{% endblock %}

```

Source Code project_result.html

```

{% extends "base.html" %}
{% load static %}

{% block page_content %}

<script type="text/javascript">

function backToHome() {
    window.location = "http://127.0.0.1:8000/";
}

</script>

<div class="header">
    <p align="center"></p>
    <h4 align="center" class="font-weight-light text-dark">Optimasi
Support Vector Machine (SVM) Menggunakan Grid Search dan Unigram Guna
Meningkatkan Akurasi Review pada Situs E-Commerce</h4>
</div>

<hr>

<div>
    <h4 align="center">Hasil Eksperimen</h4>
</div>

<table class="table table-striped">
    <thead class="thead-dark">
        <tr align="center">
            <th scope="col">No.</th>
            <th scope="col">Skenario</th>
            <th scope="col">Akurasi Training</th>
            <th scope="col">Akurasi Testing</th>
        </tr>
    </thead>
    <tbody>

        <tr>
            <th scope="row">1</th>
            <td>SVM</td>
            <td>{{scenario1.training_accuracy}}</td>
            <td>{{scenario1.testing_accuracy}}</td>
        </tr>

        <tr>
            <th scope="row">2</th>
            <td>SVM + Unigram</td>
            <td>{{scenario2.training_accuracy}}</td>
            <td>{{scenario2.testing_accuracy}}</td>
        </tr>

        <tr>
            <th scope="row">3</th>

```

```

        <td>SVM + Unigram + Grid Search</td>
        <td>

            {% for score in scenario3.training_accuracy %}
                {% if score.best_param == True %}
                    <p><strong>C: {{score.c}}, Akurasi:
                {{score.accuracy}}</strong></p>
                {% else %}
                    <p>C: {{score.c}}, Akurasi: {{score.accuracy}}</p>
                {% endif %}
            {% endfor %}

        </td>
        <td>{{scenario3.testing_accuracy}}</td>
    </tr>
</tbody>
</table>

<div id="next-section">
    <p align="center"><button type="button" id="start"
onClick="backToHome()" class="btn btn-primary">Ulangi
Perhitungan</button></p>
</div>

<hr>

{% endblock %}

```

Source Code views.py difolder profile

```

from django.shortcuts import render
from django.conf import settings

def profile_index(request):
    return render(request, "profile_index.html", {})

```

Source Code profile_index.html

```

{% extends "base.html" %}
{% load static %}
{% block page_content %}

<!-- Header -->
<header class="bg text-center py-3 mb-4">
    <h1 class="font-weight-light text-dark">Profile</h1>
</header>

<!-- Page Content -->
<div class="container">
    <div class="row">
        <!-- empty space -->

```

```

<div class="col-xl-3 col-md-6 mb-4">
  <div class="card border-0 shadow">
    </div>
  </div>

  <!-- Team Member 1 -->
  <div class="col-xl-3 col-md-6 mb-4">
    <div class="card border-0 shadow">
      
      <div class="card-body text-center">
        <h5 class="card-title mb-0">Sulistiana</h5>
        <div class="card-text text-black-50">Mahasiswa <br /></div>
      </div>
    </div>
  </div>
  <!-- Team Member 2 -->
  <div class="col-xl-3 col-md-6 mb-4">
    <div class="card border-0 shadow">
      
      <div class="card-body text-center">
        <h6 class="card-title mb-0">Bapak Much Aziz Muslim, S.Kom.,
M.Kom.</h6>
        <div class="card-text text-black-50">Dosen Pembimbing</div>
      </div>
    </div>
  </div>

  <!-- empty space -->
  <div class="col-xl-3 col-md-6 mb-4">
    <div class="card border-0 shadow">
      </div>
    </div>

  <!-- /.row -->
</div>
<!-- /.container -->

{% endblock %}

```

Source Code views.py difolder about

```

from django.shortcuts import render
from django.conf import settings

def about_index(request):
    return render(request, "about_index.html", {})

```

Source Code about_index.html

```

{% extends "base.html" %}
{% load static %}
{% block page_content %}

<!-- Header -->
<header class="bg text-center py-3 mb-4">
  <h1 class="font-weight-light text-dark">About</h1>
</header>

<!-- Page Content -->
<div class="container">
  <div class="row">
    <div class="col-xl-6 col-md-6 mb-4">
      <div class="card border-0 shadow">
        <div class="card-body text-center">
          <h4 class="card-title mb-0" align="left">Sistem</h4>
          <br />
          <div class="card-text text-black-50" align="justify">Sistem
ini dibuat sebagai alat uji untuk pengujian algoritma yang dilakukan
dalam penyusunan skripsi yang berjudul "Optimasi Support Vector Machine
(SVM) Menggunakan Grid Search dan Unigram Guna Meningkatkan Akurasi
Review pada Situs E-Commerce".</div>
        </div>
      </div>
    </div>

    <div class="col-xl-6 col-md-6 mb-4">
      <div class="card border-0 shadow">
        <div class="card-body text-center">
          <h4 class="card-title mb-0" align="left">Metode</h4>
          <br />
          <div class="card-text text-black-50"
align="justify">Algoritma klasifikasi yang digunakan adalah Support
Vector Machine (SVM). Terdapat 5 tahap preprocessing data yaitu
transform cases, punctuation removal, tokenize, stopword removal, dan
stemming. Feature extraction yang digunakan yaitu Unigram dan
digunakan optimasi parameter dengan Grid Search. Data yang digunakan
adalah sentiment labelled dataset yaitu review Amazon (berbahasa
Inggris) dan review Lazada (berbahasa Indonesia).</div>
        </div>
      </div>
    </div>

  </div>

  <div class="row">
    <div class="col-xl-6 col-md-6 mb-4">
      <div class="card border-0 shadow">
        <div class="card-body text-center">
          <h4 class="card-title mb-0" align="left">Menu Home</h4>
          <br />
          <div class="card-text text-black-50" align="justify">Menu
Home merupakan halaman awal sistem yang menampilkan judul penelitian
dan tersedia tombol untuk memulai perhitungan algoritma.
        </div>
      </div>
    </div>
  </div>

```

```
        </div>
    </div>
</div>

<div class="col-xl-6 col-md-6 mb-4">
    <div class="card border-0 shadow">
        <div class="card-body text-center">
            <h4 class="card-title mb-0" align="left">Menu Profile</h4>
            <br />
            <div class="card-text text-black-50" align="justify">Menu
Profile berisikan foto dan nama mahasiswa beserta pembimbing.</div>
        </div>
    </div>
</div>

<!-- /.row -->

</div>
<!-- /.container -->

{% endblock %}
```

Lampiran 2

Dataset Review Amazon

__label__2 Stuning even for the non-gamer: This sound track was beautiful! It paints the senery in your mind so well I would recomend it even to people who hate vid. game music! I have played the game Chrono Cross but out of all of the games I have ever played it has the best music! It backs away from crude keyboarding and takes a fresher step with grate guitars and soulful orchestras. It would impress anyone who cares to listen! ^_^

__label__2 The best soundtrack ever to anything.: I'm reading a lot of reviews saying that this is the best 'game soundtrack' and I figured that I'd write a review to disagree a bit. This in my opinino is Yasunori Mitsuda's ultimate masterpiece. The music is timeless and I'm been listening to it for years now and its beauty simply refuses to fade.The price tag on this is pretty staggering I must say, but if you are going to buy any cd for this much money, this is the only one that I feel would be worth every penny.

__label__2 Amazing!: This soundtrack is my favorite music of all time, hands down. The intense sadness of "Prisoners of Fate" (which means all the more if you've played the game) and the hope in "A Distant Promise" and "Girl who Stole the Star" have been an important inspiration to me personally throughout my teen years. The higher energy tracks like "Chrono Cross ~ Time's Scar~", "Time of the Dreamwatch", and "Chronomantique" (indefinably remeniscent of Chrono Trigger) are all absolutely superb as well.This soundtrack is amazing music, probably the best of this composer's work (I haven't heard the Xenogears soundtrack, so I can't say for sure), and even if you've never played the game, it would be worth twice the price to buy it.I wish I could give it 6 stars.

__label__2 Excellent Soundtrack: I truly like this soundtrack and I enjoy video game music. I have played this game and most of the music on here I enjoy and it's truly relaxing and peaceful.On disk one. my favorites are Scars Of Time, Between Life and Death, Forest Of Illusion, Fortress of Ancient Dragons, Lost Fragment, and Drowned Valley.Disk Two: The Draggons, Galdorb - Home, Chronomantique, Prisoners of Fate, Gale, and my girlfriend likes ZebessDisk Three: The best of the three. Garden Of God, Chronopolis, Fates, Jellyfish sea, Burning Orphange, Dragon's Prayer, Tower Of Stars, Dragon God, and Radical Dreamers - Unstealable Jewel.Overall, this is a excellent soundtrack and should be brought by those that like video game music.Xander Cross

__label__2 Remember, Pull Your Jaw Off The Floor After Hearing it: If you've played the game, you know how divine the music is! Every single song tells a story of the game, it's that good! The greatest songs are without a doubt, Chrono Cross: Time's Scar, Magical Dreamers: The Wind, The Stars, and the Sea and Radical Dreamers: Unstolen Jewel. (Translation varies) This music is perfect if you ask me, the best it can be. Yasunori Mitsuda just poured his heart on and wrote it down on paper.

__label__2 an absolute masterpiece: I am quite sure any of you actually taking the time to read this have played the game at least once, and heard at least a few of the tracks here. And whether you were aware of it or not, Mitsuda's music contributed greatly to the mood of every single minute of the whole game.Composed of 3 CDs and quite a few songs (I haven't an exact count), all of which are heart-rendering and impressively remarkable, this soundtrack is one I assure you you will not forget. It has everything for every listener -- from fast-paced and energetic (Dancing the Tokage or Termina Home), to slower and more haunting (Dragon God), to purely beautifully composed (Time's Scar), to even some fantastic vocals (Radical Dreamers).This is one of the best videogame soundtracks out there, and surely Mitsuda's best ever. ^_^

__label__1 Buyer beware: This is a self-published book, and if you want to know why--read a few paragraphs! Those 5 star reviews must have been written by Ms. Haddon's family and friends--or perhaps, by herself! I can't imagine anyone reading the whole thing--I spent an evening with the book and a friend and we were in hysterics reading bits and pieces of it to one another. It is most

definitely bad enough to be entered into some kind of a "worst book" contest. I can't believe Amazon even sells this kind of thing. Maybe I can offer them my 8th grade term paper on "To Kill a Mockingbird"--a book I am quite sure Ms. Haddon never heard of. Anyway, unless you are in a mood to send a book to someone as a joke---stay far, far away from this one!

__label__2 Glorious story: I loved Whisper of the wicked saints. The story was amazing and I was pleasantly surprised at the changes in the book. I am not normally someone who is into romance novels, but the world was raving about this book and so I bought it. I loved it !! This is a brilliant story because it is so true. This book was so wonderful that I have told all of my friends to read it. It is not a typical romance, it is so much more. Not reading this book is a crime, because you are missing out on a heart warming story.

__label__2 A FIVE STAR BOOK: I just finished reading Whisper of the Wicked saints. I fell in love with the characters. I expected an average romance read, but instead I found one of my favorite books of all time. Just when I thought I could predict the outcome I was shocked ! The writing was so descriptive that my heart broke when Julia's did and I felt as if I was there with them instead of just a distant reader. If you are a lover of romance novels then this is a must read. Don't let the cover fool you this book is spectacular!

__label__2 Whispers of the Wicked Saints: This was a easy to read book that made me want to keep reading on and on, not easy to put down.It left me wanting to read the follow on, which I hope is coming soon. I used to read a lot but have gotten away from it. This book made me want to read again. Very enjoyable.

__label__1 The Worst!: A complete waste of time. Typographical errors, poor grammar, and a totally pathetic plot add up to absolutely nothing. I'm embarrassed for this author and very disappointed I actually paid for this book.

__label__2 Great book: This was a great book,I just could not put it down,and could not read it fast enough. Boy what a book the twist and turns in this just keeps you guessing and wanting to know what is going to happen next. This book makes you fall in love and can heat you up,it can also make you so angry. this book can make you go thru several of your emotions. This is a quick read romance. It is something that you will want to end your day off with if you read at night.

__label__2 Great Read: I thought this book was brilliant, but yet realistic. It showed me that to error is human. I loved the fact that this writer showed the loving side of God and not the revengeful side of him. I loved how it twisted and turned and I could not put it down. I also loved The glass castle.

__label__1 Oh please: I guess you have to be a romance novel lover for this one, and not a very discerning one. All others beware! It is absolute drivel. I figured I was in trouble when a typo is prominently featured on the back cover, but the first page of the book removed all doubt. Wait - maybe I'm missing the point. A quick re-read of the beginning now makes it clear. This has to be an intentional churning of over-heated prose for satiric purposes. Phew, so glad I didn't waste \$10.95 after all.

__label__1 Awful beyond belief!: I feel I have to write to keep others from wasting their money. This book seems to have been written by a 7th grader with poor grammatical skills for her age! As another reviewer points out, there is a misspelling on the cover, and I believe there is at least one per chapter. For example, it was mentioned twice that she had a "lean" on her house. I was so distracted by the poor writing and weak plot, that I decided to read with a pencil in hand to mark all of the horrible grammar and spelling. Please don't waste your money. I too, believe that the good reviews must have been written by the author's relatives. I will not put much faith in the reviews from now on!

__label__1 Don't try to fool us with fake reviews.: It's glaringly obvious that all of the glowing reviews have been written by the same person, perhaps the author herself. They all have the same misspellings and poor sentence structure that is featured in the book. Who made Veronica Haddon think she is an author?

__label__2 A romantic zen baseball comedy: When you hear folks say that they don't make 'em like that anymore, they might be talking about "BY THE SEA". This is a very cool story about a young Cuban girl searching for identity who stumbles into a coastal resort kitchen gig with a zen motorcycle maintenance man, three hysterical Italian chefs and a Latino fireballing right handed pitcher who plays on the team sponsored by the resort's owner. As is often the case she 'finds' herself through honest, often comical but always emotional, interaction with this sizzling roster of players. With the perfect mix of special effects, that salsa sound and flashbacks, BY THE SEA, gets 4 BIG stars from me!

__label__2 Fashionable Compression Stockings!: After I had a DVT my doctor required me to wear compression stockings. I wore ugly white TED hose and yucky thick brown stockings. Then I found Jobst UltraSheer. They gave me the compression I needed (15-20,) but looked like regular pantyhose. Even though my blood clot has been gone for 4 years, I still buy these to wear as support stockings because they make my legs feel so nice.**Note, I have problems with the rubberized tops rolling down my thigh. I tried the Jobst adhesive, but I hated having my skin pulled all day. I bought an inexpensive garter belt and it works fine and helps keep the stockings from rolling.

__label__2 Jobst UltraSheer Thigh High: Excellent product. However, they are very difficult to get on for older people. I feel like I've had a full day workout after getting them on. Also, as the day wears on, they begin to roll down from the top and create a very deep ridge in the skin. I have to wear them, so if those two difficulties could be addressed it would be such a help.

__label__1 sizes recommended in the size chart are not real: sizes are much smaller than what is recommended in the chart. I tried to put it and sheer it!. I guess you should not buy this item in the internet..it is better to go to the store and check it

__label__1 mens ultrasheer: This model may be ok for sedentary types, but I'm active and get around alot in my job - consistently found these stockings rolled up down by my ankles! Not Good!! Solution: go with the standard compression stocking, 20-30, stock #114622. Excellent support, stays up and gives me what I need. Both pair of these also tore as I struggled to pull them up all the time. Good riddance/bad investment!

__label__2 Delicious cookie mix: I thought it was funny that I bought this product without knowing it was a mix. I read the header very quickly and just thought it was packaged cookies. But no, it is cookie MIX and I guess I should have noticed that since it is right in the title.This is the first time I have ever tried baking with a cookie mix. If you are used to the convenience of the cookie dough that you buy wrapped up in plastic logs then you might be in for a bit of a surprise. Mixing up the dough can get VERY messy (it is extremely sticky). However, with a cookie mix like this you have a lot of flexibility in the ratio of ingredients (I like to add some extra butter to make the baked cookies more chewy). Also, this mix has really large chocolate chips in it--I love that.I removed a star for the addition of 'natural flavors' in the mix.

__label__1 Another Abysmal Digital Copy: Rather than scratches and insect droppings, this one has random pixelations combined with muddy light and vague image resolution. Probably the cue should have been the packaging is straight out of your street corner bootleg dealer.If you've ever seen a reasonably good condition film copy, you know the defining visuals of his film are the crystal clear lighting contrasts in black and white. The surrounding countryside and 'old home' scenes are set with early morning ground mists or the haze of memory while the events on the bridge and in the water are bright, clear, and immediate.Here everything is dull, dark, and clouded. Or, if you remember the timbre and enunciation of Captain's commands, so are the visuals.After that, it is hard to believe this award winning, critically acclaimed film's best presentation is on YOUTUBE. Somewhere "out there" is a DVD that comes up to the standard of a 16mm public library reel.Just none of them appear to be on Amazon.

Untuk *dataset review Amazon* secara lengkap dapat dilihat di *link* berikut.

<https://www.kaggle.com/bittlingmayer/amazonreviews>

Lampiran 3

Dataset Review Lazada

bagus mantap dah sesuai pesanan	positive
Bagus, sesuai foto	positive
okkkkk mantaaaaaapppp ... goood	positive
bagus sesuai	positive
Mau tanya ini cicilnya pake apa ya,cc bkn?	negative
Apakah TV. Tsb. Suda ada anti gores..	positive
Pengirim barang tidak sesuai janji. Katanya express 1 hari. Dari hari jumat barang sampai hari rabu. SeMoga barangnya awet saja.	negative
Kualitas ok	positive
Bagaimana tv LED saya merek Sharp 24 hitam LC 24LE170i sekarang mati total baru 2 bln kemana saya harus menghubungi dan meminta bantuan mohon kebijakan dan solusinya ,, so,alnya masih garansi kepada pihak lazada saya mohon bantuannya terimakasih!!!	negative
Pengiriman super lama.. tapi datang juga sich	negative
ok barang sampai dengan Baik.. smoga terus di pertahankan...	positive
Kalau mau cicilan gmna cra nya??	positive
barang sudah sampai dan bagus thanks lazada	positive
Pengiriman cepat sipplahhh.....	positive
Sampe Sekarang masih Ok ne laptop...	positive
Laptop ini memiliki spesifikasi yang bagus dan harga yang terjangkau, namun pada tab spesifikasi utama produk itu emang intel core i5 atau salah buat? Di judul dikatakan core i7	negative
Barang sampai dengan aman. Kemampuan laptop sesuai dengan yang di iklankan. Hanya saja, laptop datang dengan DOS sehingga harus di install sendiri windows nya. Overall, it's worth the money.	positive
Barang sampai 2 hari dengan selamat dan sesuai deskripsi	positive
Harga sangat murah mantep banget	positive
Murah banget itu enga salah kasih harga tau2 nati salah kasi info barang nya	negative
Barang udah sampai tapi pas dinyalain g bisa. Gmn pertanggung jawaban lazada?	negative
Kapasitas HDD asli 1 TB, dan tanpa ada lecet di saat saya pertama kali membuka bungkus bubble nya	positive
barang sesuai dengan gambar	positive
HD diterima dg packing bubble bagus, hanya tidak tertulis, barang elektronik, atw keterangan lain, agar penanganan lebih hati2 dan istimewa, ketika paket dalam perjalanan. Finally, 4bintang lhah...	positive
Mantap gan Barang sampai dengan selamat , Packing rapih , pokok Mantap lah !	positive
Bagus ori kok cuma hanya aja 58 gb:(klo 63 gpp hilang 1 gb	positive
kapasitasy kurang bagusss aturany klw kosong g bnyakkk firusss tpi nieh bnyakkk firusss sangat kecewe	negative
Baguss bangett kurir nya jga ramah, cakep pula wkwk. Ini 64gb tp cmn 58 sekian gb tp gpp lah yah... Pokok nya recommended deh....	positive
produk dah sampai kemasan rapih pengiriman cepat barang bagus pembayaran cod tidak perlu diragukan lagi ðŸ‘ðŸ‘	positive

barangnya bagus banget.. sesuai dengan gambar dn deskripsi.. pengiriman cepat.. bagus dn recomended banget..	positive
packing tanpa bubble warp. fisik no minus. Cuma 58 GB, bukan 64. ya lumayan	positive
barang ok .. packing Rapii .. Kurir nya Ramahh .. Puass ..	positive
produk original mantap semoga awet	positive
baik. sesuai ekspektasi.	positive
barang asli dan original terimakasih maju terus toko jet store	positive
Flashdisk nya bagus, sesuai picture, barang cepat sampai, kurir ramah	positive
oke... mudah2an awet dan ga crash belum sempet d coba	positive
kapasitas cumA 58 gb ajaaaaa.... huft...	positive
katanya 64 GB nyatanya cuman 58 GB	positive
thks ya barang sdh sampai,smga cocok	positive
siipp...thank lazada	positive
udah di trima ...barang OK Moga awet...	positive
josss tenan	positive
pesanan sesuai dgn yg di gambarskrg sy lg coba smoga awet...	positive
Pengiriman tidak sesuai tanggal yang tertera	negative
sesuai... no tipu2	positive
barang ok packing rapih	positive
Barangnya sampai dengan cepat, bagus, terimakasih lazada.	positive
pengiriman cepat packirng rapih harga murah, free ongkir	positive
mantapp barang nya sesuai pesanan	positive
Jangan pernah beli disini..parah banget..barang belum sampe udh 2 mi ggu alesan nya gak jelas gangguan dll..	negative
hati hati ini tv ga bisa pake USB	negative
Memuaskan, produk sesuai deskripsi, packing bagus, dan penerimaan produk cepat untuk saya yg berada di jkt	positive
Kenapa chargernya tidak lengkap....?? G bisa langsung dipakai buat charger..., kita harus nyari dan beli lagi sambungan buat chargernya..., tolong d bantu	negative
Ini spek ram.nya 2GB atau 4GB	negative
GAN SAYA SARANKAN UNTUK LEBIH BERHATI-HATI DALAM MEMBELI, FLASHDISK YANG DIKIRIM KE SAYA ADALAH FLASHDISK REKONDISI (DAUR ULANG BEKAS). JANGAN TERTIPU OLEH HARGA. MAS DOSA MAS BOHONG. DI FLASHDISKNYA ADA TULISAN NAMA "ARYO" YANG SENGAJA DIHAPUS TP KURANG BERSIH. FILE YANG DIKIRIM KE FLASHDISK JUGA ERROR. MAKAN DUIT HARAM MAS. KENYANG TP GA BERKAH.	negative
Barang cepat sampainya. Kualitas FD kurang bagus, file yang di copy sering corrupt.	negative
hardisknya keren, pelindungnya kokoh , udah ada aplikasi transcend elitnya di dalam untuk enkripsi thanks seller and lazada	positive
Sampainya agak lama tapi puas, packing rapi dan pas dateng warnanya biru seneng banget gan hehe. Top deh semoga awet barangnya	positive
Trims Lazada & reseller, , barang sdh saya terima senin siang, penigiriman sesuai dengan yg di informasikan.	positive
sesuai spek yg saya mau.. jd oke lah	positive
barangnya bagus. rapi tidak mengecewakan. akan tetapi bagus kalau ada Bonus pouch nya. hehhe	positive
barang udah sampe gunn,, masih segell dari pabriknya,, mangstabb,,	positive

barang tlah diterima kondisi baik	positive
Barang sesuai deskripsi,penjual ramah,respon cepat..	positive
Packaging rapi, flashdisknya ORI, kapasitasnya sesuai dengan detail produk yang dicantumkan dan foto dengan barangnya sesuai + barangnya datang sesuai tepat waktu. Makasih ya.. ðŸ˜~	positive
flashdisknya sudah diterima dengan baik. kapasitas total 58,544,586,752 bytes dengan FAT32. menjadi 62,885,724,160 bytes kalau diformat dengan exFAT. write speed 10MB/s dan read 30 MB/s dengan port USB 2.0, belum dicoba dengan port usb 3.0.	positive
Terima kasih, barang bagus dan respon penjual sgt baik.	positive
Barang sudah diterima dgn baik. Packing rapih. Tp isinya cuma 58GB. Selebihnya, oke pas cocok buat anda semua. bgus. sesuai deskripsi.	positive
Sesuai deskripsi	positive
barang dah sampai, makasih	positive
mantap barang gak minus sama sekali	positive
pengiriman cepat, packing cukup bagus	positive
Biarkan bintang yang berbicara ... ðŸ˜•	positive
Barangnya ok, cepat sampai	positive
Pengirimannya cepat, produk baik	positive
Barang sesuai dengan deskripsi dan berfungsi baik.	positive
sampai cepat, barang sesuai.	positive
dikirim hari berikutnya dan barang mulus	positive
Cuma butuh waktu 12 jam unit tiba.. thks	positive
Mau cicilan gmn caranya6	negative
Nyicil dengan cara mudah prosesnya susah nggak	negative
Pokonya puas..	positive

Untuk *dataset review* Lazada secara lengkap dapat dilihat di *link* berikut.

<https://www.kaggle.com/grikomsn/lazada-indonesian-reviews#20191002-reviews.csv>