**PROPOSAL PROYEK**

**Sentimen Analisis Terhadap Review Kepuasan Pelanggan Shopee Berbahasa Inggris Menggunakan Algoritma SVM (Support Vector Machine)**



**Disusun oleh:**

|  |  |
| --- | --- |
| 12S17011 | Astri Monica Sianturi |
| 12S17013 | Mega Sari Pasaribu |
| 12S17046 | Pebri Sangmajadi Sinaga |

**11S4037 - PEMROSESAN BAHASA ALAMI**

**PROGRAM STUDI SARJANA SISTEM INFORMASI**

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO**

**INSTITUT TEKNOLOGI DEL**

**NOVEMBER 2020**

# 

# **Daftar Isi**

[Daftar Isi i](#_Toc56972667)

[Daftar Gambar ii](#_Toc56972668)

[Daftar Tabel iii](#_Toc56972669)

[1. Pendahuluan 1](#_Toc56972670)

[1.1. Latar Belakang 1](#_Toc56972671)

[1.2. Tujuan 2](#_Toc56972672)

[1.3. Manfaat 2](#_Toc56972673)

[1.4. Ruang Lingkup 2](#_Toc56972674)

[2. Isi 3](#_Toc56972675)

[2.1 Tahap Pemrosesan Bahasa Alami 3](#_Toc56972676)

[3. Rencana Kerja 6](#_Toc56972677)

[3.1 Jadwal Kegiatan 6](#_Toc56972678)

[3.2 Pembagian Tugas 8](#_Toc56972679)

[Daftar Pustaka 9](#_Toc56972680)

# 

# 

# 

# 

# **Daftar Gambar**

[Gambar 1. Tahapan pengerjaan proyek analisis sentiment 3](#_Toc56972702)

# **Daftar Tabel**

[Table 1. Jadwal Kegiatan 6](#_Toc56972812)

[Table 2. Pembagian Tugas dalam Kelompok 8](#_Toc56972813)

# 

# **1.** **Pendahuluan**

Bab ini berisi penjelasan mengenai latar belakang pengerjaan proyek, tujuan, manfaat, dan ruang lingkup yang ingin dicapai dalam proyek.

## **1.1. Latar Belakang**

Kebutuhan masyarakat semakin meningkat seiring dengan perkembangan teknologi yang mempengaruhi proses transaksi jual beli dari konvensional berubah menjadi modern dengan pemanfaatan internet. Perubahan transaksi dengan memanfaatkan internet disebut dengan *e-commerce*. *E-commerce* memungkinkan pengguna untuk melakukan pembelian barang dimanapun dan kapanpun. Perkembangan *e-commerce* yang sangat cepat menjadi peluang bagi produsen untuk memasarkan dan mempromosikan produk kepada konsumen. Saat ini telah banyak *e-commerce* yang beredar di kalangan masyarakat. Salah satu *e-commerce* yang sering digunakan masyarakat adalah Shopee.

Shopee adalah salah satu platform e-commerce yang terkemuka di Asia Tenggara dan Taiwan. Shopee diluncurkan pada tahun 2015. Shopee merupakan *platform* yang memberikan pelanggan mengenai pengalaman belanja online yang mudah, aman, dan cepat disertai metode pembayaran yang baik. Pada Shopee terdapat banyak produk yang tersedia, misalnya makanan, minuman, barang-barang elektronik, produk kesehatan dan kecantikan, pakaian, dan lain-lain. Shopee juga memungkinkan pengguna untuk melakukan pengisian pulsa, pembelian paket data, pembelian token listrik, dan lain-lain.

Shopee menyediakan banyak fitur kepada pengguna. Salah satu fitur yang diberikan adalah *review* terhadap transaksi yang dilakukan. Pada Shopee terdapat *review* dari pelanggan yang telah membeli suatu produk untuk memberikan komentarnya terhadap produk tersebut. Dengan adanya *review* akan menjadi masukan terhadap pihak Shopee serta para pembeli yang ingin membeli produk tersebut sehingga membantu pengambilan keputusan untuk membeli barang berdasarkan *review* para pengguna. Pada *review* juga terdapat banyak kata-kata yang menjadi masukan untuk pihak Shopee atau penjual barang agar dapat meningkatkan kualitas barang berdasarkan review para pengguna. Namun, banyaknya review yang ada terkadang pembeli malas untuk membaca satu per satu kolom ulasan tersebut dan hanya melihat komen-komen teratas dengan berbagai rating yang tinggi. Oleh karena itu, *review* dari pelanggan harus diolah menjadi sebuah informasi untuk memudahkan calon pembeli dan pihak Shopee/penjual dalam mengambil keputusan dengan cara melakukan analisis sentimen.

*Sentiment analysis* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung di dalam suatu kalimat opini [1]. Pada proyek ini, tim ingin membagi suatu opini ke dalam opini positif dan opini negatif pada *review* barang di Shopee. Penelitian ini bertujuan untuk menghasilkan informasi sentimen mengenai *review* pengguna terhadap suatu barang yang dipesan di Shopee yang mengarah ke sentimen positif dan negatif menggunakan algoritma *Support Vector Machine* yang merupakan salah satu teknik pembelajaran dari *Text Mining*.

## **1.2. Tujuan**

Adapun tujuan dari pengerjaan proyek dengan teknik *text* *mining* ini adalah untuk melakukan analisis sentimen terhadap *review* produk yang diberikan pelanggan Shopee yang dapat memudahkan pihak Shopee/penjual dan pembeli dalam mengetahui kepuasan pelanggan terhadap suatu produk yang dapat dimanfaatkan dalam mengambil keputusan.

## **1.3. Manfaat**

Adapun manfaat yang diperoleh dari pengerjaan proyek sentimen analisis terhadap *review* kepuasaan pelanggan Shopee yang ditujukan untuk:

1. Calon Pembeli
2. Mengetahui ulasan pelanggan terhadap produk.
3. Membantu pengambilan keputusan apakah pembelian produk akan dilakukan berdasarkan *review* dari pelanggan.
4. Pihak Shopee/Penjual
5. Mengetahui kepuasan pelanggan terhadap produk yang ditawarkan.
6. Membantu dalam meningkatkan kualitas barang berdasarkan *review* pelanggan

## **1.4. Ruang Lingkup**

Adapun batasan dari pengerjaan proyek sentimen analisis terhadap *review* kepuasaan pelanggan Shopee adalah sebagai berikut.

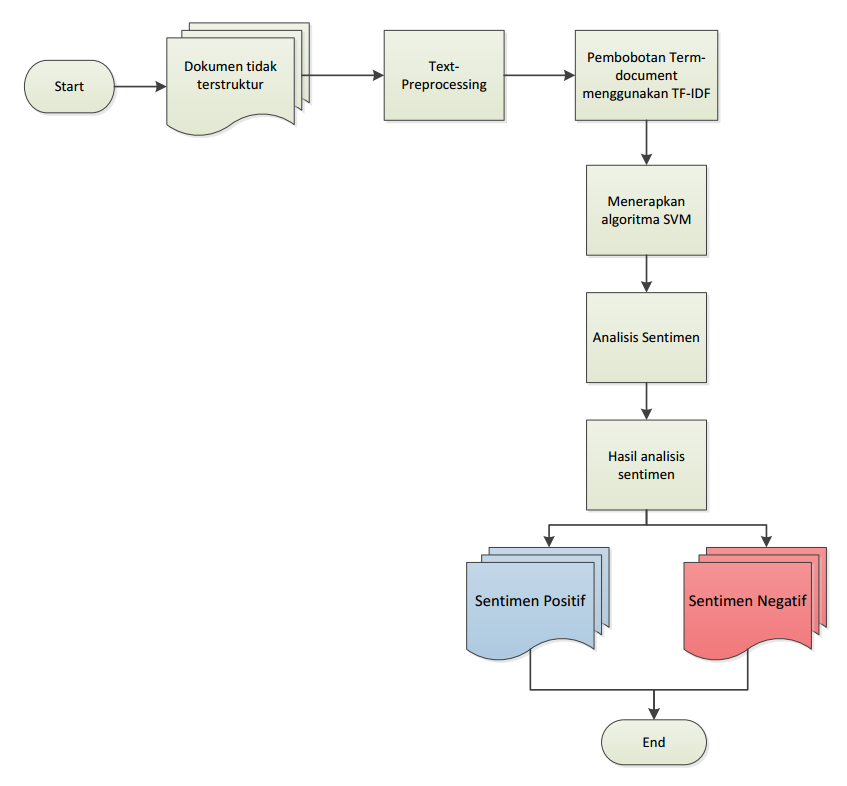
1. Algoritma yang digunakan dalam melakukan analisis sentimen adalah *Support Vector Machine* (SVM).
2. Literatur yang digunakan sebagai ruang lingkup pengerjaan proyek ini adalah *review* produk pelanggan Shopee berbahasa Inggris.
3. Data yang digunakan adalah data *review* dari pelanggan Shopee terhadap suatu barang yang dapat diakses di *website* Kaggle.com.

# **2. Isi**

Bab ini berisi penjelasan mengenai tahapan aktivitas pengembangan dasbor.

## **2.1 Tahap Pemrosesan Bahasa Alami**

Tahapan pengerjaan proyek analisis sentimen yang akan dilakukan dengan teknik *text* *mining* untuk menganalisis sentimen terhadap *review* produk menggunakan algoritma *Support Vector Machine* (SVM) adalah sebagai berikut:



**Gambar 1. Tahapan pengerjaan proyek analisis sentimen**

1. **Membaca Dokumen tidak Terstruktur**

*Unstructured Documents* merupakan langkah awal dalam desain yang digunakan untuk membaca keseluruhan *input document* yang terdapat dalam dataset *review* produk pelanggan Shopee. *Dataset* yang terdapat dalam *review* produk pelanggan Shopee masih merupakan *dataset* yang terdapat kalimat atau kata yang belum terstruktur dengan benar sehingga masih perlu dilakukan pada teks pra pemrosesanpada *dataset* tersebut.

1. **Teknik Teks Pra Pemrosesan**

Pra pemrosesan teks merupakan tahap yang dilakukan untuk mengolah teks sebelum digunakan dalam proses selanjutnya. Pra Pemrosesan teks dilakukan untuk menghilangkan data yang tidak dibutuhkan ataupun data yang terdapat dalam teks yang tidak sesuai dengan proses yang dibutuhkan. Penerapan pra pemrosesan teks ini akan menggunakan bahasa pemrograman *Python*. Beberapa tahapan untuk melakukan pra pemrosesan teks adalah *case folding*, *tokenization, stopwords removal*, *stemming*, dan *tagging*.

Berikut ini adalah tahapan proses penerapan *text mining* menurut Feldman dan Sanger (2007) [2]:

1. *Case folding* yaitu tahap untuk mengkonversi keseluruhan teks huruf kapital (*uppercase*) pada sebuah kalimat menjadi huruf kecil (*lowercase*) serta menghilangkan seluruh karakter yang dianggap tidak valid seperti angka, tanda baca, dan *Uniform Resources Locator* (URL). Pengimplementasian *case folding* tidak menggunakan library khusus pada Python namun menggunakan modul yang disediakan oleh Python itu sendiri.
2. *Tokenizing* yaitu proses memecah sebuah kalimat dalam dokumen berdasarkan tiap kata yang menyusunnya, atau dapat dikatakan memecah kalimat ke dalam satuan kata yang disebut dengan token. Cara untuk membedakan kata tersebut yaitu dengan menggunakan pemisah (delimiter) seperti spasi, enter, tabulasi, petik tunggal (‘), titik (.), semikolon (;), titik dua (:) dan lain sebagainya. Pengimplementasian *tokenizing* dapat dilakukan dengan menerapkan library NLTK.
3. *Stemming* yaitu merubah berbagai kata berimbuhan menjadi kata dasar. Stemming bertujuan untuk menghilangkan imbuhan-imbuhan seperti awalan kata (*prefixes*), sisipan kata (*infixes*), akhiran kata (*suffixes*) serta awalan dan akhiran kata (*confixes*) pada kata turunan yang terdapat dalam kata. Tahap ini pada umumnya dilakukan untuk teks dengan bahasa Inggris, karena teks dengan bahasa Inggris memiliki struktur imbuhan yang tetap.
4. *Stopwords removal* bertujuan untuk mengurangi kata yang kurang penting, seperti kata yang, di, ke, dalam, dan, dengan, ini, itu, untuk, dan lain sebagainya, sehingga mempermudah dan mempercepat pengolahan dokumen.
5. *Tagging* yaitu merubah berbagai kata dalam bentuk lampau menjadi kata awalnya, tahap ini pada umumnya dilakukan untuk teks dengan bahasa Inggris atau bahasa lainnya yang memiliki bentuk lampau.

**3. Term-document Matrix Using TF-IDF**

Pengukuran tingkat *similarity* antara 2 *object* atau *document* harus dimodelkan dalam bentuk tertentu. Pada representasi dokumen *bag of words*, setiap kata yang terdapat dalam dokumen akan dihitung bobotnya. Cara termudah yang sering digunakan adalah dengan menghitung frekuensi kemunculan kata atau *term* dalam sebuah dokumen serta menggunakannya sebagai bobot dalam *term* tersebut. Data yang berhasil melewati tahap *preprocessing* harus berbentuk angka atau numerik. Pengubahan data menjadi numerik dilakukan dengan menggunakan algoritma pembobotan *Terms Frequency - Inverse Document Frequency* (TF- IDF) yang merupakan algoritma yang membahas tentang *Term Frequency* dengan *Inverse Document Frequency*. [3]

**4. Support Vector Machine (SVM) Algorithm**

Algoritma yang digunakan dalam menganalisis sentimen yaitu dengan menggunakan *Support Vector Machine* (SVM). Metode *Support Vector Machine* (SVM) adalah metode klasifikasi *linier* dengan menemukan *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada *input space* (Saifinnuha, 2015). Prinsip dasar dari SVM adalah pengklasifikasi *linear*, kemudian dikembangkan menjadi pengklasifikasi non*linear* dengan memasukkan *kernel* trik pada ruang dimensi tinggi [4]. Hasil dari *preprocessing* tersebut akan dilakukan klasifikasi dengan *Support Vector Machine*. Nilai bias yang didapat pada proses ini akan digunakan untuk analisis sentimen.

**5. Analisis Sentimen**

Data *review* yang akan digunakan didapatkan dari situs *kaggle*.*com* kemudian disalin ke sebuah file .csv. Apabila nilai hasil perhitungan analisis sentimen bernilai lebih dari 0, maka data uji tersebut akan digolongkan menjadi komentar positif, jika nilai hasil perhitungan analisis sentimen bernilai kurang dari 0, maka data uji tersebut akan digolongkan menjadi komentar negatif.

# **3. Rencana Kerja**

Bab ini berisi penjelasan mengenai jadwal kegiatan dan pembagian tugas dalam pengerjaan proyek.

## **3.1 Jadwal Kegiatan**

Berikut adalah rencana jadwal pengerjaan proyek analisis sentimen terhadap *review* pelanggan Shopee.

**Table 1. Jadwal Kegiatan**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No.** | **Jenis Kegiatan** | **November 2020, Minggu ke-** | | | | **Desember 2020, Minggu ke-** | | | | **Januari 2021, Minggu ke-** | | | |
| **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** | **1** | **2** | **3** | **4** |
| 1 | Identifikasi masalah |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Studi Literatur |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Pengajuan Topik |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Review Topik |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Analisis dan desain |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Implementasi |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Evaluasi hasil |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | Penulisan Laporan Akhir |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 | Presentasi |  |  |  |  |  |  |  |  |  |  |  |  |

1. Identifikasi Masalah

Pada tahap ini tim akan mengidentifikasi apa yang menjadi latar belakang permasalahan sehingga perlu untuk melakukan analisis sentimen pada review produk pelanggan Shopee. Adapun latar belakang permasalah ini adalah pentingnya pihak Shopee/Penjual dan pembeli mengetahui pendapat para pelanggan terhadap produk sudah dibeli untuk dapat dimanfaatkan sebagai pertimbangan dalam mengambil keputusan.

1. Studi Literatur

Pada tahap ini tim akan mempelajari terkait studi literatur terkait dimulai dari memahami data yang akan digunakan dalam analisis sentimen, mempelajari materi atau topik yang berkaitan dengan proses analisis sentimen dan proses atau teknik yang digunakan.

1. Pengajuan Topik

Pada tahapan ini, tim akan menyusun proposal sesuai dengan masalah yang telah diidentifikasi sebelumnya. Penyusunan proposal didasarkan pada pemahaman literatur yang berkaitan dengan visualisasi data.

1. Review Topik

Pada tahap ini, tim dan dosen pengampu mata kuliah akan me-*review* topik yang telah diajukan oleh tim dalam proposal. Jika *scope* pengerjaan proyek telah memenuhi maka proyek akan dikerjakan oleh tim dalam rentang waktu yang ditentukan.

1. Analisis Proyek dan Perencanaan Teknik Analisis Sentimen

Pada tahapan ini tim akan menganalisis dan mendesain model atau teknik pengerjaan perencanaan teknik analisis sentimen yang akan dikembangkan. Pengerjaan proyek diawali dengan analisis permasalahan dan pencarian data set yang akan diuji. Untuk melakukan penerapan teknik analisis sentimen maka tim proyek akan memilih algoritma yang cocok. Tim akan menentukan algoritma dan melakukan pemahaman lebih lanjut terkait cara kerja algoritma dan bagaimana pengaplikasiannya.

1. Implementasi Teknik Analisis Sentimen

Pada tahap ini, tim akan mengimplementasikan teknik analisis sentimen yang sudah didesain pada tahap sebelumnya menggunakan bahasa pemrograman Python. Tim akan mengerjakan tahapan-tahapan tersebut untuk menghasilkan sebuah dokumen *review* yang dibagi berdasarkan sentimen positif dan sentimen negatif atau.

1. Evaluasi Hasil

Pada tahap ini, tim akan melakukan evaluasi terhadap hasil analisis sentimen yang telah dilakukan. Evaluasi dilakukan untuk menentukan apakah hasil analisis sentimen sesuai dengan rancangan dan tujuan awal yang telah ditentukan sebelumnya.

1. Pengerjaan Laporan Akhir

Setelah hasil analisis sentimen yang diterapkan dievaluasi, maka tim akan menuliskan laporan akhir dari proyek yang telah selesai diimplementasi. Pada laporan akan terdapat kesimpulan dari hasil analisis sentimen di review kinerja Shopee. Pada laporan juga akan disertakan hasil analisis sentimen yang dikembangkan.

1. Presentasi

Pada tahap ini, tim akan melakukan presentasi hasil akhir dari proyek analisis sentimen *review* produk pelanggan Shopee. Presentasi akan disajikan dalam bentuk video presentasi.

## **3.2 Pembagian Tugas**

Dalam pengerjaan proyek, kelompok yang terdiri atas 3 orang akan mengerjakan tugas masing-masing untuk menyelesaikan proyek analisis sentimen terhadap *review* pelanggan Shopee. Berikut adalah tabel pembagian tugas dalam kelompok.

**Table 2. Pembagian Tugas dalam Kelompok**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Kegiatan** | **Penanggung Jawab** | **Luaran** |
| 1 | Penyusunan Proposal | Astri, Mega, Pebri | Proposal |
| 2 | Analisis Proyek dan Perencanaan Teknik Analisis Sentimen | Astri, Mega, Pebri | Desain Visualisasi |
| 3 | Implementasi Teknik Analisis Sentimen | Astri, Mega, Pebri |  |
| 4 | Penulisan Laporan Akhir | Astri, Mega, Pebri | Laporan Akhir |

# 

# 

# **Daftar Pustaka**

|  |  |
| --- | --- |
| [1] | A. S. A. F. Alvi Pranandha Syah, "ANALISIS SENTIMEN PADA DATA ULASAN PRODUK TOKO ONLINE DENGAN METODE MAXIMUM ENTROPY," *e-Proceeding of Engineering,* vol. IV, p. 4633, 2017. |
| [2] | S. F. d., Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data, New York: Feldman dan Sanger, 2007. |
| [3] | D. a. J. H. M. Jurafsky, Speech and Language Processing, Prentice Hall, 2008. |
| [4] | L. M. M. A. F. Dimas Joko Haryanto, "Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode Support Vector Machine Dan Query Expansion," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN,* vol. II, pp. 2909-2916, 2018. |