

## KLASIFIKASI TEKS *REVIEW* PADA *E-COMMERCE* TOKOPEDIA MENGGUNAKAN ALGORITMA SVM

Ipan Saepul Milal<sup>1</sup>, M. Hasanudin<sup>2</sup>, M. Aliffiallathifa Nur Azhari<sup>3</sup>, Rifki Aditya Nugraha<sup>4</sup>, Nova Agustina<sup>5</sup>,  
Sri Erina Damayanti<sup>6</sup>  
Program Studi Teknik Informatika<sup>1, 2, 3, 4, 5, 6</sup>  
Sekolah Tinggi Teknologi Bandung<sup>1, 2, 3, 4, 5, 6</sup>  
ipanmarqalenta93@gmail.com<sup>1</sup>, hasan080801@gmail.com<sup>2</sup>, mochalif222@gmail.com<sup>3</sup>, rifkiadityan9@gmail.com<sup>4</sup>,  
nova@sttbandung.ac.id<sup>5</sup>, srierina@sttbandung.ac.id<sup>6</sup>

### Abstrak

Perkembangan teknologi saat ini berdampak bagus kepada bidang ekonomi, salah satunya adalah dalam berbelanja. Banyak *platform E-commerce* yang dibuat untuk memudahkan konsumen dalam berbelanja secara *online*. Dengan adanya *review* produk akan mempermudah pemberian informasi kepada pembeli sehingga memberikan kepercayaan terhadap produk atau jasa yang akan digunakan dan memberikan evaluasi bagi para penjual. Ketika penjualan produk meningkat dan *review* terhadap produk dari konsumen juga makin banyak, penjual mengalami kesulitan dalam segi memilah *review* mana yang dapat dikatakan positif atau negatif berdasarkan komentar yang diberikan konsumen. Salah satu metode untuk algoritma untuk melakukan analisis berupa teks pada *review* produk yaitu melalui analisis sentimen. Data yang didapat dari hasil *review* konsumen diproses dengan *text mining*, selanjutnya melakukan klasifikasi data *review* ke dalam enam kelas yaitu positif kuat, positif, cenderung positif, cenderung negatif, negatif dan negatif kuat. Dalam penelitian ini proses klasifikasi menggunakan algoritma *Support Vector Machine (SVM)* yang mampu menghasilkan model klasifikasi yang baik melalui hasil data *training* meskipun yang digunakan relatif sedikit. Hasil penelitian menyatakan bahwa model klasifikasi *review* produk *handphone* dengan model SVM memiliki nilai akurasi 75%. Pada penelitian ini *tools* yang digunakan berupa *Google Colab* dan *Visual Studio Code*, dengan hasil pengklasifikasian dilakukan *deploy* pada tahap *code* akhir setelah *testing* data dan dibuat tampilan *page* berupa website agar pengguna lebih mudah untuk melakukan cek input dalam melihat hasil klasifikasinya secara langsung berdasarkan text ulasan pada tokopedia produk *handphone*. Kata kunci : *Deploy Model, E-Commerce, Review, SVM*.

### Abstract

*Current technological developments have a good impact on the economic sector, one of which is in shopping. Many E-commerce platforms have been created to make it easier for consumers to shop online. With product reviews it will make it easier to provide information to buyers, so giving confidence in the products or services that will be used, and provide evaluations for sellers. When product sales increase and product reviews from consumers also increase, sellers experience difficulties in sorting out which reviews can be said to be positive or negative based on comments given by consumers. One method for analyzing product reviews is sentiment analysis. The data obtained from the results of consumer reviews is processed by text mining, then classifies the review data into six classes, namely strong positive, positive, tends to be positive, tends to be negative, negative, strong negative. In this study the classification process uses the Support Vector Machine (SVM) algorithm which is capable of producing a good classification model through the results of training data although relatively few are used. The results of the study stated that the Hp product review classification model with the SVM model has an accuracy value of 75%. In this research the tools used were Google Colab and Visual Studio Code, with the results of the classification being deploy in the final step after testing data and created a page view in the form of website so that it is easier for the user to do check input to see directly the result of classification based on the review text on tokopedia mobile phone products.*

*Keyword: Deploy Model, E-Commerce, Review, SVM.*

### I. PENDAHULUAN

Perkembangan teknologi saat ini memberikan banyak kemudahan termasuk dalam berbelanja. Saat ini banyak *E-Commerce* yang menyediakan *platform* untuk berbelanja secara *online*, sehingga konsumen tidak perlu beranjak dari rumah untuk membeli sesuatu yang dibutuhkannya. Bukan hanya konsumen yang berbelanja, penjual juga mendapatkan keuntungan dari adanya *platform* belanja *online* ini, mudah menemukan konsumen hanya dengan bermodalkan konten yang menarik untuk membuat konsumen tertarik membeli produk yang dijual. *Platform E-Commerce* memiliki sisi negatifnya di mana konsumen dan pedagang tidak dapat bertransaksi secara *face to face*, konsumen tidak dapat secara langsung melihat produk yang hendak dibeli begitu juga penjual tidak bisa mengetahui secara detail keinginan konsumen [1]. Solusi dari masalah yang dihadapi dapat di selesaikan dengan adanya *review* produk yang dilakukan oleh konsumen kepada penjual. *Review* produk bertujuan untuk memberikan informasi kepada pembeli, memberikan kepercayaan terhadap produk atau jasa yang akan digunakan dan memberikan evaluasi bagi para penjual [2]. Ketika penjualan produk meningkat dan *review* terhadap produk dari konsumen juga makin banyak, penjual mengalami kesulitan dalam segi memilah *review* mana yang dapat dikatakan positif atau negatif berdasarkan komentar yang diberikan konsumen.

Banyaknya penilaian/*review* yang tersusun secara acak menyebabkan kesulitan bagi penjual untuk mengetahui penilaian positif atau negatif. Hal ini dapat di selesaikan dengan memanfaatkan cara mengklasifikasi *review* konsumen sehingga diperoleh informasi hasil *review* dengan mudah. Salah satu metode untuk menganalisis *review* produk yaitu

analisis sentimen. Data yang didapat dari hasil *review* konsumen diproses dengan *text mining*, selanjutnya melakukan klasifikasi data *review* ke dalam lima kelas yaitu positif kuat, positif, cenderung positif, negatif kuat, negatif, cenderung negatif. Dalam penelitian ini proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM)

## II. TINJAUAN PUSTAKA

Pada bagian ini akan dibahas teori yang berhubungan dengan penelitian yang kami lakukan, di antaranya sebagai berikut:

### 1. Klasifikasi

Menurut KBBI klasifikasi adalah penyusunan atau menggolongkan menurut jenis, kaidah atau standar yang telah ditetapkan [3]. Klasifikasi merupakan proses menggabungkan kumpulan model untuk membedakan label kelas, tujuan akhir dari proses ini untuk memprediksi kelas dari suatu objek dengan tepat. Data yang akan di klasifikasi harus data berupa *supervised learning* atau sederhananya data yang berlabel dan termasuk ke dalam model yang prediktif (*output* berupa nilai variabel target) [4]. Dalam penelitian kami klasifikasi digunakan untuk mengelompokan *review e-commerce* dari pembeli, di mana *output* yang dihasilkan berupa sentimen sangat positif, positif, negatif dan sangat negatif.

### 2. Support Vector Machine

Salah satu algoritma klasifikasi teks yaitu *Support Vector Machine* (SVM). SVM pertama dikenalkan pada tahun 1992 oleh *Valdimir Vapnik* bersama rekannya *Bernard Boser* dan *Isabelle Guyon* [5][2]. SVM bekerja dengan mencari *hyperline* terbaik untuk membagi kelas pada ruang *input*, dari penelitian [6] mengatakan bahwa algoritma SVM memiliki nilai AUC lebih baik daripada *naive bayes* [6]. SVM memiliki fungsi kernel yang membuat algoritma ini istimewa, dengan sistem kernel ini SVM dapat menerapkan pemisah *input non-linear* berdimensi tinggi. Kernel yang sering digunakan di antaranya *Linear*, *Polynomial* dan *Radial Basis Function*(RBF) [4]. *Hyperline* adalah sebuah fungsi untuk memisahkan antar kelas [2]. Rumus dalam menentukan *hyperline*:

$$w \cdot x + b = 0 \rightarrow w_0 + w_1 x_1 + w_2 x_2 \quad (1)$$

Dimana:

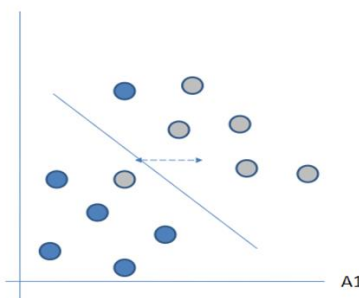
$W = \{w_1, w_2, \dots, w_n\} \rightarrow$  *vector* bobot

$n$  = jumlah atribut

$b$  = skalar yang digunakan sebagai bias

$x$  = nilai atribut

Pada kasus data yang terpisah secara linear, persamaan yang dipakai yaitu menggunakan SVM linear. Persamaan ini memungkinkan optimasi terpenuhi dengan menggunakan teknik *soft margin*. Pada gambar di bawah merupakan sebaran data yang tidak terpisah secara linear.



Gambar 1. Sebaran Data yang Tidak Terpisah Secara Linear

Dengan penjelasan di atas, persamaan linear dirumuskan ulang menggunakan teknik *soft margin*:

$$y_i(x_i \cdot w + b) \geq 1 - \varepsilon_i \quad (2)$$

$$\min_w \tau(w) = \frac{1}{2\|w\|^2} + C \sum_{i=1}^n \varepsilon_i \quad (3)$$

Dimana:

$W$  = *vector* bobot

$C$  = parameter

$\varepsilon$  = nilai *error*

### 3. Review / Penilaian Barang

Kata *review* artinya penilaian atau ulasan terhadap kualitas suatu produk atau jasa yang didasari oleh pengalaman nyata pengguna. Teknis dalam memberikan ulasan suatu produk atau jasa yang telah digunakan yaitu dengan pengguna memberikan penilaian berdasarkan tingkat kepuasan pengguna. Tujuan dari *review* untuk memberikan informasi kepada pembeli, memberikan kepercayaan terhadap produk atau jasa yang akan digunakan dan memberikan evaluasi bagi para penjual [7].

#### 4. *E-Commerce*

*Electronic commerce* merupakan kepanjangan dari *E-Commerce*. Mulai dibuat pada tahun 1970-an, diperkenalkan bersamaan dengan layanan pengiriman uang secara elektronik. Seiring perkembangan internet pada tahun 1990 perdagangan elektronik ini kian membesar di seluruh dunia. *E-commerce* adalah layanan transaksi melalui jaringan internet. *E-commerce* tidak selalu berhubungan dengan aktivitas jual-beli, aktivitas lain seperti pengecekan ketersediaan barang di gudang, pengambilan uang di mesin ATM dll. [8]

#### 5. *Text Mining*

Salah satu bidang yang sering dipakai dalam *data mining* yaitu *text mining*. Bidang ini merupakan suatu proses untuk menggali informasi dari sekumpulan dokumen menggunakan aplikasi analisis. *Text mining* dapat dilakukan menggunakan komponen klasifikasi, dimana *text* dapat dikategorikan (*text categorization*) dan di kelompokkan (*text clustering*)[4]. Salah satu contoh yang diaplikasikan menggunakan *text mining* yaitu pada *customer service* perusahaan yang berhubungan langsung dengan konsumen, dengan demikian mereka memahami respon pelanggan terhadap perusahaan mulai dari panggilan, telepon, *chat*, *review*, misalnya dengan penggunaan teknologi *chat bot* atau pesan otomatis dengan mengidentifikasi pertanyaan yang paling sering ditanyakan dengan menyediakan jawabannya secara langsung[9].

#### 6. Analisis Sentimen

Analisis sentimen adalah bidang studi yang menganalisis opini, evaluasi, perilaku dan emosi yang ditunjukkan kepada entitas tertentu, seperti instansi, produk dan sebagainya [10]. Analisis sentimen memiliki tujuan untuk menganalisis data teks, dapat berupa pendapat, evaluasi, penilaian dan sentimen seseorang terhadap suatu barang, orang dan organisasi. Analisis sentimen merupakan bagian ilmu dari *data mining* [6]. Analisis sentimen biasanya digunakan untuk pelabelan teks kedalam tiga kelas sentimen (sentimen positif, negatif dan netral), suatu ulasan bisa di anggap pada kelas sentimen positif jika mengandung pernyataan pujian, ungkapan terima kasih atau ulasan positif untuk suatu produk[11].

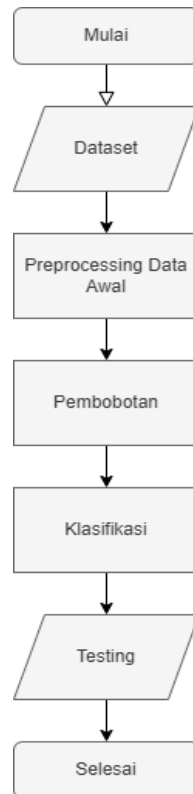
#### 7. Bahasa Pemrograman *Python*

Salah satu bahasa pemrograman yang mudah dipelajari adalah bahasa pemrograman *python*. Karena bahasa pemrograman ini memiliki struktur bahasa yang sederhana namun memiliki struktur data tingkat tinggi yang efisien serta efektif untuk pemrograman berorientasi objek. Bahasa pemrograman *python* memiliki sifat yang diinterpretasikan sehingga ideal dalam pengembangan aplikasi secara cepat di sebagian banyak *platform*. *Interpreter Python* dapat diimplementasikan dalam C dan C++ serta mudah dikembangkan dengan fungsi dan tipe data baru (bisa dipanggil dari C) [12].

### III. ANALISIS DAN PERANCANGAN

#### 1. Metode Penelitian

Terdapat beberapa tahapan yang dilakukan pada penelitian. Alur tahapan dapat dilihat pada gambar *flowchart* di bawah ini:



Gambar 2. Tahapan Penelitian

a. Pengambilan *Dataset*

Data pada penelitian diambil dari *website zenodo.org* [13], data yang diambil merupakan kumpulan teks *review* produk dari salah satu *E-Commerce* (Tokopedia). Data penelitian terdiri dari dua jenis data, yaitu *data tagging review* dan *data tagging category*. Terkait *data tagging review* memuat 9 atribut yang berisi (teks ulasan, bintang, *positif\_kuat*, *positif*, *cenderung\_positif*, *cenderung\_negatif*, *negatif*, *negatif\_kuat* dan *check\_sum*). Sedangkan pada *data tagging category* memiliki dua *field* atau kolom dengan atribut *text*, dan *category*. Diantara keduanya terdapat kegunaan pada masing-masing data, model dilatih melalui data tersebut sebagai *data testing* dan *data training*

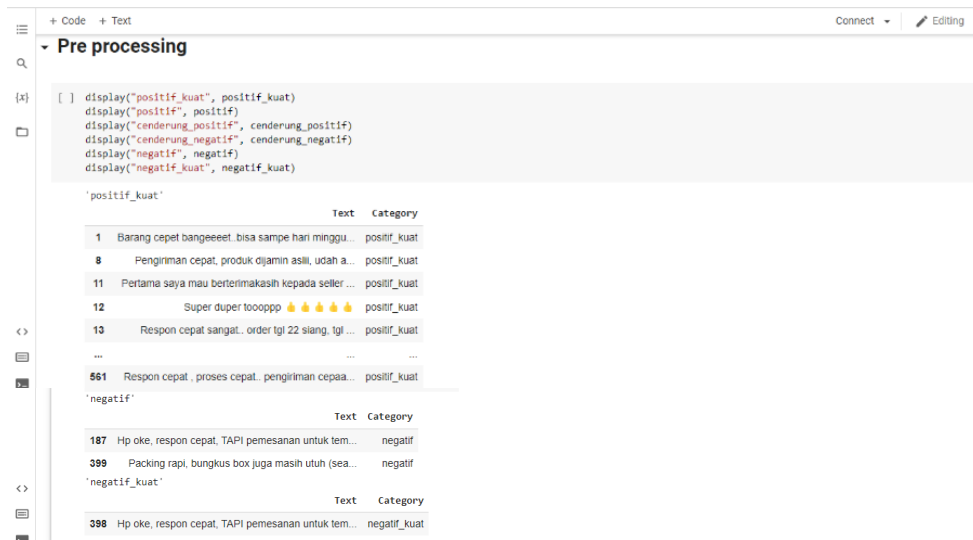
b. Pengolahan Data Awal (*Preprocessing*)

Pada tahapan ini, data yang masih memiliki struktur sembarangan atau tidak beraturan dilakukan terlebih dahulu pembersihan data. Proses membersihkan data bertujuan untuk menghilangkan *noise* dan mengubah data menjadi lebih terstruktur. Tahapan paling awal sebelum dilakukannya *data mining* karena banyak diantara beberapa data memiliki permasalahan yang bisa mengganggu saat pemrosesan berlangsung. Terdiri dari beberapa tahapan yang secara singkat mampu membuat setiap ulasan menjadi lebih *clean* seperti menghilangkan tanda baca, simbol dan kata yang tidak perlu.

Tahapan *Preprocessing* yaitu sebagai berikut:

a. Klasifikasi *Category*

Pada proses ini dilakukan klasifikasi dari seluruh data yang ditampilkan dalam bentuk tabel sesuai enam jenis kategori, yaitu *positif kuat*, *positif*, *cenderung positif*, *negatif kuat*, *negatif* dan *cenderung negatif*. Maka akan didapatkan lebih mudah jumlah ulasan yang termasuk dalam setiap jenis kategori yang tersedia. Klasifikasi bertujuan mempermudah pembagian data agar dapat diambil sesuai jenis kategori tanpa melakukan klasifikasi pencarian *manual* dengan jumlah data yang banyak, serta berguna mengidentifikasi tahapan paling awal kata yang muncul disetiap jenis kategori.



Gambar 3. Klasifikasi Jenis Kategori

b. Data *Cleaning* dan Normalisasi

Data *cleaning* dan normalisasi ini terdapat beberapa tahapan. Penjelasan tahapannya data *cleaning* dilakukan untuk melakukan proses pada data dengan menghilangkan data duplikat serta pada normalisasi yang akan berguna untuk menghapus *text* dengan kata yang tidak diperlukan. Beberapa tahapan untuk proses data *cleaning* dan normalisasi dengan hasilnya, diantaranya :

a. Data *cleaning*

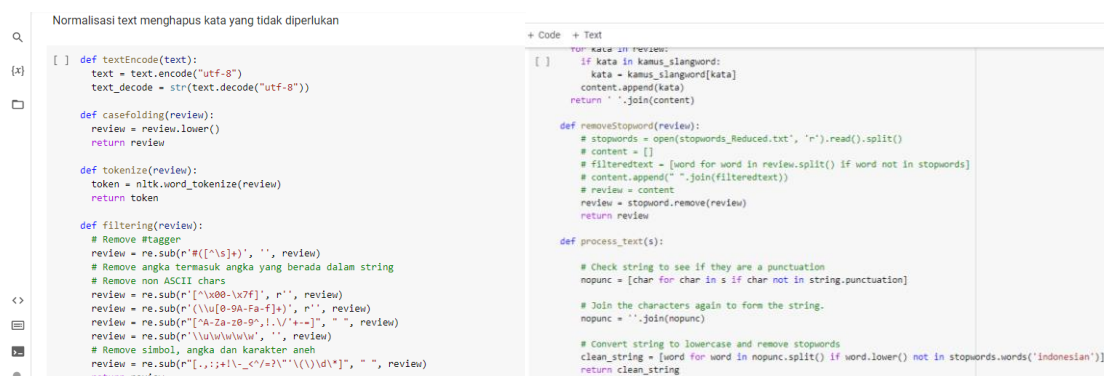
Hasil yang ditampilkan pada data *cleaning* dengan hasil sebagai berikut.



Gambar 4. Data *Cleaning*

b. Filter *Stopword*

Tahapan yang dilakukan untuk menghapus kata-kata yang tidak relevan yang dihasilkan dari *dictionary slangword* sehingga hasil dari *output* teks mampu berkaitan dan memiliki arti yang jelas dengan klasifikasi sentimen.



Gambar 5. Filter *Stopword*

c. *Tokenize*

Tahapan yang digunakan untuk melakukan proses menghitung token pada setiap kata *dataset* yang telah dinormalisasi sebelumnya. Proses menghilangkan tanda baca, simbol, ataupun yang bukan huruf,

sehingga mampu menyajikan sekumpulan kata menjadi kalimat yang jelas dan utuh dipahami. Proses yang terjadi :

```

Tahap pertama menghitung token setiap kata pada dataset yang sudah dinormalisasi.

[ ] # test = data.iloc[0, data.columns.get_loc('text')]
    # # test = 'km apa khabar'
    normalize = []
    positif_kuat = []
    positif = []
    cenderung_positif = []
    cenderung_negatif = []
    negatif = []
    negatif_kuat = []

    for index, row in data.iterrows():
        test = row['Text']
        test = casefolding(test)
        # print(test)
        test = filtering(test)
        # print(test)
        test = replaceThreeOrMore(test)
        # print(test)
        test = removeDoubleSpaces(test)
        # print(test)
        test = convertToSlangword(test)
        # print(test)
        test = removeStopword(test)

+ Code + Text
[ ] test = removeStopword(test)
    # print(test)
    test = tokenize(test)
    # print(test)
    normalize += test
    if row['Category'] == "positif_kuat":
        positif_kuat += test
    elif row['Category'] == "positif":
        positif += test
    elif row['Category'] == "cenderung_positif":
        cenderung_positif += test
    elif row['Category'] == "cenderung_negatif":
        cenderung_negatif += test
    elif row['Category'] == "negatif":
        negatif += test
    else:
        negatif_kuat += test

    # print(fake)
    print("NORMALIZE TOKEN", len(normalize))
    #remove Duplicates
    normalize_clean = list(dict.fromkeys(normalize))
    print("NORMALIZE REMOVE DUPS", len(normalize_clean))

NORMALIZE TOKEN 5969
NORMALIZE REMOVE DUPS 1859
    
```

Gambar 6. Tokenize Process

Hasil dari proses tersebut dilanjutkan dengan menghitung jumlah perbandingan pada data sebelum dan sesudah dilakukan normalisasi yang disajikan melalui bentuk grafik.

Proses dan hasil setelah dilakukan normalisasi :



Gambar 7. Proses dan Hasil Setelah Dinormalisasi

Proses dan hasil sebelum dilakukan normalisasi :

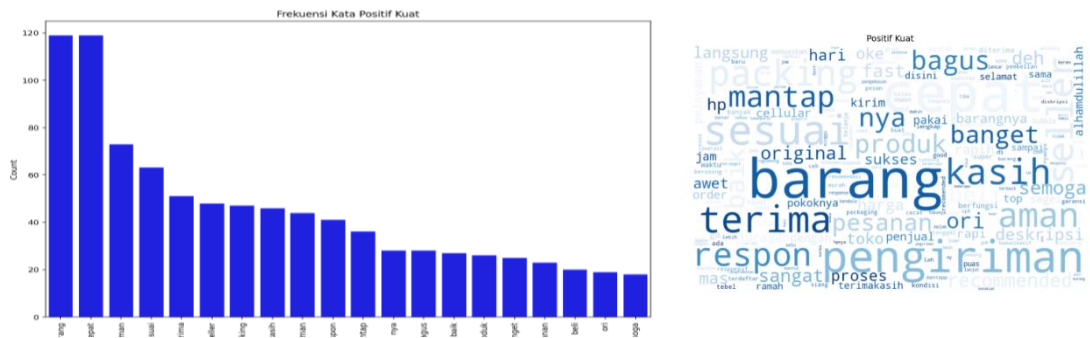


Gambar 8. Proses dan Hasil Sebelum Dinormalisasi

c. Menghitung kata yang sering muncul

Pada bagian ini dilakukan perhitungan pada setiap kata yang sering muncul disetiap kategori ulasan dengan menghasilkan grafik frekuensi pada ulasan positif atau negatif.

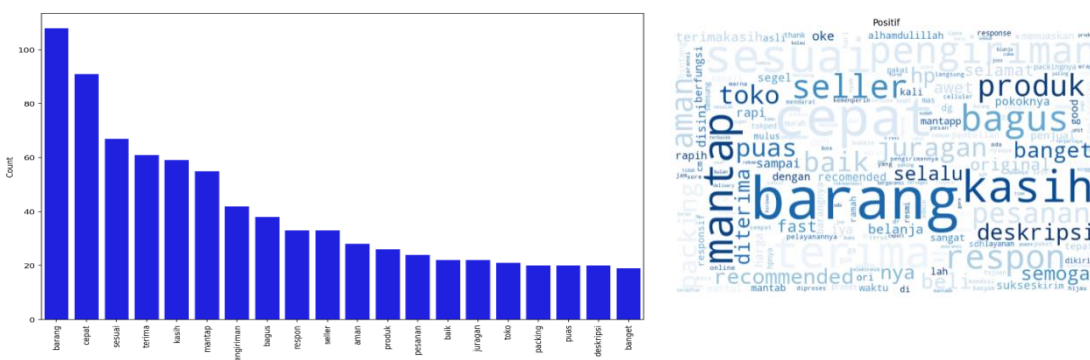
a. Positif Kuat



Gambar 9. Frekuensi Kata Positif Kuat

Dari hasil gambar 9 dapat diambil kesimpulan, bahwa kata yang sering muncul pada kategori positif kuat adalah barang, cepat dengan jumlah terbanyak yaitu 120, pengiriman, sesuai, terima, seller, packing, respon, mantap bagus, baik, produk, banget, pesanan, beli, ori, dan semoga. Dan terlihat satu kata yang pada wordcloud di positif kuat ini didukung dengan kata ‘sangat’, ‘banget’, jika dilihat dari KBBI (Kamus Besar Bahasa Indonesia) kedua kata tersebut menunjukkan hal yang lebih. Dapat dikatakan kata yang muncul pada positif kuat ini akan selalu muncul pada kalimat atau ulasan yang menunjukkan rasa senang.

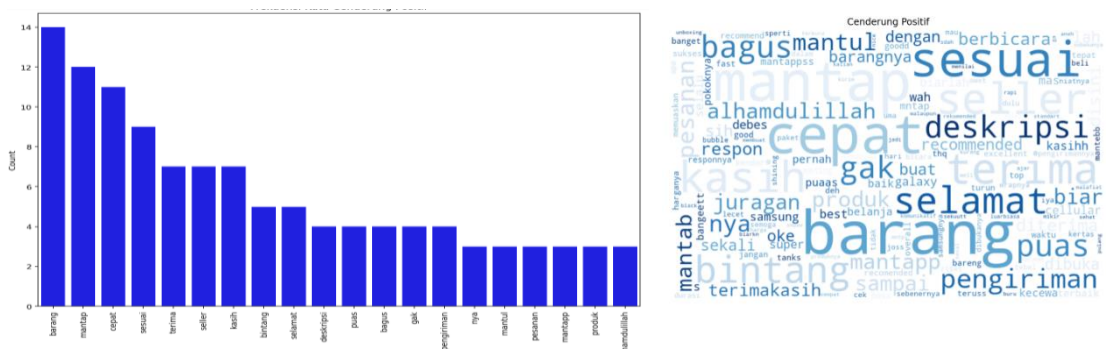
b. Positif



Gambar 10. Frekuensi Kata Positif

Dari hasil gambar 10 maka dapat dijelaskan frekuensi kata yang sering muncul pada jenis kategori positif adalah barang dengan jumlah paling tinggi diatas 100, cepat, sesuai, terima, kasih, mantap, pengiriman, bagus, respon, seller, aman, produk, pesanan, baik, juragan, toko, packing puas, deskripsi, banget. Jika dilihat dari gambar sebelumnya di gambar 9 banyak persamaan dapat dilihat sekilas kepuasan akan produk nya memberikan respon positif, hanya saja dengan tingkat kepuasan sedikit dibawah positif kuat.

c. Cenderung positif

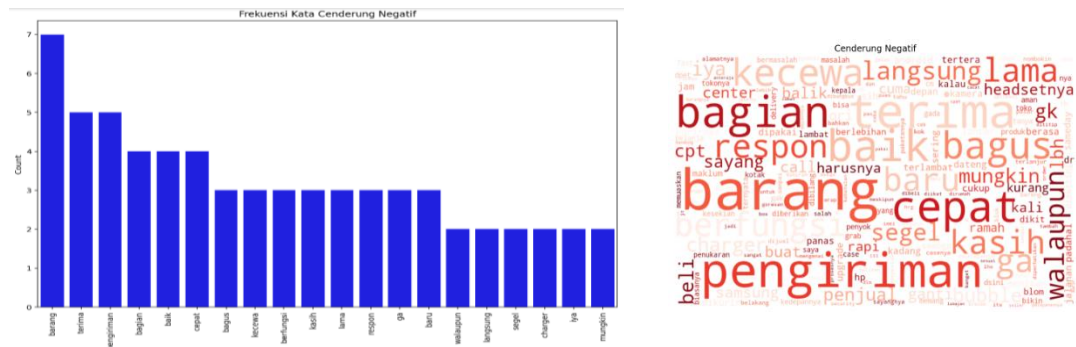


Gambar 11. Frekuensi Kata Cenderung Positif

Cenderung positif ini menunjukkan tingkat kepuasan yang standar atau biasa saja, dapat dilihat pada gambar frekuensi kata gambar 11 hampir sama dari kata yang ada di positif sebelumnya, akan tetapi mengingat standar kepuasan ulasan cenderung positif ini pada tingkat standar atau bisa, maka nilai frekuensi kata yang digunakan juga menjadi lebih menurun jumlahnya dengan kata paling tinggi ‘barang’ sebanyak 14.

Dapat diambil kesimpulan dari ketiga jenis kategori positif (kuat, positif, dan cenderung) memiliki satu persamaan, yaitu kata ‘bagus’, ‘sesuai’, ‘barang’, selalu muncul pada wordcloud dengan teks berukuran besar yang artinya memiliki frekuensi kata terbanyak.

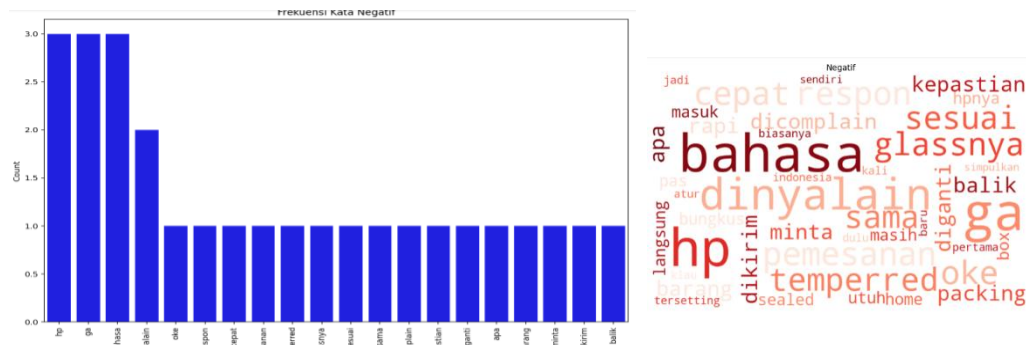
d. Cenderung Negatif



Gambar 12. Frekuensi Kata Cenderung Negatif

Hasil dari frekuensi pada gambar 12 di atas memiliki nilai tertinggi 7 pada kata ‘barang’, jika dilihat kembali kata pada cenderung negatif ini ada kesamaan dengan ulasan positif, dengan perbedaan jumlah kata ‘baik’ yang lebih sedikit sebanyak 4, dan kata-kata lain seperti ‘kecewa’, ‘lama’, dan ‘walaupun’ yang menunjukkan rasa tidak puas terhadap seller dengan produk yang dia beli tidak seperti yang diharapkan, dan muncul kata seperti ‘ga’ atau ‘gk’ artinya ‘tidak’ menunjukkan penolakan.

e. Negatif



Gambar 13. Frekuensi Kata Negatif

Dapat dilihat dari data frekuensi negatif pada gambar 13 jumlah frekuensi kata yang ada menjadi lebih berkurang menjadi nilai tertinggi 3.0 di kata ‘hp’, ‘ga’, ‘bahasa’. Sebetulnya ada beberapa kata seperti ‘oke’, ‘cepat’, dan ‘sesuai’, tetapi jumlah frekuensi yang sangat kecil 1.0. Yang dapat menunjukkan kata ketidakpuasan pembeli yaitu kata ‘dikomplain’, ‘diganti’, dan kata ‘minta’, ‘dikirim’, ‘balik’, seperti satu kalimat utuh “minta dikirim balik” artinya pelanggan ingin *cancel* batal atau meminta uang kembali.



```
[x] ▾ Algoritma SVM
[ ] #Classifier - Algorithm - SVM
# fit the training dataset on the classifier
SVM = svm.SVC(C=1.0, kernel='linear', degree=3, gamma='auto')
SVM.fit(Train_X_Count,Train_Y)
# predict the labels on validation dataset
predictions_SVM = SVM.predict(Test_X_Count)
# Use accuracy_score function to get the accuracy
print("SVM Accuracy Score -> ",accuracy_score(predictions_SVM, Test_Y)*100)

SVM Accuracy Score -> 75.40983606557377
```

Gambar 16. Running model algoritma SVM

e. *Testing*

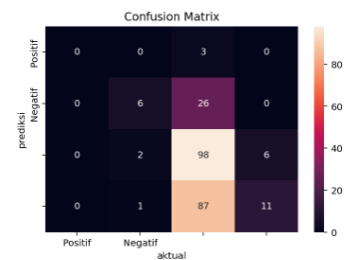
Selanjutnya tahapan pengujian dari hasil klasifikasi. Pengujian dilakukan dengan *Confusion Matrix* dan *Test 1 Data* dari *dataset*. Berikut *testing* yang dilakukan:

a. *Confusion Matrix*

```
[x] ✓ 1s [ ] from sklearn.metrics import confusion_matrix
cm = confusion_matrix(status_test, predictions)
ax = plt.subplot()
sns.heatmap(cm, annot=True, ax=ax)

ax.set_xlabel('aktual')
ax.set_ylabel('prediksi')
ax.set_title('Confusion Matrix')
ax.xaxis.set_ticklabels(['Positif', 'Negatif'])
ax.yaxis.set_ticklabels(['Positif', 'Negatif'])
print('Confusion Matrix: \n', confusion_matrix(status_test,predictions))
```

```
Confusion Matrix:
[[ 0  0  3  0]
 [ 0  6 26  0]
 [ 0  2 98  6]
 [ 0  1 87 11]]
```



Gambar 17. Confusion Matrix

b. *Test 1 Data*

*Test 1 data* adalah proses yang dilakukan sebagai uji coba pada model menggunakan satu buah teks atau kalimat ulasan untuk mencoba model apakah sudah berjalan sesuai dan hasil yang ditampilkan pada proses ini yaitu jenis kategori ulasan tersebut dari 6 jenis kategori (positif kuat, positif, cenderung positif, cenderung negatif, negatif, negatif kuat).

▾ **Test 1 Data**

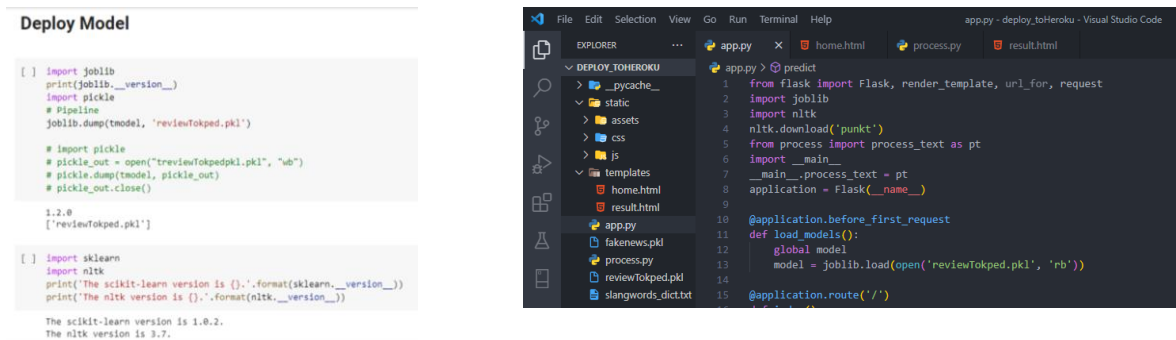
```
[ ] test = news_test.iloc[0]
test = 'Barang sesuai deskripsi, ORI asli buatan Samsung Indonesia. Senin malam pesen, rabu siang tiba di rumah..'
pre = tmodel.predict([test])
pre

array(['positif_kuat'], dtype=object)
```

Gambar 18. Test 1 Data

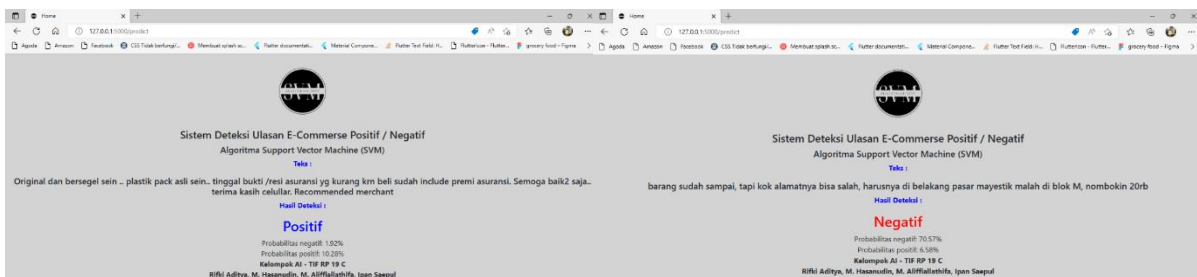
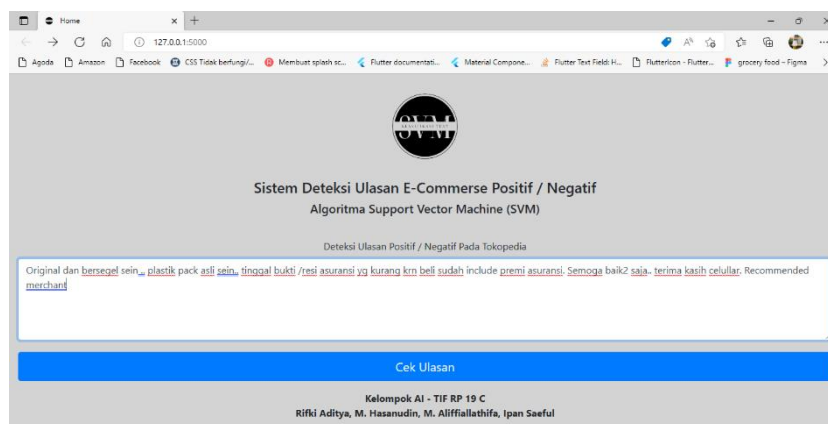
f. *Deploy Program*

*Deployment model* adalah proses membuat model agar tersedia pada lingkungan produksi dengan model yang dapat memberikan prediksi ke sistem perangkat lunak yang lain menjadikan model tersebut mempunyai peran dengan penerapan secara efektif pada sistem. Hasil *deploy model* pada *google colab* berbasis pickle, dan *VSCode* untuk pembuatan model pada tampilan *website*, sebagai berikut :



Gambar 19. Deploy Program

Hasil tampilan halaman website :



Gambar 20. Tampilan Website

#### IV. KESIMPULAN

Berdasarkan hasil analisis penelitian dapat disimpulkan, Algoritma *Support Vector Machine* (SVM) masih relevan untuk digunakan sebagai pengklasifikasian *Text Review* dengan menentukan ulasan yang termasuk kelas positif atau negatif. Terdapat 6 tahapan yang digunakan pada penelitian ini yaitu *preprocessing*, pembobotan, klasifikasi dan *testing*. Model klasifikasi dengan algoritma *Support Vector Machine* (SVM) ini memiliki nilai akurasi yang cukup baik. Hasil pengujian mendapat persentase 75% untuk akurasi model klasifikasi SVM ini. Hasil klasifikasi juga dapatkan nilai *precision* paling tinggi pada kelas cenderung positif dengan nilai 1.00, nilai *recall - f1score - support* paling tinggi ada pada kelas yang sama yaitu positif dengan nilai masing masing 0,76 – 0,57 – 106. Sistem prediksi di *deploy* berbasis *website*, sebelum di implementasikan model terlebih dahulu di rubah ke format *.pkl* (*pickle*). Tujuan *deployment model* agar pengguna lebih mudah dalam melihat hasil klasifikasi berdasarkan *text review* berupa komentar yang diberikan konsumen. Dalam proses *deploy* penulis menggunakan *code editor Visual Studio Code*, didukung dengan *framework Python, Flask* dan desain tampilan menggunakan *CSS Bootstrap*. Hasil dari klasifikasi pada *website* menampilkan persentase probabilitas negatif dan probabilitas positif.

Adapun saran dari penulis untuk penelitian ini, diharapkan penelitian ini dapat dikembangkan dengan algoritma lainnya, seperti *Naive Bayes*, *KNN*, *Decision Tree*, dan algoritma klasifikasi lainnya, dengan tujuan meningkatkan persentase akurasi dalam klasifikasi

## REFERENSI

- [1] (diakses pada 2023) website accurate. [online]. Tersedia pada: <https://accurate.id/bisnis-ukm/apa-itu-e-commerce>
- [2] Nova Agustin, "Support Vector Machine", Matakuliah Artificial Intelligence., Sekolah Tinggi Teknologi Bandung., 2022.
- [3] (2023) pada website KBBI. [Online]. Tersedia pada: <https://kbbi.web.id>.
- [4] E. Erlinda, Z. Bahri. "Klasifikasi Berita Menggunakan Support Vector Machine Untuk Mengetahui Jenis Konten E-Government". Repository Universitas Riau, pp. 1-14, Okt. 2020
- [5] F. Fahri Irfani, Mohamad Triyanto, Anggit Dwi Hartanto, Kusnawi, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine", Jurnal Bisnis, Manajemen dan Informatika., vol. 16, No. 3, hal 258-266, Feb. 2020
- [6] Fajar Sodik Pamungkas, Iqbal Kharisudin., "Analisis Sentimen dengan SVM, Naive Bayes dan KNN untuk studi tanggapan masyarakat Indonesia terhadap pandemi Covid-19 pada media sosial Twitter", Prosiding Seminar Nasional Matematika., PRISMA 4, hal. 628-634. 2021
- [7] (diakses pada 2023) sosialights website. [online]. Tersedia pada: <https://journal.socialights.id/menulis-review-produk/>
- [8] (diakses pada 2023) ditulis oleh Khairina F. Hidayanti pada website glints. [online]. Tersedia pada: <https://glints.com/id/lowongan/e-commerce-adalah/>
- [9] (diakses pada 2023) algorit.ma. [online]. Tersedia pada: <https://algorit.ma/blog/teknik-text-mining-2022/>
- [10] Yoga Tika Pratma. "Analisis Sentimen Opini Pelanggan Terhadap Aspek Pariwisata Pantai Malang Selatan Menggunakan TF-IDF dan Support Vector Machine". Univ Brawijaya, Fk. Ilmu Komputer. 2018
- [11] Ditia Yosmita Pratiwi. "Analisis Sentimen Online Review Pengguna E-Commerce Menggunakan Metode Support Vector Machine Dan Maximum Entropy", Univ. Islam Indonesia, Fk Matematika dan Ilmu Pengetahuan Alam, 2018
- [12] (diakses pada 2023) website python. [online]. Tersedia pada: <https://docs.python.org/id/3.10/tutorial/index.html>
- [13] "Dataset Review Konsumen HP untuk Analisis Kata pada Korpus Analisis Sentimen", zenodo.org, Maghfirani, Nida Inayah; Utomo, Girraz Karyo; Hadi, Rizal Maulana, Indonesia. (link: <https://zenodo.org/record/3833283#.Y9TAdnZBzrf>)