

Analisis Bigdata Sentimen Ulasan Konsumen Masker Kesehatan Pada Marketplace

Dody Indra Sumantiawan

Universitas Nasional Karangturi Semarang

Jl. Raden Patah No.182-192 Rejomulyo, Semarang Timur Kota Semarang

Email : dodyindrass@gmail.com

Abstract. Big data is a collection of data that has a large volume, so traditional data processing technology is unable to handle it well. Marketplace is a platform that most people often use to shop online. On this platform there is a comments column for writing reviews of products that have been purchased. Consumer reviews have an important role in understanding customer perceptions and sentiments towards the products being sold. Classification uses the Support Vector Machine method. The goal is to classify consumer reviews into positive or negative sentiment categories. The test uses data from more than 300 data samples with the assumption of independence between the features in the data. The results of big data analysis of consumer sentiment reviews of health masks on the marketplace used the support vector machine method with an accuracy value of 88%. The results of the analysis can be concluded that the dominant results of scraping reviews on health mask products lead more to positive reviews. The results on wordcloud of negative reviews provide insight to improve the quality of masks which are still lacking in terms of thinness, straps breaking easily, tears, holes, rubber quality and product packaging.

Keywords: Sentiment Analysis, Support Vector Machine, Classification

Abstrak. Bigdata merupakan sekumpulan data yang memiliki volume besar, sehingga teknologi pemrosesan data tradisional tidak mampu untuk menanganinya dengan baik. Marketplace merupakan platform yang sering digunakan kebanyakan orang untuk berbelanja online, didalam platform ini terdapat kolom komentar untuk menuliskan hasil ulasan produk yang telah dibeli. Ulasan konsumen memiliki peran penting dalam memahami persepsi dan sentimen pelanggan terhadap produk yang dijual. Klasifikasi menggunakan metode *Support Vector Machine*. Tujuan mengklasifikasi ulasan-ulasan konsumen menjadi kategori sentimen positif atau negatif. Pengujian menggunakan data lebih dari 300 sampel data dengan asumsi independensi antara fitur-fitur yang ada dalam data. Hasil analisis bigdata sentimen ulasan konsumen masker kesehatan pada marketplace menggunakan metode *support vector machine* dengan nilai akurasi sebesar 88%. Hasil analisis dapat disimpulkan dominannya hasil *scraping* ulasan pada produk masker kesehatan lebih banyak mengarah ke ulasan positif. Hasil pada *wordcloud* ulasan negatif memberikan insigt untuk memperbaiki kualitas masker yang masih kurang dalam segi tipis, tali mudah putus, sobek, bolong, kualitas karet dan pengemasan produk.

Kata kunci: Analisis Sentimen, *Support Vector Machine*, Klasifikasi

LATAR BELAKANG

Bigdata merupakan sekumpulan data yang memiliki volume besar, sehingga teknologi pemrosesan data tradisional tidak mampu untuk menanganinya dengan baik. Aspek terpenting dalam bigdata bukan sekedar terletak pada berapa besarnya data yang bisa disimpan dan diolah, tetapi kegunaan dan nilai tambah yang terkandung dalam data (D. I. Sumantiawan, 2023).

Berkembangnya era digital, semakin banyak pengguna internet yang menggunakan berbagai platform digital. Marketplace adalah salah satu platform yang sering digunakan kebanyakan orang untuk berbelanja online, didalam platform marketplace digital ini terdapat kolom komentar untuk menuliskan hasil ulasan produk yang telah dibeli dalam platform marketplace (Al-Smadi, M. et al., 2018). Ulasan konsumen memiliki peran penting dalam memahami persepsi dan sentiment pelanggan terhadap produk yang dijual. Oleh sebab itu,

Received: Januari 31, 2024, Accepted: Februari 27, 2024, Published: Februari 29, 2024

* Dody Indra Sumantiawan, dodyindrass@gmail.com

analisis sentimen menjadi penting untuk mengidentifikasi pandangan positif, negatif atau netral yang terkandung dalam ulasan yang dituliskan (Bin Alias, M.S. et al., 2021) (Baek, Y. et al., 2020).

Masker adalah salah satu produk yang sering dijual online dan banyak mendapatkan ulasan. Masker telah menjadi kebutuhan saat pandemic COVID-19, maka analisis sentimen produk masker ini dapat memberikan wawasan yang berharga bagi produsen, penjual dan pengguna dalam memahami kualitas, kinerja dan kepuasan pelanggan terhadap berbagai jenis masker yang tersedia diplatform marketplace.

Klasifikasi dalam penelitian ini akan digunakan untuk melakukan klasifikasi analisis sentimen ulasan produk masker yang dijual dalam aplikasi marketplace. Pendekatan metode menggunakan algoritma *Support Vector Machine*, yang merupakan salah satu metode klasifikasi yang populer dalam pemrosesan Natural Language Processing (NLP) (Borg, A. and Boldt, M., 2020). SVM memiliki kelebihan dalam menentukan jarak pemisah antar kelas dengan menggunakan support vector sehingga proses komputasi yang dilakukan pada data yang cukup besar bisa menjadi lebih cepat (Borg, A. and Boldt, M., 2020) (SUMANTIAWAN and et al, 2023).

Tujuan mengklasifikasi ulasan-ulasan konsumen menjadi kategori sentimen positif atau negatif. Hasil dari analisis ini dapat memberikan pemahaman yang lebih baik tentang preferensi dan kebutuhan terkait masker, serta memberikan informasi pada produsen dan penjual untuk meningkatkan kualitas produk masker (S. Burns,2018) (Bourequat, et al, 2021).

Dataset yang terkumpul kemudian melalui preprocessing data, proses selanjutnya transformasi data dilakukan dan membagi data menjadi data training yang akan masuk dalam proses melatih model support vector machine dan data test untuk menguji performa data. Selanjutnya melakukan evaluasi kinerja model dan menganalisis hasil yang diperoleh.

Kontribusi hasil yang diharapkan dalam penelitian ini adalah pemahaman sentimen konsumen terkait produk masker diaplikasi marketplace dan memberikan wawasan yang berguna bagi produsen dan penjual dalam meningkatkan kualitas produk.

KAJIAN TEORITIS

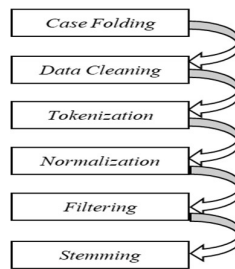
Analisis sentimen

Analisis sentimen merupakan studi komputasi yang menganalisis terkait opini/pendapat, sentimen, dan emosi yang diekspresikan melalui teks (Al-Smadi, M. et al., 2018). Penelitian terkait hal tersebut mulai populer pada tahun 2002 dan terus berkembang. Analisis sentimen memberikan luaran yaitu informasi yang dikategorikan menjadi nilai positif

dan negatif (Bin Alias, M.S. et al., 2021). Proses untuk mendapatkan informasi tersebut, analisis sentimen menciptakan sebuah sistem yang kemudian dapat melakukan klasifikasi terhadap teks dalam suatu dokumen. Pemanfaatan analisis sentimen dilakukan untuk memeriksa pendapat terhadap suatu produk atau suatu kejadian (Borg dan Boldt, 2020).

Text Mining

Text mining adalah proses mengubah teks tidak terstruktur menjadi format terstruktur untuk mengidentifikasi pola yang bermakna dan wawasan baru dengan menerapkan teknik analitik tingkat lanjut, seperti *Naïve Bayes*, *Support Vector Machines (SVM)* dan algoritma lainnya (D. I. Sumantiawan, et al., 2023). Teks adalah salah satu tipe data yang paling umum dalam database (Sumantiawan, D. I. 2024). Masalah yang paling menantang dalam *text mining* adalah kompleksitas dan ambiguitas bahasa manusia. Kata yang sama yang digunakan dalam konteks yang berbeda dalam dokumen yang sama akan memiliki arti yang berbeda karena interpretasi yang berbeda (Al-Smadi, M. et al., 2018). Tahapan dalam text mining dapat dilihat pada gambar 1.



Gambar 1 Tahapan Text Mining

TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan salah satu metode untuk melakukan pembobotan kata dari proses ekstraksi kata dengan menerapkan perhitungan kata umum di information retrieval. Metode pembobotan ini merupakan penggabungan antara *term frequency* dan *inverse document frequency*. *Term frequency* merupakan jumlah kemunculan sebuah *term* pada sebuah dokumen. Besarnya jumlah term yang muncul berbanding lurus dengan pembobotan yang diberikan. *Inverse document frequency* adalah proses untuk mengukur seberapa penting kata dalam suatu dokumen (Hunt, 2021).

Support Vector Machine (SVM)

Support vector machine termasuk dalam *supervised learning* yang berarti model atau mesin mempelajari terlebih dahulu untuk melakukan klasifikasi, dengan membagi data menjadi dataset yaitu data training dan data testing (Borg, A. and Boldt, M., 2020). Metode ini diperkenalkan pertama kali pada tahun 1992 oleh Vapnik. *Support Vector Machine* dapat

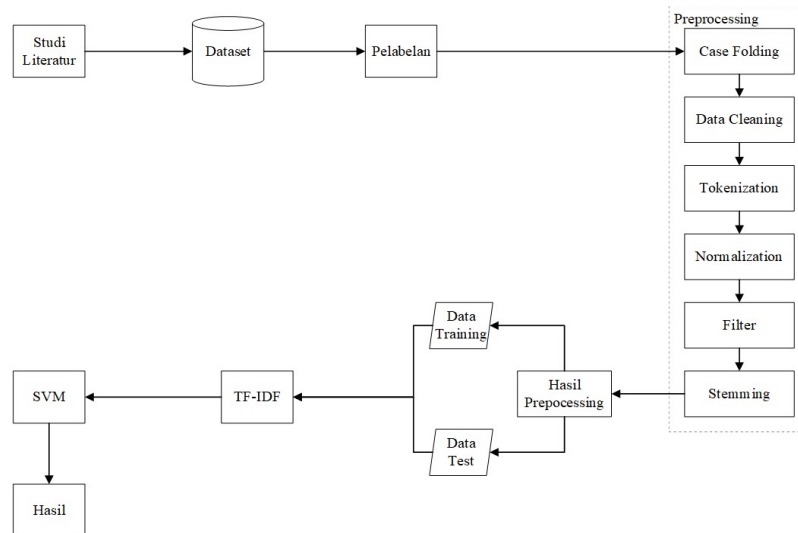
menemukan fungsi pemisah yang bisa memisahkan dua dataset dari dua kelas yang berbeda. *Support Vector Machine* mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas pada input space dengan memaksimalkan jarak antar kelas (Boureauqat, W. and Mourad, H., 2021).

SVM bertujuan untuk menggambar *hyperplane* (garis lurus dalam 2D, atau setara dengan dimensi lebih tinggi di ruang yang lebih kompleks) yang paling baik memisahkan setiap kategori. *Hyperplane* ini disebut batas keputusan. Margin mengacu pada jarak antara *hyperplane* dan titik data terdekat dari setiap kategori, yang juga dikenal sebagai vektor pendukung. Vektor pendukung ini sangat penting untuk mendefinisikan *hyperplane* dan pada dasarnya bertindak sebagai data pelatihan untuk SVM.

METODE PENELITIAN

Prosedur penelitian

Prosedur penelitian dalam penelitian ini meliputi tahapan studi literatur, *preprocessing*, pelabelan, NLP (*Natural Language Processing*), menghitung kata dengan TF-IDF, sentimen ulasan positif dan negatif, klasifikasi dengan SVM serta kesimpulan. Prosedur penelitian dapat dilihat pada Gambar 2.



Gambar 2 Prosedur Penelitian

Pelabelan

Dataset yang diambil dalam penelitian ini masih dalam keadaan data mentah yang belum memiliki kelas target, sehingga perlu dilakukan pelabelan pada dataset. Proses dalam tahap pelabelan ini dilakukan dengan fitur peringkat untuk menentukan pola data ulasan pada

sentimen positif atau negatif. Pengklasifikasian ini dilakukan pada ulasan yang memiliki rating 4 dan 5 sebagai sentimen positif dan rating 3 sampai 1 sebagai sentimen negatif.

NLP (*Natural Language Processing*)

Komputer menggunakan teknik-teknik seperti tokenisasi (pemisahan teks menjadi unit-unit yang lebih kecil seperti kata-kata), analisis morfologi (memahami struktur kata), analisis sintaksis (memahami struktur kalimat), analisis semantik (memahami makna teks), dan analisis pragmatik (memahami konteks dan tujuan teks).

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

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode yang digunakan dalam pemrosesan bahasa alami untuk memberikan bobot pada kata-kata dalam sebuah dokumen atau korpus teks. Metode ini mengukur seberapa penting sebuah kata dalam suatu dokumen atau korpus berdasarkan frekuensi kemunculan kata tersebut dalam dokumen tersebut serta sejauh mana kata tersebut umum atau jarang muncul dalam seluruh korpus.

Sentimen ulasan

Sentimen ulasan digunakan untuk melihat jumlah dari masing-masing ulasan positif dan negatif terhadap produk masker yang terjual, pada tahap ini bertujuan untuk melihat seberapa besar rasio ulasan yang diberikan oleh pelanggan setelah bertransaksi di *marketplace*.

HASIL DAN PEMBAHASAN

Hasil analisis bigdata sentimen ulasan konsumen produk masker menggunakan metode klasifikasi *support vector machine* dengan nilai akurasi yang bagus. Pengujian menggunakan data lebih dari 300 sampel data dengan asumsi independensi antara fitur-fitur yang ada dalam data.

Pelabelan

Pelabelan data dilakukan untuk membagi data menjadi data ulasan dengan label sentimen positif dengan angka 1 dan label sentimen negatif dengan angka 0. Kriteria pemberian label dilakukan dengan klasifikasi ulasan yang memiliki rating 4 dan 5 sebagai label ulasan sentimen positif dan rating 3 sampai 1 sebagai label ulasan sentimen negatif. Hasil pelabelan dapat dilihat pada gambar 3.

| Unnamed: 0 | Nama_Produk | Akun | Ulasan | rate | Ulasan_tokenize | Ulasan_normalized | Ulasan_stop_removed | Ulasan_Stemmed | Ulasan_clean | label |
|------------|---|-----------|---|------|---|---|---|---|---|-------|
| 0 | Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese... | Stephanus | dua kali beli sayang lambat respon order pagi ... | 3 | [yang, 'dua', 'kali', 'beli', 'sayang', 'lam... | [yang, 'dua', 'kali', 'beli', 'sayang', 'lam... | ['dua', 'kali', 'beli', 'sayang', 'lambat', 'r... | ['dua', 'kali', 'beli', 'sayang', 'lambat', 'r... | dua kali beli sayang lambat respon order pagi ... | 0 |
| 1 | Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese... | Yati | bagus kirim cepat | 5 | ['bagus', 'kirim', 'cepat'] | ['bagus', 'kirim', 'cepat'] | ['bagus', 'kirim', 'cepat'] | ['bagus', 'kirim', 'cepat'] | bagus kirim cepat | 1 |
| 2 | Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese... | Husin | barang sesuai pesan | 5 | ['barang', 'sesuai', 'pesan'] | ['barang', 'sesuai', 'pesan'] | ['barang', 'sesuai', 'pesan'] | ['barang', 'sesuai', 'pesan'] | barang sesuai pesan | 1 |
| 3 | Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese... | Febby | panas pakai | 2 | ['panas', 'pakai'] | ['panas', 'pakai'] | ['panas', 'pakai'] | ['panas', 'pakai'] | panas pakai | 0 |
| 4 | Masker KF94 4 Ply Isi 10 Pcs Masker Medis Kese... | Rahmat | barang sesuai minta bagus komunikasi jual beli... | 3 | ['barang', 'sesuai', 'minta', 'bagus', 'komuni... | ['barang', 'sesuai', 'minta', 'bagus', 'komuni... | ['barang', 'sesuai', 'minta', 'bagus', 'komuni... | ['barang', 'sesuai', 'minta', 'bagus', 'komuni... | NaN | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

Gambar 3 Hasil Pelabelan

TF-IDF (Term Frequency-Inverse Document Frequency)

Pemrosesan bahasa alami untuk melakukan pembobotan kata dalam sebuah dokumen atau korpus teks dalam penelitian ini menggunakan TF-IDF. Metode TF-IDF digunakan untuk mengukur seberapa penting sebuah kata dalam suatu dokumen atau korpus berdasarkan frekuensi kemunculan suatu kata dalam sebuah dokumen tersebut serta sejauh mana kata tersebut sering muncul atau jarang muncul dalam seluruh korpus. Hasil jumlah kata dapat dilihat pada gambar 4.

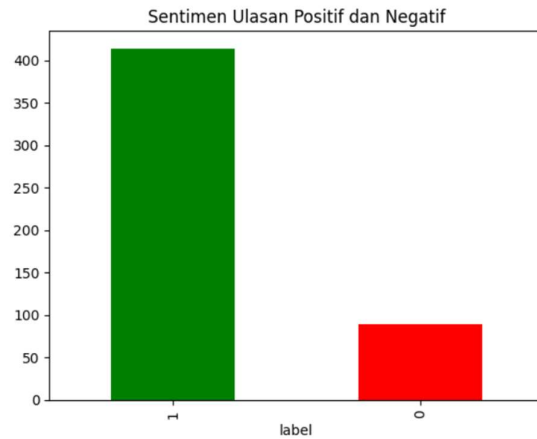
```
{'dua': 202, 'kirim': 341, 'putus': 594,
 'kali': 304, 'sore': 686, 'pendek': 542,
 'beli': 64, 'besok': 78, 'estimasi': 212,
 'sayang': 641, 'barang': 52, 'baik': 44,
 'lambat': 373, 'bagus': 36, 'karet': 309,
 'respon': 615, 'sesuai': 669, 'headloop': 255,
 'order': 507, 'harga': 249, 'istri': 274,
 'pagi': 520, 'cepat': 114, 'untung': 774,
 'kirim': 341, 'pesan': 563, 'anak': 21,
 'sore': 686, 'panas': 529, 'hrnga': 263,
 'besok': 78, 'pakai': 521, 'murah': 470,
 'barang': 52, 'tidak': 745, 'layan': 383,
 'bagus': 36, 'ada': 8, 'muas': 467,
 'sesuai': 669, 'komentar': 348, 'medis': 436,
 'harga': 249, 'masker': 431, 'lumayan': 401,
 'cepat': 114, 'pengiriman': 548, 'taju': 763,
 'pesan': 563, 'produk': 585, 'murahansmoga': 471, ....}
```

Gambar 4 Jumlah kata yang muncul

Preprocessing

Preprocessing dalam penelitian ini menggunakan NLP (natural language processing). Tokenisasi sebagai teknik untuk melakukan pemisahan teks menjadi unit-unit yang lebih kecil seperti kata-kata, memahami struktur kata (morfologi), sintaksi atau memahami struktur kalimat, semantik (memahami makna dari teks), dan analisis pragmatis untuk memahami konteks dan tujuan teks. Sehingga dapat dihasilkan jumlah antara ulasan positif dan ulasan

negatif yang diberikan oleh konsumen dalam memberikan *feedback* atau ulasan komentar. Hasil perolehan ulasan positif dan negatif dapat dilihat pada gambar



Gambar 5 Jumlah perbandingan ulasan

Visualisasi Data

Visualisasi data dalam penelitian ini menggunakan wordcloud. Fungsi librari wordcloud ini digunakan untuk mengidentifikasi sentimen data dalam memahami sentimen data teks berpola positif atau negatif terhadap kata yang sering muncul dalam masing-masing label positif maupun negatif. Warna biru digunakan untuk mengidentifikasi label positif terhadap kata yang sering muncul, kemudian warna merah digunakan untuk mengidentifikasi label negatif terhadap kata yang sering muncul. Visualisasi data wordcloud dapat dilihat pada gambar



Gambar 6 Visualisasi Ulasan Negatif dan Positif

Klasifikasi

Proses klasifikasi dilakukan setelah proses pembagian data train dan data test dilakukan dengan pembagian persentasi data train 90% dan data test 10%. Setelah melakukan pembagian data dan pembobotan dengan TF-IDF selanjutnya melakukan proses klasifikasi menggunakan model SVM dengan menggunakan librari. Berikut *code* pemanggilan librari untuk klasifikasi.

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(data_label['Ulasan_clean'],
data_label['label'],
test_size=0.1, stratify=data_label['label'], random_state=30)
#Import svm model
from sklearn import svm
#Create a svm Classifier
clf = svm.SVC(kernel='linear') # Linear Kernel
#Train the model using the training sets
clf.fit(X_train, y_train)
#Predict the response for test dataset
y_pred = clf.predict(X_test)
#Import scikit-learn metrics module for accuracy calculation
from sklearn import metrics
# Model Accuracy: how often is the classifier correct?
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))

```

Hasil klasifikasi menggunakan SVM :

Accuracy: 0.8823529411764706

Precision: 0.8913043478260869

Recall: 0.9761904761904762

KESIMPULAN DAN SARAN

Hasil analisis bigdata sentimen ulasan konsumen masker kesehatan pada marketplace menggunakan metode *support vector machine* dengan tujuan untuk mengetahui sentimen ulasan menghasilkan nilai akurasi sebesar 88%. Berdasarkan jumlah pembagian ulasan positif dan negatif didapatkan lebih banyak ulasan positif, yang menyimpulkan bahwa produk masker kesehatan sudah cukup baik.

Hasil analisis dapat disimpulkan dominannya hasil *scraping* ulasan pada produk masker kesehatan lebih banyak mengarah ke ulasan positif daripada ulasan negatif. Hasil pada wordcloud ulasan negatif memberikan *insigt* untuk memperbaiki kualitas masker yang masih kurang dalam segi tipis, tali mudah putus, sobek, bolong, kualitas karet dan pengemasan produk.

DAFTAR REFERENSI

- D. I. Sumantiawan, J. E. Suseno, and W. A. Syafei, "Sentiment Analysis of Customer Reviews Using Support Vector Machine and Smote-Tomek Links For Identify Customer Satisfaction," *Jurnal Sistem Informasi Bisnis*, vol. 13, no. 1, pp. 1-9, Jun. 2023. <https://doi.org/10.21456/vol13iss1pp1-9>
- Sumantiawan, D. I. (2024). METODE ANALISIS MENGGUNAKAN ALGORITMA RANDOM FOREST UNTUK PREDIKSI BIAYA ASURANSI KESEHATAN. *JICode: Jurnal Informatika Dan Komputer*, 1(1), 1–8. <https://doi.org/10.30599/jicode.v1i1.3294>
- Al-Smadi, M. et al., 2018, Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews, *Journal of Computational Science* 27386–393.
- Alamoodi, A.H. et al., 2021, Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy, *Computers in Biology and Medicine* 139 (August), 104957.
- Bin Alias, M.S. et al., 2021, Improved sampling data workflow using smtmk to increase the classification accuracy of imbalanced dataset, *European Journal of Molecular and Clinical Medicine* 8 (2), 91–99.
- Baek, Y. et al., 2020, Erasable pattern mining based on tree structures with damped window over data streams, *Engineering Applications of Artificial Intelligence* 94 (October 2018), 103735.
- Borg, A. and Boldt, M., 2020, Using VADER sentiment and SVM for predicting customer response sentiment, *Expert Systems with Applications* 162113746.
- Bourequat, W. and Mourad, H., 2021, Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine, *International Journal of Advances in Data and Information Systems* 2 (1), 36–44.
- SUMANTIAWAN, Dody Indra and Suseno, Jatmiko Endro and Syafei, Wahyul Amien (2023) *ANALISIS SENTIMEN ULASAN KONSUMEN MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN KOMBINASI SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE) DENGAN TOMEK LINKS*. Masters thesis, UNIVERSITAS DIPONEGORO.
- J. Supranto, *Statistik, Teori dan Aplikasi*. Surabaya: Penerbit Erlangga, 2016.
- S. Burns, *Python Machine Learning Deep Learning Tensorflow*. 2018.
- Sumantiawan, D.I., Astuti, S. and Kom, M., *ANALISIS KELAYAKAN PROMOSI JABATAN PIMPINAN BAGIAN PD. BPR BANK PEMALANG DENGAN MENGGUNAKAN METODE PROFILE MATCHING ANALYSIS ON RANK PROMOTION OF DIVISION MANAGER IN PD. BPR BANK PEMALANG USING PROFILE MATCHING*.