

ANALISIS SENTIMEN PADA ULASAN PEMBELIAN PRODUK DENGAN METODE *SUPPORT VECTOR MACHINE*

Felina Khasanah¹, Muhammad Hatta², Suwandi³.

^{1,2,3}Fakultas Teknologi Informasi, Sistem Informasi, Universitas Catur Insan Cendekia, Cirebon, Indonesia
Email: ¹felina.khasanah.si.20@cic.ac.id, ²muhammad.hatta@cic.ac.id, ³suwandi@cic.ac.id

Abstrak

Dibalik pertumbuhan pesat pada *marketplace* juga menghadapi persaingan yang semakin ketat salah satunya dalam menghadapi persaingan antar Toko yang menjual produk yang sama. Dalam menilai kualitas produk dilihat dari rating dan ulasan yang diberikan pelanggan, selain itu ulasan memiliki pengaruh yang besar terhadap minat pelanggan baru. Namun terdapat pelanggan yang memberikan ulasan positif tetapi memberikan rating kecil dan sebaliknya, sehingga menyebabkan firtur penilaian produk berdasarkan rating menjadi kurang baik karena tidak dapat merepresentasikan nilai sesungguhnya dan juga dengan jumlah ulasan yang banyak untuk menilai produk itu memiliki kualitas yang baik menjadi sulit. Penelitian ini membahas berkaitan dengan analisis sentimen ulasan pembelian produk pada Toko JJ Grosir Hijab di Shopee. Penelitian ini bertujuan agar dapat membantu Toko dalam memahami perspektif pelanggan dengan mengkategorikan ulasan positif dan ulasan negatif. Jumlah data ulasan sebanyak 812 ulasan produk yang dikumpulkan dengan bantuan *web scraper*. Analisis sentimen dilakukan dengan pelatihan model menggunakan metode pembobotan TF-IDF dan *Support Vector Machine*. Hasil dari penelitian yang dilakukan menghasilkan akurasi sebesar 93%. Hal ini menunjukkan model bekerja dengan baik dalam mengklasifikasikan ulasan. Informasi yang didapatkan dari hasil analisis dapat digunakan sebagai bahan evaluasi dalam meningkatkan kualitas dan inovasi produk.

Kata Kunci: Analisis Sentimen, *Marketplace Shopee*, Ulasan Produk, TF-IDF, SVM.

Abstract

Behind the rapid growth in the marketplace also faces increasingly fierce competition, one of which is in facing competition between stores that sell the same product. In assessing product quality, it can be seen from the ratings and reviews given by customers, besides that reviews have a big influence on new customer interest. However, there are customers who give positive reviews but give small ratings and vice versa, thus causing the product assessment firtur based on the rating to be unfavorable because it cannot represent the real value and also with a large number of reviews to assess the product has good quality becomes difficult. This research discusses related to sentiment analysis of product purchase reviews at the JJ Wholesale Hijab Shop at Shopee. This research aims to help stores understand customer perspectives by categorizing positive reviews and negative reviews. The number of review data is 812 product reviews collected with the help of a web scraper. Sentiment analysis is done by training the model using TF-IDF weighting method and Support Vector Machine. The results of the research conducted resulted in an accuracy of 93%. This shows that the model works well in classifying reviews. The information obtained from the analysis results can be used as evaluation material in improving product quality and innovation.

Keywords: Sentiment Analysis, *Shopee Marketplace*, Product Reviews, TF-IDF, SVM.

1. PENDAHULUAN

Pengaruh penggunaan *internet* pada masa pandemi Covid-19 mendorong peningkatan penggunaan aplikasi *digital* dan perubahan perilaku konsumen, salah satunya pada *platform marketplace* yang menjadi pemenuhan kebutuhan

pokok pada masa pandemi, hingga saat ini pelanggan mulai terbiasa berbelanja kebutuhan pokok secara online [1]. Dibalik pertumbuhan pesat pengguna, *marketplace* juga menghadapi persaingan ketat, hingga banyak *platform marketplace* terus mengembangkan inovasi untuk mempertahankan posisi pada pasar *online*, salah

satunya Shopee yang melakukan strategi promosi melalui iklan yang menarik, menghadirkan beragam fitur dan promo hingga dapat membuat banyak pengguna baru yang tertarik dan terdorong menggunakan Shopee. Dan pada tahun 2023 Shopee masuk dalam kategori platform *marketplace* dengan kunjungan situs terbanyak [2]

Saat ini *marketplace* Shopee menjadi pilihan yang paling digemari masyarakat baik bagi penjual maupun pembeli, salah satunya pada penelitian ini yaitu toko JJ Grosir Hijab yang bergabung dengan Shopee sejak tahun 2018. Toko menjual produk hijab dari usia anak-anak hingga dewasa. Tergabung dengan *marketplace* juga memiliki tantangan salah satunya dalam menghadapi persaingan antar toko yang menjual produk sama dengan harga yang kompetitif, agar dapat menarik minat pembeli baru salah satunya dapat dilakukan evaluasi dengan meningkatkan kualitas dan inovasi terhadap produk karena ulasan produk dari pembeli menjadi penting sebagai sumber informasi tambahan dalam membuat keputusan pembelian [3]. Pembeli dapat memberikan ulasan setelah menerima barang, ulasan pembelian produk terdiri dari bintang dan isi komentar ulasan berupa tanggapan, apresiasi maupun kritik dan masukan terhadap produk yang dibeli. Ulasan pembelian produk memiliki pengaruh yang signifikan terhadap minat pembeli dari pembeli lain [4].

Dalam menilai kualitas dari suatu produk dapat dilihat dari rating yang diberikan namun rating tidak selalu merepresentasikan nilai sesungguhnya karena terdapat rating dengan nilai 4 mempunyai ulasan negatif dan rating dengan nilai 2 mempunyai ulasan positif, sehingga untuk menilai kualitas produk menjadi sulit [5] Oleh karena itu dibutuhkan analisa lebih lanjut dari ulasan yang diberikan yaitu dapat dilakukan dengan analisis sentimen pada ulasan komentar yang diberikan. Analisis sentimen berfungsi untuk menilai opini, sikap dan emosi dalam teks. Proses analisis sentimen bertujuan untuk mengubah teks mentah menjadi wawasan yang berharga seperti preferensi pelanggan terhadap produk. Pada analisis sentimen ulasan produk dapat diartikan sebagai proses dalam memahami pendapat dan pengalaman dari pembeli terhadap produk tersebut [6].

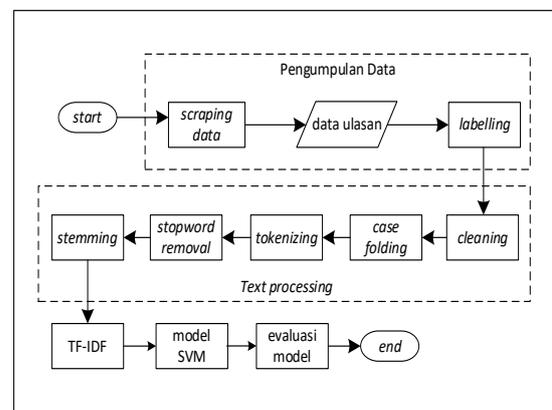
Pada penelitian ini menggunakan metode *Support Vector Machine* merupakan algoritma *machine learning* yang digunakan dalam *data mining*, metode ini berfungsi dalam mencari *hyperplane* yang memisahkan data ke dalam dua kelas yang digunakan untuk klasifikasi pada ulasan, dengan mengidentifikasi ulasan tersebut termasuk dalam kelas positif atau kelas negatif [7]. Beberapa penelitian yang dilakukan dibidang analisis sentimen menggunakan metode *Support Vector Machine* sebelumnya yang dilakukan oleh (Putri, Dea Safryda, and Taufik Ridwan, 2023) penelitian

ini membahas terkait dengan analisis sentimen ulasan yang diambil dari aplikasi Pospay pada *Google Play Store*, hal ini bertujuan dalam pengambilan keputusan penggunaan aplikasi dan evaluasi dengan hasil akurasi sebesar 95% [8]. Penelitian yang dilakukan oleh (Moch Saoki Syahlan, Dede Irmayanti, dan Syariful Alam, 2023) yang membahas dalam rangka menginformasikan kepada Dinas Pariwisata Purwakarta terkait potensi peningkatan destinasi di Purwakarta pada Air Mancur Sri Baduga dengan melakukan analisis sentimen dengan akurasi nilai sebesar 81% [9]. Penelitian yang dilakukan oleh (Siti Azza Amira, Satria Utama, dan Muhammad Hanif Fahmi, 2020) Pada penelitian ini berfokus pada analisis sentimen dari informasi ulasan hotel yang didapatkan dari situs Tripadvisor sebagai acuan bagi calon pengunjung dalam menentukan hotel mana yang terbaik untuk mereka kunjung saat bepergian ke luar kota, dengan hasil pengujian didapatkan akurasi nilai 88% [10]. Pada penelitian yang dilakukan oleh (Sholikun.S, 2023) penelitian ini membahas analisis sentimen kepuasan pengguna aplikasi *Halodoc* dengan hasil pengujian didapatkan akurasi nilai 99% [11]. Dan pada penelitian yang dilakukan oleh (Pradhana, R.M, 2021) membahas berkaitan dengan ini membahas berkaitan dengan kebijakan PPKM mikro menimbulkan pro-kontra, menggunakan data dari Twitter untuk memahami respon dan persepsi masyarakat, dengan hasil pengujian didapatkan akurasi nilai 97% [12].

Berdasarkan beberapa penelitian diatas, metode *Support Vector Machine* dipilih karena memiliki nilai akurasi yang cukup tinggi pada bidang pengklasifikasian dibanding metode lain, dan mempunyai kinerja lebih baik dalam mengklasifikasikan data pada banyak kasus [13].

2. METODOLOGI PENELITIAN

Metodologi penelitian yang dilakukan pada penelitian ini melalui beberapa tahapan dapat dilihat pada gambar 2.1.



Gambar 2. 1 Tahapan Penelitian

2.1 Pengumpulan Data

Dalam penelitian ini proses pengumpulan data dilakukan dengan teknik *scraping* data dengan bantuan *web scraper extension*, dalam proses ekstraksi data dilakukan dengan membuat *sitemap* yang berisi *link* halaman produk hijab dari JJ Grosir Hijab di Shopee. Setelah proses *scraping* didapatkan kumpulan ulasan dari produk hijab sebanyak 812 data, dan dari ulasan tersebut dilakukan pelabelan dengan kategori positif atau negatif pada tiap ulasan komentar secara manual.

2.2 Text PreProcessing

Preprocessing merupakan tahapan dalam mempersiapkan data dari yang tidak terstruktur menjadi data yang lebih baik dan siap untuk dilakukan analisis. Tujuan utama dari *preprocessing* pada data untuk meningkatkan kualitas data, memastikan keakuratan hasil analisis, dan mengatasi masalah atau kekurangan yang muncul pada data. Tahapan *preprocessing* sebagai berikut:

- Cleaning* merupakan pengurangan noise pada teks dengan cara menghilangkan tanda baca, angka, atau karakter khusus yang tidak diperlukan dalam analisis.
- Case Folding* merupakan salah satu tahap penting dalam *text processing*, pada tahap ini semua huruf akan diubah menjadi huruf kecil untuk mencegah kasus sensitif dalam pengenalan pada teks untuk memastikan kata-kata yang sama dalam bentuk huruf besar dan kecil tidak dianggap berbeda.
- Tokenizing* dapat didefinisikan sebagai pemecahan teks menjadi bagian praktis yang biasa disebut dengan token seperti kata frasa, lain agar pengerjaan teks lebih efektif. Misalnya kalimat 'Hijab mudah dibentuk' akan dipecah menjadi ('Hijab', 'mudah', 'dibentuk').
- Stopword Removal* atau yang biasa disebut dengan *filtering* merupakan tahap pemilihan kata yang penting dari hasil *tokenizing*. Tahap ini akan menghilangkan kata yang tidak memiliki arti dan pengaruh dalam menganalisis sebuah data. Kata-kata yang tidak lebih dari tiga huruf tidak memiliki arti yang berarti seperti 'di', 'ke'.
- Stemming* merupakan proses merubah kata-kata yang ditemukan dalam teks menjadi bentuk dasarnya dengan menghilangkan prefix (menghilangkan awalan kata) dan sufiks (menghilangkan akhiran kata), misalnya kata ('bahannya', 'jahitannya', 'dibentuk'), akan ditransformasikan menjadi kata ('bahan', 'jahitan', 'bentuk').

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

Setelah dataset melalui tahapan *preprocessing*, selanjutnya dilakukan pembagian data latih dan data uji dengan perbandingan 70:30, dilanjutkan dengan pembobotan TF-IDF untuk menemukan kata-kata yang paling penting dalam dokumen tertentu berdasarkan seberapa sering kata tersebut muncul dalam seluruh kumpulan dokumen. Pembobotan TF-IDF digunakan untuk menghitung bobot (w) masing masing dokumen terhadap kata kunci dengan persamaan berikut.

$$TF_{t,d} = F(t, d) \quad (1)$$

$$IDF_t = \log\left(\frac{N}{DF_t}\right) \quad (2)$$

$$W_{t,d} = TF_{t,d} \times IDF_t \quad (3)$$

Keterangan pada persamaan rumus (1), (2), (3) sebagai berikut [14]:

$TF_{t,d}$: Mencari bobot dari sebuah dokumen.

$F(t, d)$: Frekuensi munculnya *term t* pada dokumen

IDF_t : Menghitung penyebaran *term* pada dokumen.

N : Frekuensi dari semua kumpulan dokumen

DF_t : Frekuensi dari dokumen memiliki *term t*.

$W_{t,d}$: Pembobotan TF-IDF.

2.4 Klasifikasi SVM *Support Vector Machine*

Pada tahap ini *dataset* sebelumnya yang telah dibagi dua yaitu data *training* dan data *testing*, data *training* digunakan untuk melatih algoritma atau model yang dibangun untuk mempelajari dataset dan data *testing* digunakan untuk mengevaluasi performa dari model yang dilatih. Pada klasifikasi ini menggunakan metode *Support Vector Machine*. Dalam kasus SVM dengan kernel linear, model mencoba untuk menemukan *hyperplane* yang memisahkan dua kelas data dengan margin maksimum. *Hyperplane* ini dapat dinyatakan dalam bentuk persamaan [15]:

$$w \cdot x_i + b = 0 \text{ atau } w_1x_1 + w_2x_2 + b = 0 \quad (4)$$

Dari persamaan (4) dengan keterangan:

w : vektor bobot

x : vektor fitur

b : bias

2.5 Evaluasi Model

Pada proses evaluasi model ini menggunakan *confusion matrix*, dari model yang sudah di bangun untuk memperkirakan objek yang

benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas sebenarnya atau dengan kata lain berisi informasi nilai sebenarnya dan prediksi pada klasifikasi. Rumus yang digunakan untuk menghitung *confusion matrix* dalam menentukan akurasi, presisi, *recall*, dan *F1-Score* sebagai berikut [16]:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Presisi = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - score = \frac{2 \times Presisi \times Recall}{Presisi + Recall} \quad (8)$$

Dari persamaan rumus (5), (6), (7), (8) sebagai berikut:

- True Positive* (TP): kasus positif yang diprediksi benar.
- True Negative* (TN): kasus negatif yang diprediksi benar.
- False Positive* (FP): kasus negatif yang salah diprediksi sebagai positif
- False Negative* (FN): kasus positif yang salah diprediksi sebagai negatif

3. HASIL DAN PEMBAHASAN

Pada bagian ini berisikan hasil dan pembahasan dari tahapan analisis yang telah dilakukan sebagai berikut:

3.1. Pengumpulan Data

Gambar 3. 1 Tampilan Proses Pelabelan Data

Pada gambar 3.1 menampilkan data ulasan hasil dari proses *scraping* yang selanjutnya dilakukan proses pemberian label pada setiap ulasan secara manual dengan mengidentifikasi ulasan apakah termasuk dalam kategori ulasan positif atau ulasan negatif.

3.2. Hasil Text Preprocessing

Gambar 3. 2 Tampilan Hasil Text Preprocessing

3.3. Hasil Pembobotan TF-IDF

| Kata | IDF |
|--------|----------|
| bagus | 1.899842 |
| bahan | 2.857773 |
| sesuai | 2.160697 |
| tampil | 1.824225 |
| tidak | 2.626787 |
| warna | 1.712999 |

Gambar 3. 3 Hasil Pembobotan TF-IDF

Pada gambar 3.3 menampilkan beberapa fitur kata disertai dengan bobot nilai TF-IDF pada masing-masing kata.

3.4. Hasil Pengujian dan Evaluasi Model

| Classification Report: | | | | |
|------------------------|-----------|--------|----------|---------|
| | precision | recall | f1-score | support |
| negatif | 0.81 | 0.87 | 0.84 | 54 |
| positif | 0.96 | 0.94 | 0.95 | 190 |
| accuracy | | | 0.93 | 244 |
| macro avg | 0.89 | 0.91 | 0.90 | 244 |
| weighted avg | 0.93 | 0.93 | 0.93 | 244 |

| Confusion Matrix: | |
|-------------------|--|
| [[47 7] | |
| [11 179]] | |

Gambar 3. 4 Classification Report

Pada gambar 3.4 dari hasil *Classification Report* didapatkan:

- Pada *accuracy* menunjukkan nilai proporsi dari semua prediksi yang benar dengan data asli didapatkan nilai 0.93, model memprediksi dengan benar 93% dari ulasan.
- Pada *macro avg precision* menunjukkan nilai dari persentase kinerja model dalam memprediksi semua ulasan dengan persentase 89%.
- Pada *macro avg recall* menunjukkan nilai dari persentase kinerja model dalam mengidentifikasi setiap ulasan berdasarkan kelasnya dengan persentase 91%.

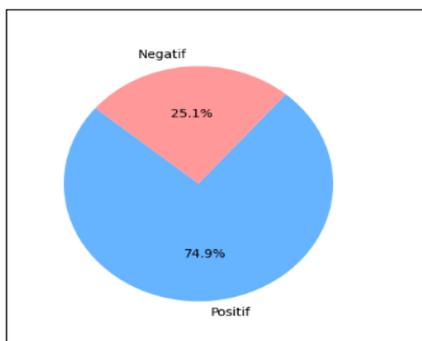
- d. Pada *macro avg F1-Score* menunjukkan nilai dari persentase kinerja model dengan persentase keseimbangan yang baik antara *precision* dan *recall* dengan persentase 90%.

Pada gambar 3.4 dari hasil *Confusion Matrix* didapatkan:

- a. Nilai *true negatif* atau jumlah ulasan yang diprediksi negatif dengan benar oleh model sebanyak 47 ulasan.
- b. Nilai *false positif* atau jumlah ulasan positif yang diprediksi secara salah sebagai kelas negatif sebanyak 7 ulasan.
- c. Nilai *false negatif* atau jumlah ulasan negatif yang diprediksi secara salah sebagai kelas positif sebanyak 11 ulasan.
- d. Nilai *true positif* atau jumlah ulasan yang diprediksi positif dengan benar oleh model sebanyak 179 ulasan.

3.5. Visualisasi

- a. Diagram Lingkaran



Gambar 3. 5 Hasil Persentase Ulasan Berdasarkan Label

Pada gambar 3.5 menunjukkan persentase ulasan pada masing-masing kelas, dimana jumlah kelas positif sebanyak 74,9% dan kelas negatif sebanyak 25,1%.

- b. *Word Cloud*



Gambar 3. 6 Tampilan *Word Cloud*

Pada gambar 3.6 menampilkan sebaran kata pada ulasan positif dan ulasan negatif, dimana untuk setiap kata yang sering ditemukan dalam

dokumen akan memiliki ukuran dominan dan sebaliknya.

4. KESIMPULAN

Dari hasil analisis sentimen pada penelitian berhasil dalam mengklasifikasi ulasan produk dengan jumlah ulasan positif lebih banyak dari pada negatif. Dimana dapat disimpulkan 74,9% pelanggan merasa puas dengan kualitas produk berdasarkan ulasan positif yang diberikan dari dataset 812 ulasan produk dengan kategori ulasan positif sebanyak 615 dan ulasan negatif sebanyak 197. Berdasarkan hasil dari analisis dengan menggunakan model *Support Vector Machine* memiliki akurasi tinggi sebesar 93% dengan perbandingan 70:30, hal ini membuktikan model dapat bekerja dengan baik dalam mengklasifikasikan ulasan.

5. REFERENCES

- [1] M. Fadillaha, M. N., & Subchan, "DAMPAK COVID-19 TERHADAP PERILAKU KONSUMEN DALAM PENGGUNAAN MARKETPLACE DI INDONESIA Muhammad Nur Fadillah," *Jurnal mitra manajemen*, 12(1), 123-130., 2021.
- [2] M. S. Dwiriansyah, "Analisis Perbandingan Strategi Promosi pada Marketplace Shopee dan Lazada," vol. 2, no. 3, pp. 206–215, 2023.
- [3] E. Sera, H. Hazriani, M. Mirfan, and Y. Yuyun, "Analisis Sentimen Ulasan Produk di E-Commerce Bukalapak Menggunakan Natural Language Processing," *Prosiding SISFOTEK*, pp. 237–243, 2023.
- [4] E. H. Muktafin, K. Kusriani, and E. T. Luthfi, "Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing," *Jurnal Eksplora Informatika*, vol. 10, no. 1, pp. 32–42, 2020, doi: 10.30864/eksplora.v10i1.390.
- [5] R. Kosasih and A. Alberto, "Analisis Sentimen Produk Permainan menggunakan Metode TF-IDF dan Algoritma K-Nearest Neighbor," *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 6, no. 1, pp. 134–139, 2021, [Online]. Available: <https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/3893>
- [6] B. Goodey, "Analisis Sentimen Pelanggan." [Online]. Available: <https://www.sentisum.com/customer-sentiment-analysis>

- [7] F. Sulianta, *Basic Data Mining from A to Z*. 2023.
- [8] D. Safryda Putri and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi Pospay Dengan Algoritma Support Vector Machine,” *Jurnal Ilmiah Informatika*, vol. 11, no. 01, pp. 32–40, 2023, doi: 10.33884/jif.v11i01.6611.
- [9] M. S. Syahlan, D. Irmayanti, and S. Alam, “MENGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) (Studi Kasus : Taman Air Mancur Sri Baduga Purwakarta),” vol. 8, no. 2, pp. 315–319, 2023.
- [10] S. A. Amira, S. Utama, and H. Fahmi, “Penerapan Metode Support Vector Machine untuk Analisis Sentimen pada Review Pelanggan Hotel,” vol. 7, no. 2, pp. 40–48, 2020, doi: <https://doi.org/10.15294/edukomputika.v7i2.42608>.
- [11] Sholikun, “Analisis Sentimen Kepuasan Pengguna Aplikasi Halodoc Menggunakan Metode Support Vector Machine,” 2023.
- [12] R. M. Pradhana, “Analisis Sentimen Publik terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Skala Mikro Menggunakan Algoritma Support Vector Machine Studi Kasus Twitter,” Universitas Dinamika, 2021.
- [13] R. Ramadhan, M. A., & Andarsyah, *Klasifikasi Text SPAM Menggunakan Metode Support Vector Machine Dan Naive Bayes*. Penerbit Buku Pedia, 2022.
- [14] P. B. R. Ninuk Wiliani, Nuke L. Chusna, *Analisis Sentimen Terhadap Pro Kontra Aksi Unjuk Rasa Mahasiswa dengan Naive Bayes dan Information Gain*. Penerbit NEM, 2023.
- [15] Deny Jollyta (et al, *Algoritma Klasifikasi Untuk Pemula Solusi Python dan RapidMiner*. 2023.
- [16] S. Juniarsih, E. F. Ripanti, and E. E. Pratama, “Implementasi Naive Bayes Classifier pada Opinion Mining Berdasarkan Tweets Masyarakat Terkait Kinerja Presiden dalam Aspek Ekonomi,” *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 3, p. 239, 2020, doi: 10.26418/justin.v8i3.39118.