

Analisis Sentimen Masyarakat Terhadap Penerima Beasiswa Kartu Indonesia Pintar Kuliah Dengan Metode Support Vector Machine

(Analysis of Community Sentiment Toward Indonesia Smart Card Scholarship Recipients with Support Vector Machine Method)

M. Fahriz Zain Jannan^{*1)}, Yudha Dwi Putra Negara²⁾

¹⁾Program Studi Sistem Informasi, Fakultas Teknik, Universitas Trunojoyo Madura
Jl. Raya Telang, PO BOX 02 Kecamatan Kamal, Bangkalan Jawa Timur 69162 Indonesia

Riwayat: Copyright ©2024, JITU, Submitted: 26 Juni 2024; Revised: 15 Agustus 2024;
Accepted: 22 Agustus 2024; Published: 30 September 2024
DOI: 10.32938/jitu.v4i2.7598

Abstrac - The smart Indonesia card is the government's effort to help outstanding students who have economic limitations to continue their studies at university. The crowded discussion of the smart indonesia card on twitter made the author interested in knowing the public's sentiment towards the recipients of the smart indonesia card scholarship for college with the keyword "KIPK". Sentiment consists of negative, positive, and neutral sentiments. The method used to analyze public opinion on social media twitter uses Support Vector Mechine (SVM). The Term Frequency-inverce document Frequency (TF-IDF) is used to measure how important words are in a piece of data. Meanwhile, to measure the performance of the SVM classification using the Confusion Matrix method. A linear carnal is a carnal that we use in data analysis. The results of the evaluation obtained were 81.82%, precession 80%, recall 82%, and F1-score 78%, it can be concluded that public sentiment on social media twitter towards the recipients of the Indonesia Smart College Card Scholarship is more inclined to neutral sentiment.

Keywords - kipk, svm, tf-idf, confuntion matrix.

Abstrak - Kartu indonesia pintar merupakan upaya pemerintah membantu para siswa berprestasi yang memiliki keterbatasan ekonomi untuk melanjutkan studi ke perguruan tinggi. Ramainya pembahasan kartu indonesia pintar di twitter membuat penulis tertarik untuk mengetahui sentimen masyarakat terhadap penerima beasiswa kartu indonesia pintar kuliah dengan keyword "KIPK". Sentimen terdiri dari sentimen negatif, positif, dan netral. Metode yang di gunakan menganalisis pendapat masyarakat pada media sosial twitter menggunakan Support Vector Mechine (SVM). Term Frequency-inverce document Frequency (TF-IDF) kami gunakan untuk melakukan pengukuran seberapa penting kata dalam suatu data. Sedangkan untuk mengukur kinerja klasifikasi SVM menggunakan metode Confusion Matrix. Karnel linear

merupakan karnel yang kami gunakan pada analisis data. Hasil evaluasi yang di dapat yaitu 81,82%, precession 80%, recall 82%, dan F1-score 78%, dapat di simpulkan bahwa sentimen masyarakat pada media sosial twitter terhadap penerima beasiswa kartu indonesia pintar kuliah lebih condong ke sentimen netral.a.

Kata kunci - kipk, svm, tf-idf, confuntion matrix.

I. PENDAHULUAN

Kartu Indonesia Pintar Kuliah (KIP-K) merupakan bantuan biaya pendidikan yang di berikan pemerintah melalui program kartu indonesia pintar kuliah agar di peroleh akses yang luas dalam belajar untuk peserta didik dan mahasiswa baru dari keluarga miskin atau rentan miskin[1]. Kebijakan KIP kuliah yang di dikeluarkan oleh pemerintah di bawah wewenang kementerian dan kebudayaan bertujuan untuk menyediakan bantuan khusus untuk mahasiswa miskin sehingga dapat melanjutkan kuliah, dengan demikian sasaran dari kebikan ini adalah untuk mencegah mahasiswa putus kuliah[2]. Ada beberapa kriteria mahasiswa yang bisa mendapatkan KIP Kuliah yaitu prestasi, penghasilan orang tua, jumlah tanggungan, luas tanah dan penerima bantuan sosial berupa KIP/KSS. Karena kurangnya penyeleksian penerima beasiswa KIP-K mengakibatkan banyak mahasiswa yang lebih membutuhkan bantuan terabaikan[3].

Dengan kurang tepatnya penerima beasiswa terdapat minimnya tanggungjawab atas tujuan serta moral dalam pemakaian dana sehingga masih banyak kasus penyalahgunaan KIP-K di luar fungsi, menjadikannya sarana pemenuhan gaya hidup hedonisme serta banyak kasus salah sasaran[4]. Hedonisme dapat di artikan sebagai filosofi atau pandangan hidup yang mengedepankan kesenangan dan kenikmatan sebagai tujuan utama kehidupan[5]. Fenomena ini menyebabkan terjadinya perbincangan yang sangat masif terjadi pada media sosial twitter.

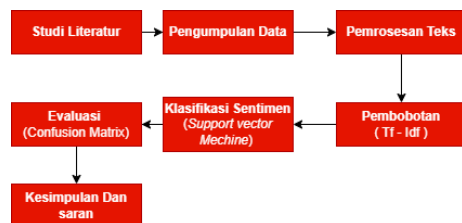
^{*)} M. Fahriz Zain Jannan
Email: zeinbaim4@gmail.com

Twitter merupakan media sosial yang merupakan sebuah media informasi yang menghubungkan antar pengguna dengan informasi – informasi tentang hal – hal sesuai minat[6]. Dari semua ulasan yang di ungkapkan oleh pengguna twitter dapat di kelompokkan menjadi ulasan positif, negatif dan netral. Sehingga peneliti tertarik melakukan analisis sentimen untuk meneliti ulasan yang paling banyak di ungkapkan masyarakat pada platform twitter. Untuk melakukan analisis sentimen peneliti membutuhkan dataset untuk melakukan pengujian. Dataset ini di peroleh dengan cara melakukan crawling menggunakan kata kunci ‘KIPK’ pada media sosial twitter.

Dalam penelitian ini peneliti mengusulkan pendekatan untuk melakukan analisis sentimen terhadap penerima beasiswa kartu indonesia pintar kuliah dengan menggunakan algoritma Support Vector machine. Peneliti melakukan crawling data pada twitter menggunakan library tweet-harvest dan menggunakan auth-token sebagai kunci untuk mendapatkan dataset dengan kata kunci yang telah kita tulis di kolom pencarian. Kemudian dilakukan preprocessing data seperti stopword, tokenization, dan stemming setelah itu melakukan pembobotan pada frekuensi kemunculan kata menggunakan Term frequency – inverse document frequency (tf-idf) sebelum menerapkan Support vector machine dalam mengukur tingkat keakuratan pada teknik metode yang di gunakan.

II. METODE PENELITIAN

Penelitian ini merupakan analisis sentimen yang menggunakan metode pemrosesan dan pembobotan teks serta menggunakan algoritma support vector machine untuk memperoleh hasil penelitian dan menggunakan confusion matrix untuk mengevaluasi hasil yang telah di peroleh. Berikutnya adalah Gambar 1 yang merupakan alur penelitian analisis sentimen penerima beasiswa KIPK.



Gambar 1. Alur Penelitian Analisis sentimen Penerima Beasiswa KIPK

A. Studi Literatur

Pada tahapan ini peneliti melakukan pencarian referensi, teori, dan temuan – temuan dari penelitian yang terkait dengan penelitian. Studi literatur merupakan langkah awal dalam perumusan masalah penelitian. Pentingnya studi literatur adalah untuk membantu pemahaman peneliti terhadap konteks penelitian, menemukan ide – ide baru, serta memanfaatkan informasi dan pemikiran relevan yang telah ada[7].

B. Pengumpulan Data

Pengumpulan data pada penelitian ini menggunakan metode crawling pada media sosial twitter. Untuk melakukan crawling peneliti menggunakan tools yang bernama tweet – harvest kami tidak menggunakan sncrape dikarenakan twitter sudah mengubah API dan fiturnya sehingga kami tidak menggunakannya[8].

Untuk menggunakan tools tweet – harvest kita terlebih dahulu harus menginstal node-js karena tweet - harvest dibuat menggunakan node-js. Tweet yang kami crawling yaitu ulasan penerima bantuan kartu indonesia pintar dengan menggunakan kata kunci ‘KIPK’, kami mengambil tweet dalam rentan waktu tanggal 7 april 2024 sampai 29 april 2024.

C. Pemrosesan Teks

Pemrosesan teks atau lebih dikenal dengan text preprocessing merupakan proses pengolahan teks yang bertujuan untuk mengurangi noise pada dataset serta mengubah dataset menjadi bentuk yang lebih terstruktur. Tahapan pemrosesan teks pada penelitian ini yaitu cleaning, casefolding, tokenization, dan stopword.

Cleaning

Cleaning merupakan proses membersihkan kata dan menyeleksi kata yang tidak di perlukan seperti html, emoticon, hastag, mention dan url.

Case folding

Case folding merupakan salah satu text preprocessing yang paling sederhana dan efektif. Tujuan dari case folding untuk mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf ‘a’ samapai ‘z’ yang di terima. Karakter selain huruf di hilangkan dan dianggap delimiter[9].

Stopword

Stopword merupakan proses menghilangkan kata yang kurang efektif

Stemming

Stemming merupakan proses penyaringan kata yang terdapat kata sambung, kata ganti, kata depan, menjadi kata dasar dengan menghilangkan kata awalan atau akhiran.

Tokenization

Tokenization merupakan proses seleksi pemotongan kata dalam kalimat. Diberikan pemisah seperti tanda koma (,), tanda titik (.), dan tanda pemisah lainnya.

D. Pembobotan kata

Tf-idf merupakan metode yang kami gunakan untuk mencari bobot suatu kata dalam dokumen kunci setiap kategori dan mencari kata kunci yang hampir mirip dengan kategori yang tersedia. Pembobotan ini menggunakan teknik Term Frequency dan Inverse

Documen Vrekuensi (TF – IDF) untuk mengekstraksi kata dari suatu teks dengan cara menggabungkan metode TF dan IDF [10].

Term Frequency adalah semakin besar frekuensi kemunculan kata pada sebuah dokumen maka akan semakin besar nilai bobot untuk kata tersebut. Inverse Document Frequency (IDF) adalah semakin besar frekuensi kemunculan kata maka nilai bobot kata tersebut akan semakin kecil[11]. Persamaan TF-IDF yang di gunakan pada penelitian ini yaitu :

$$W_{i,j} = t_{f,i,j} \times \log\left(\frac{N}{df_i}\right)$$

Keterangan:

$t_{f,i,j}$ = banyaknya kata-i pada dokumen ke-j

N = total dokumen

df_i = banyaknya kata dokumen yang mengandung kata ke i

E. Klasifikasi Support Vector Mechine

Support vector machine merupakan machine learning yang dapat memungkinkan perhitungan untuk masalah linear dengan menerapkan transformasi matematis untuk ruang belajar menggunakan fungsi kernel[12]. Support vector machine memiliki konsep sentral dalam mengklasifikasikan data yaitu menentukan hyperplane terbaik untuk memberi jarak atau pemisah antar dua kelas[13]. Dalam penelitian kali ini kernel yang akan di gunakan adalah kernel linear. Berikut adalah Tabel 1 yang merupakan Tabel persamaan setiap kernel support vector machine.

Tabel 1 Persamaan setiap kernel SVM

No	Jenis kernel	model
1	Linear	$k(x, x') = x \cdot x'$
2	Polinomial	$k(x, x') = (x \cdot x' + c)'$
3	RBF Gaussian	$k(x, x') = \exp(-\gamma \ x - x'\ ^2)$
4	Sigmoid	$k(x, x') = \tanh(ax \cdot x' + \beta)$

F. Evaluasi

Pada tahapan ini akan di lakukan evaluasi model di mana performa model di nilai berdasarkan metrik kesalahan untuk memperoleh tingkat akurasi. Penilaian kinerja model melibatkan penerapan matriks kebingungan untuk mendapatkan ukuran-ukuran seperti akurasi, presisi, recall dan F1-score[14]. Confusion matrix di gunakan untuk mengukur kinerja model pada setiap iterasi. Terdapat empat elemen utama dalam

confusion matrix, yaitu True Positif (TP), True Negatif (TN), False Positif (FP), False Negatif (FN)[15]. Komponen ini digunakan untuk mengukur performa model dalam mengklasifikasi data. Berikut adalah persamaan untuk menghitung nilai akurasi , presisi, recall, dan f1-score dengan persamaan berikut:

$$\text{Akurasi} = (TP+TN)/(TP+TN+FP+FN) \times 100\% \quad (3)$$

$$\text{Presisi} = TP/(TP+FP) \quad (4)$$

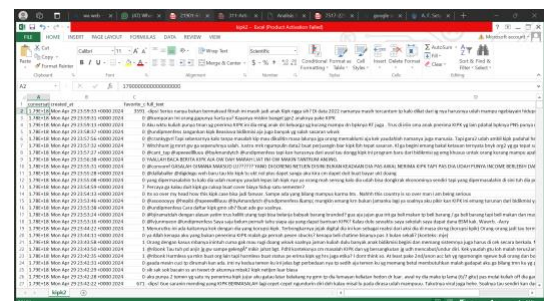
$$\text{Recall} = TP/(TP+FN) \quad (5)$$

$$\text{f1-Score} = (2 \times \text{presisi} \times \text{Recall}) / (\text{Presisi} + \text{Recall}) \quad (6)$$

III. HASIL DAN PEMBAHASAN

A. Dataset

Penelitian ini menggunakan data twitter yang di simpan dalam bentuk CSV menggunakan tools tweet-harvest yang di jalankan menggunakan google collab dengan kata kunci yang di gunakan adalah 'kipk', data yang di peroleh pada penelitian ini sebanyak 3044 data dengan rentan waktu dari tanggal 7 april 2025 sampai 29 april 2024. Berikut gambar 2 merupakan dataset yang di peroleh dari hasil crawling.



Gambar 2. Dataset yang diperoleh dari proses crawling

B. Pemrosesan Data

Dataset yang di peroleh selanjutnya dilakukan pemrosesan teks dengan langkah *cleaning* yaitu membersihkan kata, selanjutnya kata yang telah di bersihkan dilakukan *tokenization* dan *case folding* , *stopword*, dan langkah terakhir dalam pemrosesan data yaitu melakukan *stemming* atau menghilangkan kata imbuhan yang dilakukan secara berturut-turut. Berikut adalah Tabel 2 yang merupakan hasil pemrosesan teks

Tabel 2 Hasil pemrosesan teks

No	Kata sebelum pemrosesan teks	Kata setelah pemrosesan teks
1	[Untirfess] halo aku maba dpt ukt gol.3 ada kesempatan bisa dpt kipk ga ya gol.3?	untirfess halo maba dpt ukt gol sempat dpt kipk ga ya gol

- 2 @undipmenfess tapi undipmenfess temen aku juga punya temen kipk benerbener kipk yang bener-bener hedon treatment inila hedon treatment inila beli hp iphone itulah beli hp iphone terbaru kayak aku dapet kipk gitulo mikir dia mampu kok tapi kok bisa dapet kipk gitulo padahal masih banyak yang kurang mampu yang bener-bener butuh biaya

C. Pembobotan Kata

Pembobotan pada penelitian ini dilakukan menggunakan metode term Frequency – Inverse Document Frequency (TF-IDF). Berikut merupakan hasil pembobotan kata menggunakan TF-IDF. Berikutnya adalah Tabel 3 yang merupakan hasil dari pembobotan kata.

Tabel 3. Hasil pembobotan kata

No	Komentar	Bobot kata
1	untirfess halo maba dpt ukt gol sempat dpt kipk ga ya gol	untirfess = 7.314552832 halo = 5.70511492 mab = 5.235111291 dpt = 3.237015388 ukt = 2.94510498 gol = 5.299649812 sempat = 5.522793363 kipk = 0.22364301 ga = 1.543111709 ya = 1.791093912

D. Klasifikasi Support Vector Mechine

Implementasi Support Vector Machine (SVM) dalam penelitian ini menggunakan bahasa pemrograman python. Python memiliki library Sklearn yang dapat digunakan untuk mengimplementasikan SVM. Berikut adalah gambar 3 yang merupakan hasil klasifikasi menggunakan *support vector support mechine*.

```
print(classification_report(y_test, y_pred))
```

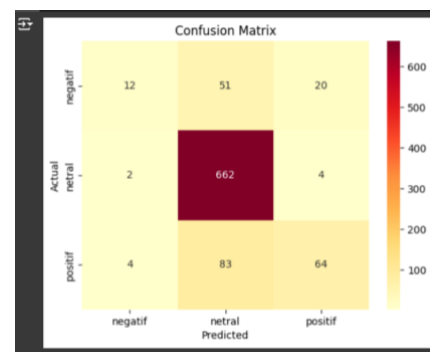
	precision	recall	f1-score	support
negatif	0.67	0.14	0.24	83
netral	0.83	0.99	0.90	668
positif	0.73	0.42	0.54	151
accuracy			0.82	902
macro avg	0.74	0.52	0.56	902
weighted avg	0.80	0.82	0.78	902

Gambar 3 Hasil klasifikasi menggunakan Support Vecctor Mechine

Pada Gambar 3 telah di tampilkan bahwa hasil akurasi metode *support vector mechine* pada data uji sebesar 81.82%, presisi sebesar 80%, recall sebesar 82% f1-score sebesar 78% yang selanjutnya akan di lakukan tahapan evaluasi.

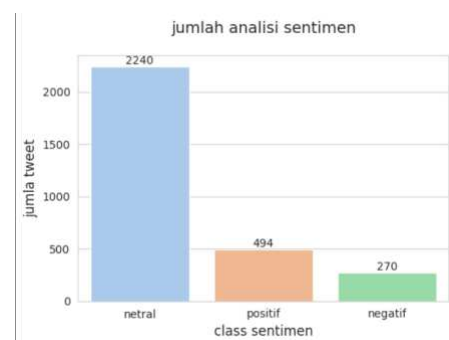
E. Evaluasi

Untuk mengevaluasi model yang telah dibuat, digunakan metode confusion matrix dan cross validation score. Confusion matrix digunakan untuk menunjukkan jumlah hasil klasifikasi yang benar dan salah dari model. Berikutnya adalah Gambar 4 yang merupakan visualisasi *confusion matrix*.



Gambar 4 Hasil confusion matrix

Hasil evaluasi menggunakan confusion matrix yang di dapatkan pada data training yang di tunjukkan pada gambar 4 adalah hasil true positif sejumlah 64 komentar, True netral sejumlah 662 komentar, True negatif sejumlah 12 komentar. Hasil evaluasi pada confusion matrix juga menunjukkan nilai false positif 24 komentar, false netral 134 dan false negatif sebanyak 14 komentar. Berikutnya adalah gambar 5 yang merupakan digram klasifikasi dari data yang di peroleh.



Gambar 5 diagram klasifikasi

Dari keseluruhan data sejumlah 3044 yang telah di kumpulkan lalu dilakukan pemrosesan, pembobotan teks serta klasifikasi menggunakan metode naive bayes mendapatkan hasil seperti pada gambar 5.

IV. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian metode Support Vector Mechine untuk mengklasifikasi penerima

beasiswa kartu indonesia pintar kuliah dengan tweet menggunakan bahasa indonesia menggunakan ekstraksi fitur Tf-Idf di dapat kesimpulan bahwa penggunaan metode support vector machine tingkat nilai akurasi sebesar 81.82%, precision sebesar 80%, recall sebesar 82%, F1-Score sebesar 78%. Yang artinya cukup baik dalam penelitian mengenai penerima beasiswa kartu indonesia pintar kuliah.

Dalam penelitian ini juga di dapat bahwa sentimen netral lebih dominan dari pada sentimen positif dan negatif, serta sentimen positif lebih dominan daripada sentimen negatif. Hal ini menunjukkan bahwa lebih banyak masyarakat yang berpendapat positif tentang penerima beasiswa kartu indonesia pintar.

DAFTAR PUSTAKA

- [1] Suganda, Marsani Asfi, Ridho Taufiq Subagio, and Ricky Perdana Kusuma, "Penentuan Penerima Bantuan Beasiswa Kartu Indonesia Pintar (Kip) Kuliah Menggunakan Naïve Bayes Classifier," *JSiI (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 193–199, 2022, doi: 10.30656/jsii.v9i2.4376.
- [2] A. Amin, R. N. Sasongko, and A. Yuneti, "Kebijakan Kartu Indonesia Pintar untuk Memerdekakan Mahasiswa Kurang Mampu," *J. Adm. Educ. Manag.*, vol. 5, no. 1, pp. 98–107, 2022, doi: 10.31539/alignment.v5i1.3803.
- [3] D. T. Yuliana, M. I. A. Fathoni, and N. Kurniawati, "Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah Dengan Menggunakan Metode K-Means Clustering," *J. Focus Action Res. Math. (Factor M)*, vol. 5, no. 1, pp. 127–141, 2022, doi: 10.30762/f_m.v5i1.570.
- [4] P. Dana et al., "DAN DAMPAKNYA TERHADAP PENDIDIKAN DI INDONESIA," vol. 7, no. 1, pp. 28–36, 2024.
- [5] * Ciek et al., "Analisis Fenomena Hedonisme di Kalangan Mahasiswa Penerima Beasiswa KIP Kuliah: Perspektif Sosio-Ekonomi Baru," *J. Penelit. Mhs.*, vol. 3, no. 2, pp. 2962–116, 2024, [Online]. Available: <https://doi.org/10.58192/populer.v3i2.2134>
- [6] Rahmania Mustaqilillah, Okky Widyaningtyas, and Tri Wantoro, "Efektivitas Penggunaan Twitter Sebagai Sarana Peningkatan Berpikir Kritis Mahasiswa Ilmu Komunikasi," *MUKASI J. Ilmu Komun.*, vol. 2, no. 1, pp. 18–28, 2023, doi: 10.54259/mukasi.v2i1.1346.
- [7] M. Ismail, "Studi Literatur: Pengertian, Ciri, dan Teknik Pengumpulan Data." [Online]. Available: <https://ismailview.com/studi-literatur-pengertian-ciri-dan-teknik-pengumpulan-data/#>
- [8] H. Satria, "cara Crawl Data Twitter Thread atau Komentar." [Online]. Available: <https://helmisatria.com/blog/cara-crawl-data-twitter-thread-komentar>
- [9] Kuncahyo Setyo Nugroho, "Dasar Text Preprocessing dengan Python," <https://ksnugroho.Medium.Com/Dasar-Text-Preprocessing-Dengan-Python-a4Fa52608Ffe>. Accessed: Jul. 23, 2024. [Online]. Available: <https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe>
- [10] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [11] A. Aziz, "Analisis Sentimen Identifikasi Opini Terhadap Produk, Layanan dan Kebijakan Perusahaan Menggunakan Algoritma TF-IDF dan SentiStrength," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 6, no. 1, p. 115, 2022.
- [12] M. F. Azhari and R. Fajriyah, "IDEKTIFIKASI GEN MARKER PBMCS ISCHEMIC STROKE MENGGUNAKAN," vol. XIII, no. 1, pp. 73–81, 2024.
- [13] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 650, 2020, doi: 10.30865/mib.v4i3.2181.
- [14] D. Ananda and R. R. Suryono, "Analisis Sentimen Publik Terhadap Pengungsi Rohingya di Indonesia dengan Metode Support Vector Machine dan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 8, no. April, pp. 748–757, 2024, doi: 10.30865/mib.v8i2.7517.
- [15] A. F. Setyaningsih, D. Septiyani, and S. R. Widiyari, "Implementasi Algoritma Naïve Bayes untuk Analisis Sentimen Masyarakat pada Twitter mengenai Kepopuleran Produk Skincare di Indonesia," *J. Teknol. Inform. dan Komput.*, vol. 9, no. 1, pp. 224–235, 2023, doi: 10.37012/jtik.v9i1.1409.