

ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP KASUS DANA HIBAH JAWA TIMUR MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM)

Oleh:

Tirta Arya Bimantoro

Yunianita Rahmawati

Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

Juli, 2025



Pendahuluan

- Kasus dugaan penyalahgunaan dana hibah di Jawa Timur memicu reaksi publik di YouTube. Media ini kini menjadi wadah masyarakat menyampaikan opini terhadap isu sosial-politik secara terbuka.
- Komentar publik dapat dianalisis melalui *sentiment analysis* menggunakan algoritma Support Vector Machine (SVM), yang terbukti efektif dalam klasifikasi opini. Namun, masih minim penelitian yang fokus pada sentimen publik terhadap kasus korupsi di tingkat daerah. Penelitian ini bertujuan mengisi kekosongan tersebut.

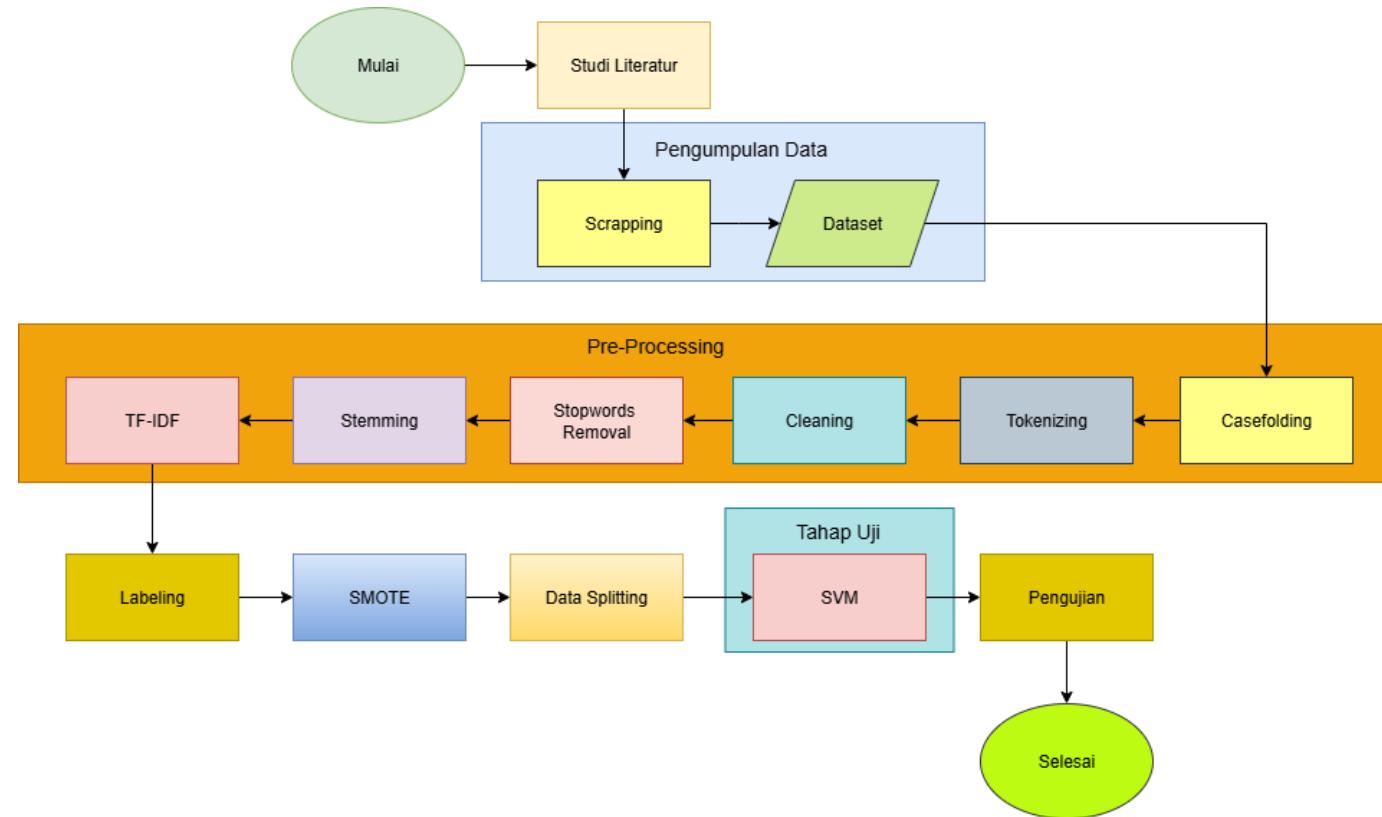
Pertanyaan Penelitian (Rumusan Masalah)

1. Bagaimana persepsi atau sentimen masyarakat terhadap kasus korupsi dana hibah Jawa Timur yang diungkapkan melalui komentar di platform YouTube?
2. Bagaimana proses klasifikasi sentimen komentar YouTube terkait kasus korupsi dana hibah Jawa Timur menggunakan algoritma Support Vector Machine (SVM)?



Metode

- Flowchart



Metode

- Hasil pengumpulan data disimpan ke file csv.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Video_URL	Komentar													
2		https://www.youtube.com/watch?v=UqMGrOcc0QM&t=73s,Jatim cuma 9M saja kapan giliran jabar 45M cuma 1 yayasan saja. Di 2021 mencapai triliunan kpk harus se													
3		https://www.youtube.com/watch?v=UqMGrOcc0QM&t=73s,dftr hey kota ngawi dn													
4		https://www.youtube.com/watch?v=UqMGrOcc0QM&t=73s,*LANGGANAN KORUPSI*													
5		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,Jawa hama : korupsiðŸ˜,													
6		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,"HALAH! Apa kabar MEGA KORUPSI PERTAMINA?													
7		Yang dipush kasus lain mulu, perasaan ga sebanding daripada kasus MEGA KORUPSI PERTAMINA deh ya?													
8		Kasus selingkuhnya RK, Tudingan pencitraan KDM, Ijazah JKW, Asusila oknum Kedokteran, ada-ada aja pengalihan issuenya ðŸ˜œ													
9		Tindakan dari yang diberi kewenangan dikemanakan? Perut kalian banyak yg buncit dari duit kami sebagai rakyat miskin yang dimiskinkan oleh PARA CALON PENGHUNI													
10		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,"KPK ITU SALAH ALAMAT.													
11		GUBERNURE SING GEBLEK ITU HARUSNYA SEGERA DITANGKAP, PASTI SEMUA AKAN TERUNGKAP.													
12		KERJAMU KIE OPO TO PAH KOMPIPAH,....."													
13		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,Jatim gae bancaan koruptorðŸ˜,ðŸ˜,ðŸ˜, piye bu gubernurnya													
14		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,Knp Dana Hibah ke Jatim banyak bermasalah ... Salah siapa.????													
15		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,Yg dulu merasa aman kini duduk pun tak tenang yo rasakno kusus jatim yg dulunya ayem2 saja													
16		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,Berantas semua penghianat negeri ini monyet2 dableg rakus dan tamak													
17		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,"Apes apes wayae bu Rizma sing dadi gubernur,eh kok malah oleh kofifahit"													
18		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,Di era jokowi bobork rakyat bawa masi di rampok demi kekuasan.itula sekma ponsi sitem poramida													
19		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,Skeptis aja wong kpk punya jokowi ?													
20		https://www.youtube.com/watch?v=UqMGrOcC0QM&t=73s,Dwet. Pajek do. Go Bancakan. Deso. Deso. Do. Ra di urus. Karo pejabate. Pejabate. Sibuk. Golek dwet. Jln													



Metode

Pre-processing

1. Casefolding

- Semua huruf dalam komentar diubah menjadi huruf kecil (*lowercase*) guna menghindari perbedaan penulisan pada kata yang sama namun ditulis dengan kapitalisasi berbeda.

Casefolding	
jatim cuma 9m saja kapan giliran jabar 45m cum...	
	dftr hey kota ngawi dn
	langganan korupsi
	jawa hama : korupsi 😊
halah! apa kabar mega korupsi pertamina?\r\nnya...	
	kpk itu salah alamat.\n gubernure sing geblek i...



Metode

2. Tokenizing

- Proses ini memisahkan kalimat atau paragraf menjadi unit kata tunggal (*tokens*). Tokenizing penting untuk memecah informasi menjadi elemen-elemen dasar yang akan dianalisis.

Tokenizing

[jatim, cuma, 9m, saja, kapan, giliran, jabar,...]

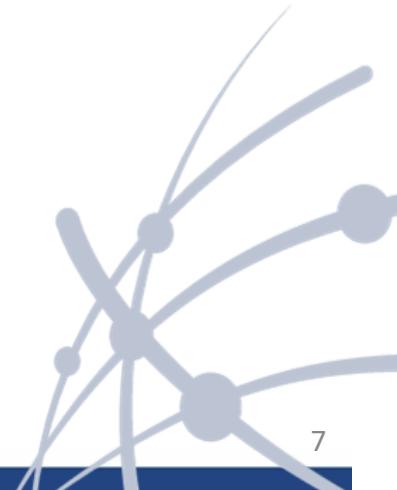
[dftr, hey, kota, ngawi, dn]

[*, langganan, korupsi, *]

[jawa, hama, :, korupsi😊]

[halah, !, apa, kabar, mega, korupsi, pertamin...]

[kpk, itu, salah, alamat, ., gubernure, sing, ...]



Metode

3. Cleaning

- Tahap ini melibatkan penghapusan karakter atau elemen yang tidak relevan dalam teks seperti tanda baca, angka, URL, mention (@username), hashtag, emotikon, serta karakter non-alfabet lainnya. Cleaning penting agar hanya informasi inti yang diproses lebih lanjut.

Cleaning

[jatim, cuma, m, saja, kapan, giliran, jabar, ...]

[dftr, hey, kota, ngawi, dn]

[langganan, korupsi]

[jawa, hama, korupsi]

[halah, apa, kabar, mega, korupsi, pertamina, ...]

[kpk, itu, salah, alamat, gubernure, sing, geb...]

Metode

4. Stopword Removal

- Stopword Removal adalah tahap pembersihan teks dengan cara menghilangkan kata-kata yang bersifat umum dan tidak memiliki pengaruh signifikan terhadap makna atau konteks utama dalam sebuah teks seperti “yang”, “di”, “dengan”, “pada”, “untuk”. Stopword dihapus agar model hanya fokus pada kata-kata bermakna utama. Daftar stopword bahasa Indonesia dapat diambil dari pustaka NLP seperti Sastrawi atau NLTK.

Stopwords Removal

[jatim, m, giliran, jabar, m, yayasan, mencapa...]

[dftr, hey, kota, ngawi, dn]

[langganan, korupsi]

[jawa, hama, korupsi]

[halah, kabar, mega, korupsi, pertamina, dipus...]

[kpk, salah, alamat, gubernure, sing, geblek, ...]



Metode

5. Stemming

- Proses stemming mengubah kata turunan menjadi bentuk dasarnya. Misalnya, kata "menyalahgunakan", "penyalahgunaan", dan "disalahgunakan" semuanya akan distem menjadi "salahguna". Teknik ini membantu mengurangi dimensi fitur dan menyatukan kata dengan akar yang sama. Proses ini dilakukan menggunakan stemmer bahasa Indonesia seperti Sastrawi.

stemming

[jatim, m, gilir, jabar, m,
yayasan, capai, tr...]

[dftr, hey, kota, ngawi, dn]

[langgan, korupsi]

[jawa, hama, korupsi]

[halah, kabar, mega,
korupsi, pertamina,
dipus...]

[kpk, salah, alamat,
gubernure, sing, geblek,

Metode

6. TF-IDF

- Setelah teks dibersihkan dan distem, langkah selanjutnya adalah mengubah kumpulan kata diubah menjadi vektor numerik. Pendekatan Term Frequency-Inverse Document Frequency (TF-IDF) dimanfaatkan untuk menetapkan bobot tinggi pada kata-kata yang sering muncul dalam satu komentar, tetapi jarang ditemukan di seluruh kumpulan data. Strategi ini bertujuan untuk mengidentifikasi kata-kata yang paling mewakili isi masing-masing dokumen. Nilai TF-IDF inilah yang kemudian digunakan sebagai fitur dalam model klasifikasi

hasil TF-IDF (10 komentar pertama):

	aamiin	aamin	abah	abah	abangkuh	abangkuuh	abis	abiz	abs	...	yme	ying	yo	yq	yra	yth	yuk	zet	zubair	zulhas
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.309361	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0

10 rows x 2807 columns

Hasil TF-IDF disimpan sebagai 'hasil_tfidf.csv'

Hasil

- ✗ Evaluasi Model: SVM (Support Vector Machine)
- ✓ Akurasi SVM: 0.9426229508196722

Model SVM menghasilkan akurasi sebesar 94,26%, yang berarti model berhasil mengklasifikasikan 94% data uji dengan benar. Nilai ini menunjukkan performa yang cukup baik secara keseluruhan, namun tetap perlu didukung oleh metrik lain seperti precision, recall, dan f1-score untuk evaluasi yang lebih menyeluruh.

Hasil

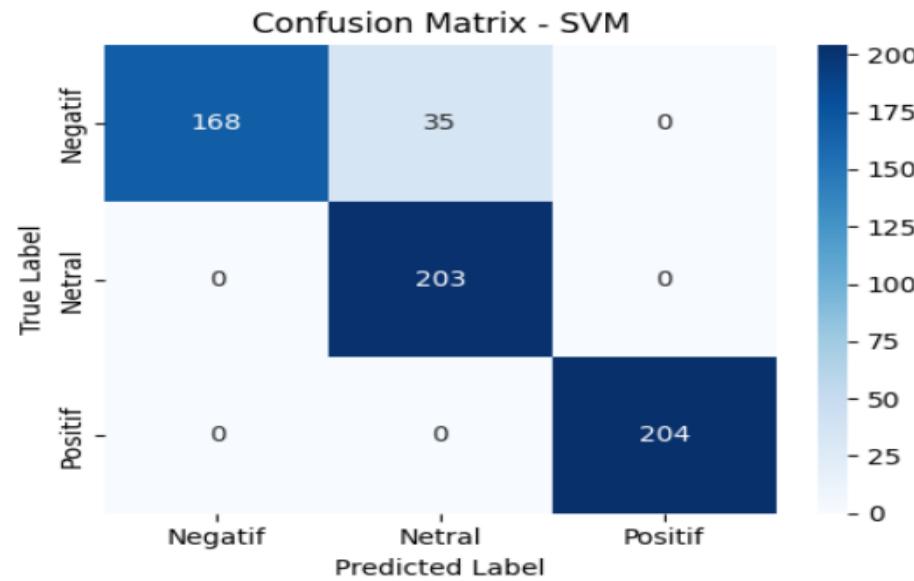
Classification Report:

	precision	recall	f1-score	support
-1	1.00	0.83	0.91	203
0	0.85	1.00	0.92	203
1	1.00	1.00	1.00	204
accuracy			0.94	610
macro avg	0.95	0.94	0.94	610
weighted avg	0.95	0.94	0.94	610

Hasil classification report menunjukkan bahwa model SVM memiliki performa sangat baik, dengan akurasi sebesar **94,26%**. Kelas **positif** diklasifikasikan sempurna (precision dan recall 1.00), sementara kelas **netral** memiliki recall tinggi (1.00) namun precision lebih rendah (0.78), karena beberapa prediksi berasal dari kelas lain. Kelas **negatif** juga memiliki precision tinggi (1.00), namun recall-nya 0.83 akibat sebagian komentar negatif terklasifikasi sebagai netral. Secara keseluruhan, nilai macro average F1-score tinggi, mencerminkan kinerja model yang kuat dan seimbang dalam mengenali ketiga kelas sentimen.



Hasil



Confusion matrix menunjukkan kinerja model SVM dalam mengklasifikasikan sentimen komentar YouTube ke dalam tiga kelas: **Negatif**, **Netral**, dan **Positif**. Model berhasil mengklasifikasikan sebagian besar data dengan benar, terutama pada kelas **Netral** dan **Positif** yang keduanya mencapai prediksi sempurna tanpa kesalahan. Untuk kelas **Negatif**, sebanyak 168 data terkласifikasi dengan benar, sementara 35 lainnya salah diklasifikasikan sebagai netral. Secara keseluruhan, model memperoleh **akurasi sebesar 94,26%**, yang menunjukkan performa klasifikasi yang sangat baik dan akurat dalam menangkap persepsi publik.

Manfaat Penelitian

- Mengungkap Persepsi Publik Secara Otomatis
- Dukungan Pengambilan Keputusan Pemerintah
- Pengembangan Sistem Pemantauan Digital



www.umsida.ac.id



[umsida1912](https://www.instagram.com/umsida1912)



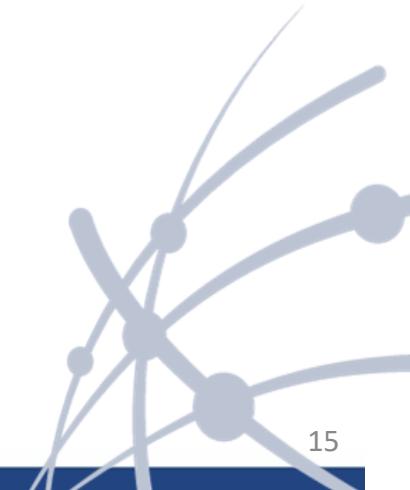
[umsida1912](https://twitter.com/umsida1912)



universitas
muhammadiyah
sidoarjo



[umsida1912](https://www.youtube.com/umsida1912)



Referensi

- [1] E. Tohidi, R. Perdana Herdiansyah, And E. Wahyudin, “Analisa Sentimen Komentar Video Youtube Di Channel Tvonenews Tentang Calon Presiden Prabowo Subianto Menggunakan Support Vector Machine,” 2024. [Online]. Available: Https://Www.Youtube.Com/Watch?V=Wu_Wqhg1ylk
- [2] R. T. Adek, Z. Fitri, And S. Chairani Siegar, “Analisis Sentimen Komentar Pada Saluran Youtube Beauty Vlogger Berbahasa Indonesia Menggunakan Metode Support Vector Machine,” *Jurnal Algoritme*, Vol. 5, No. 2, Pp. 164–175, 2025, Doi: 10.35957/Algoritme.V5i2.9692.
- [3] Desti Mualfah, “Analisis Sentimen Komentar Youtube Tvincen Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma Svm,” 2023.
- [4] R. R. Putri And N. Cahyono, “Analisis Sentimen Komentar Masyarakat Terhadap Pelayanan Publik Pemerintah Dki Jakarta Dengan Algoritma Super Vector Machine Dan Naive Bayes,” 2024.
- [5] F. Caroline, R. G. S. Budi, And M. E. Al Rivan, “Analisis Sentimen Masyarakat Terhadap Kasus Korupsi Pt. Timah Menggunakan Metode Support Vector Machine,” *Jurnal Ilmu Komputer Dan Informatika*, Vol. 4, No. 1, Pp. 43–50, Jul. 2024, Doi: 10.54082/Jiki.141.
- [6] P. Fremmuzar And A. Baita, “Uji Kernel Svm Dalam Analisis Sentimen Terhadap Layanan Telkomsel Di Media Sosial Twitter,” *Komputika : Jurnal Sistem Komputer*, Vol. 12, No. 2, Pp. 57–66, Sep. 2023, Doi: 10.34010/Komputika.V12i2.9460.
- [7] Hayati, “Analisis Sentimen Komentar Pengguna Youtube Terhadap Kebijakan Baru Badan Penyelenggara Jaminan Kesehatan Sosial Menggunakan Naïve Bayes,” 2024.
- [8] A. Irma Purnamasari And I. Ali, “Analisis Sentimen Komentar Berita Detik.Com Menggunakan Algoritma Suport Vektor Machine (Svm),” 2024.



Referensi

- [9] N. Kenneth, "Maraknya Kasus Korupsi Di Indonesia Tahun Ke Tahun," *Nathanael Kenneth*, Vol. 2, No. 1, 2024.
- [10] F. F. Abdulloh And I. R. Pambudi, "Analisis Sentimen Pengguna Youtube Terhadap Program Vaksin Covid-19," *Csrid (Computer Science Research And Its Development Journal)*, Vol. 13, No. 3, P. 141, Nov. 2021, Doi: 10.22303/Csrid.13.3.2021.141-148.
- [11] R. Asrianto And M. Herwinanda, "Analisis Sentimen Kenaikan Harga Kebutuhan Pokok Dimedia Sosial Youtube Menggunakan Algoritma Support Vector Machine," *Jurnal Coscitech (Computer Science And Information Technology)*, Vol. 3, No. 3, Pp. 431–440, Dec. 2022, Doi: 10.37859/Coscitech.V3i3.4368.
- [12] Aprilia And A. R. Isnain, "Analisis Sentimen Terhadap Media Sosial Twitter Dengan Kasus Kampanye Anti-Korupsi Di Indonesia Menggunakan Naive Bayes," *Jurnal Media Informatika Budidarma*, Vol. 8, No. 2, P. 695, Apr. 2024, Doi: 10.30865/Mib.V8i2.7582.
- [13] Muhammad Hilmy Riwanto, "Analisis Sentimen Komentar Youtube Terkait Penerapan Makan Bergizi Gratis Menggunakan Model Algoritma Svm," 2025.
- [14] L. Rofiqi And M. Akbar, "Analisis Sentimen Terkait Ruu Perampasan Aset Dengan Support Vector Machine," *Jekin - Jurnal Teknik Informatika*, Vol. 4, No. 3, Pp. 529–538, Aug. 2024, Doi: 10.58794/Jekin.V4i3.824.
- [15] H. Hidayat, F. Santoso, And L. F. Lidimillah, "Analisis Sentimen Pengguna Youtube Tentang Rohingya Menggunakan Algoritma Svm (Support Vector Machine)," *G-Tech: Jurnal Teknologi Terapan*, Vol. 8, No. 3, Pp. 1729–1738, Jul. 2024, Doi: 10.33379/Gtech.V8i3.4497.



