

PREDIKSI SCAMMER PADA PLATFORM MEDIA SOSIAL DENGAN PENDEKATAN ANOMALY DETECTION

GAYUH PUJI RAHAYU¹, RAHMAT FAUZI²

¹MAHASISWI PROGRAM STUDI TEKNIK INFORMATIKA, UNIVERSITAS PUTERA BATAM

²DOSEN PROGRAM STUDI TEKNIK INFORMATIKA, UNIVERSITAS PUTERA BATAM

email: pb200210037@upbatam.ac.id

ABSTRACT

Peranan media sosial sangat berpengaruh dalam kehidupan digital ini. Begitu banyak kegiatan yang dibagikan langsung oleh sesama pengguna di media sosial, mulai dari kegiatan sehari-hari, mengungkapkan ekspresi, berjualan dan kegiatan lain yang berkaitan dengan media massa. Keuntungan dan manfaat dari media sosial banyak dirasakan oleh penggunanya, namun dibalik itu semua terdapat beberapa oknum yang menjadikan media sosial sebagai wadah dalam melancarkan aksinya untuk melakukan kejahatan digital atau scamming. Oleh karena itu, penelitian ini dilakukan agar dapat mengantisipasi hal tersebut terjadi serta menambah wawasan dan pengetahuan terutama kepada pengguna media sosial untuk lebih waspada dan peduli terhadap keamanan software dan hardware yang mereka gunakan.

Keywords : *Anomaly Detection, Media sosial, Scamming, Support Vector Machine (SVM), Text Mining*

PENDAHULUAN

Di era digital ini, media sosial tidak hanya digunakan untuk hal-hal yang bermanfaat terdapat beberapa oknum yang menyalahgunakannya untuk keuntungan pribadi dan merugikan orang lain. Banyak kejahatan digital yang terjadi media sosial seperti *Scamming, Spoofing, Phishing* dan kejahatan lainnya. Oleh karena itu, penelitian ini dilakukan untuk mengantisipasi masalah tersebut dan mencegah dari awal kejahatan terutama *scammer* agar tidak terjadi. Penelitian ini bertujuan untuk mendeteksi lebih dini *scammer* dengan pendekatan *Anomaly Detection*. Dalam mendeteksi *scammer* diperlukan beberapa metode dalam

mengklasifikasi faktor yang dinilai menjadi cikal bakal *scammer* tersebut. Dalam menyelesaikan permasalahan ini peneliti menggunakan beberapa metode yang lain yaitu Algoritma *K-Means* dalam menyaring data uji dari sosial media serta *Support Vector Machine (SVM)* dalam mengelola data latih. Pada penelitian lain disebutkan bahwa *Support Vector Machine (SVM)* mempunyai nilai akurasi terbaik¹. (Mueller at al 2020), Meneliti peran *emotional intelligence* dan menunjukkan bahwa skor yang lebih tinggi terhadap minat dalam menanggapi permintaan penipuan adalah orang dewasa yang usianya lebih tua². Dengan adanya penelitian ini penulis berharap

para pengguna sosial media lebih waspada dan peduli terhadap media serta perangkat yang mereka gunakan dalam kegiatan sehari-hari. Penelitian ini juga diharapkan dapat menambah wawasan dan pengetahuan kepada masyarakat awam yang belum familiar dengan kejahatan digital.

KAJIAN TEORI

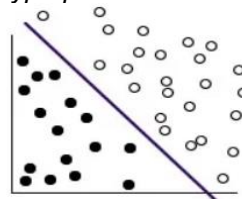
2.1 Scamming

scam adalah kegiatan yang dilakukan oleh seseorang maupun sekelompok bertujuan untuk memanfaatkan serta menipu seseorang dengan cara kerja yang tidak diketahui dan merugikan korban melalui sosial media, korban dan pelakunya bisa siapa saja orang lokal bahkan orang asing³.

2.2 Anomaly Detection

Anomaly detection bersifat *heterogen* yaitu kelas *anomaly* dapat menunjukkan karakteristik yang berbeda dan menunjukkan kelainan dari kelas *anomaly* yang lain. Dalam menganalisis *anomaly*

yang terdapat pada dataset, penelitian ini menggunakan 2 metode yaitu dengan Algoritma *K-Means* dan *Support Vector Machine (SVM)*. Cara kerja Algoritma *K-Means* adalah dengan membagi objek data pada beberapa kelompok atau *cluster* yang berbeda dan mempunyai tingkat kesamaan terkecil (Xie dkk, 2020). Sedangkan cara kerja SVM adalah dengan mengklasifikasikan data dan mengelompokkannya, memisahkannya dengan *hyperplane*.⁴



Gambar 1 SVM Hyperplane

2.3 Text Mining

Text mining memiliki tujuan yaitu menemukan info yang sangat berharga dan tersembunyi diantara sumber yang terstruktur maupun tidak terstruktur.

English

Negative	Indonesia-Negatif	English-Positive	Indonesia-Positif
2-faced	bermuka dua	a+	
2-faces	2 wajah	abound	berlimpah
abnormal	abnormal	abounds	berlimpah
abolish	menghapuskan	abundance	kelimpahan
abominable	mengerikan	abundant	melimpah
abominably	sangat buruk	accessible	dapat diakses
abominate	merasa jijik	accessible	dapat diakses
abomination	kekejangan	acclaim	tepuk tangan sorak
abort	menggugurkan	acclaimed	terkenal

Tabel 1 Text Mining

2.4 Python

Python merupakan pemrograman komputer yang biasa digunakan dalam membangun dan membuat situs, *software* maupun aplikasi. Berdasarkan *Stackoverflow 2023 python* berhasil masuk dalam 10 besar menjadi bahasa pemrograman yang populer digunakan.⁵

2.5 Machine Learning

Machine learning adalah ilmu komputer yang memungkinkan bisa berperilaku seperti layaknya manusia, Komputer diharapkan dapat meningkatkan pemahamannya didasarkan pengalaman dan berjalannya waktu dengan otomatis.⁶

memahami kasus dengan mencari tahu penyebab terbesar hal tersebut dapat terjadi.

2. Pengumpulan data

Memanfaatkan media sosial melalui beberapa kasus yang terkait dengan kejahatan digital dan difokuskan pada topik scamming dan khususnya pada platform whatsapp dan telegram.

3. Perancangan

Penulis membangun suatu program yang dapat menyaring data-data yang dianggap anomaly dengan mendeteksi kata negative pada file yang diinput.

4. Pengujian

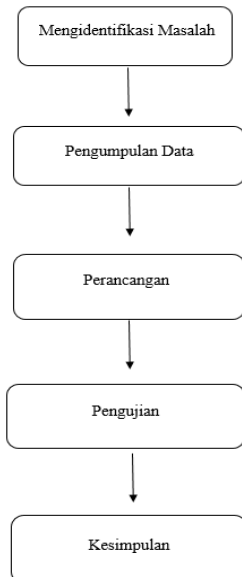
Setelah program selesai, penulis kemudian akan menguji dengan ,menginput file dengan ekstensi TXT dan JSON agar dapat diproses oleh program. Program akan mengeluarkan output sesuai dengan file apabila file yang diinput terdapat banyak teks negatif maka akan terdeteksi *scammer*.

5. Kesimpulan

Di Tahap terakhir ini penulis berharap program dapat dimanfaatkan oleh para pengguna media sosial sesuai dengan yang dibutuhkan.

METODE PENELITIAN

3.1 Desain Penelitian



Gambar 3 Desain Penelitian

1. Identifikasi masalah

Peneliti menemukan terdapatnya kasus kejahatan digital yang sering terjadi pada media sosial. Peneliti mencoba

HASIL DAN PEMBAHASAN

Data yang dikumpulkan kemudian diolah dengan menggunakan metode, didapatkan lah hasil yang dapat dijalankan yang sesuai dengan kriteria program. Pada Algoritma *K-Means* hasil data yang digunakan tersebut fokus pada platform yang digunakan karena berdasarkan perhitungan dengan menggunakan rumus *Euclidean distance* dan hasilnya adalah platform Whatsapp dan Telegram.

Data	Media sosial	Jumlah	Normal	Terduga Scam	Cluster
1.	Facebook	150	93	23	Rendah
2.	Instagram	100	87	12	Rendah
3.	Tiktok	50	39	6	Rendah
4.	Twitter	100	76	18	Rendah
5.	Telegram	320	278	56	Tinggi
6.	Whatsapp	280	243	29	Sedang

Gambar 4 K-Means Media sosial

Setelah dilakukan proses hitung dengan menggunakan algoritma K-Means tepatnya dengan menggunakan rumus Euclidean distance, didapatkanlah hasil seperti gambar diatas ini yang menjelaskan

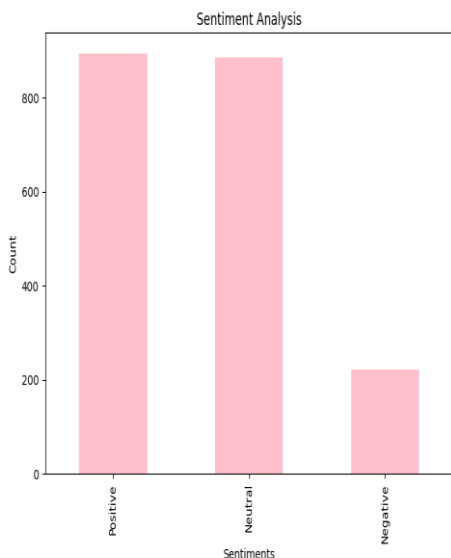
penulis akhirnya menggunakan platform Telegram dan Whatsapp. Sedangkan pada metode Support Vector Machine data yang diolah kemudian ditampilkan dalam bentuk bar grafik seperti gambar dibawah.

	Text
0	@kunalb11 I'm an alien
1	@ID_AA_Carmack Ray tracing on Cyberpunk with H...
2	@joerogan @Spotify Great interview!
3	@gtera27 Doge is underestimated
4	@teslacn Congratulations Tesla China for amazing execution last year. Now on to the next for even more!!
5	Happy New Year of the Ox! https://t.co/9WFKMYu2oj

Gambar 5 Data Latih

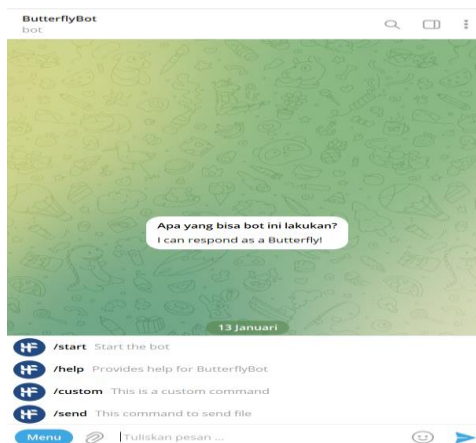
Diatas merupakan data yang mentah sebelum melalui proses *support vector machine* (SVM), Pada metode SVM banyak proses yang dilalui oleh data prosesnya seperti Load data, Uncleaned data, Load stop word, Data cleaning.

setelah melakukan proses tersebut maka didapatkanlah hasil seperti dibawah ini.



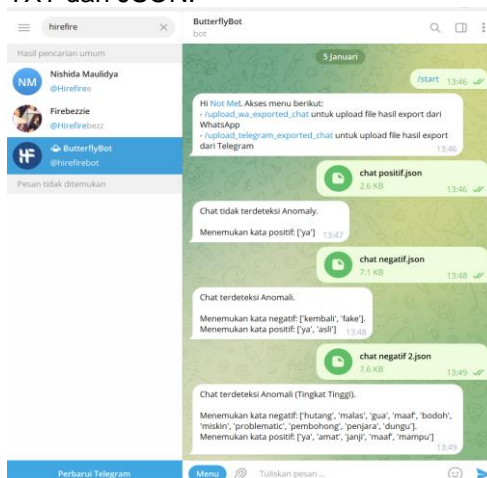
Gambar 6 Analysis text

Berdasarkan data grafik bar diatas, hasil persentase angka yang didapatkan adalah sebagai berikut :
 Negative Tweets : 10,4 %
 Positive Tweets : 36,1%
 karena yang digunakan hanyalah data dari teks positif dan negative, jadi persentase angka untuk netral tidak dicantumkan dan sesuai data di grafik bar diketahui hasil positif dan netral sama. Setelah data latih dan data uji didapatkan, maka data tersebut akan digunakan kedalam program untuk diinput agar dapat mengeluarkan output sesuai dengan yang diharapkan. Data hanya dapat dijalankan dengan ekstensi TXT dan JSON, hal ini dikarenakan untuk melindungi hardware dan media yang digunakan untuk memproses prediksi scammer tersebut.



Gambar 7 Menu Telebot

Telebot merupakan platform pilihan penulis sebagai media untuk media penginputan ekspor chat yang akan dieksekusi oleh program. Gambar diatas merupakan tampilan menu yang disediakan dan dapat di proses oleh telebot tersebut. Pada menu tersebut terdapat pesan yang menyampaikan kepada user apabila ingin menginput data hanya dapat dieksekusi dengan ekstensi TXT dan JSON.



Gambar 8 Chat Ekspor

Gambar diatas merupakaan salah satu hasil yang keluar apabila user telah menginput eksor chat yang terduga scam. Output yang keluar.

SIMPULAN

Berdasarkan penelitian yang telah dilakukan dan program yang telah dirancang, ditariklah beberapa poin kesimpulan seperti berikut :

1. Telebot yang dibuat berfungsi untuk mendeteksi teks negative dan positif yang terdapat pada chat ekspor yang diinput. Jumlah dan kata yang terdapat pada chat tersebut akan keluar sebagai output.
2. Teks negative dan positif tersebut digunakan sebagai data latih pada program, dan media sosial atau data ekspor digunakan sebagai data uji. Data uji yang dijalankan deprogram hanya dapat dijalankan dengan ekstensi TXT dan JSON.
3. Program yang dibuat diharapkan dapat bermanfaat bagi orang yang awam mengenai scam dan kejahatan lainnya serta menambah wawasan dan dapat dikembangkan dikemudian hari oleh programmer dengan fitur yang lebih maju dan *friendly*.

DAFTAR PUSTAKA

W, Anggraini. Utami. M BJ. Klasifikasi Sentimen Masyarakat Terhadap Kebijakan Kartu Prakerja di Indonesia. *homepage*. 2020;13(4).

https://www.researchgate.net/publication/357025470_KLASIFIKASI_SENTIMEN_MASYARAKAT_TERHADAP_KEBIJAKAN_KARTU_

PEKERJA_DI_INDONESIA

Yaniv HWS. The Scams Among Us: Who Falls Prey and Why. *Pshylogical Sci*. 2021;30.

doi:<https://doi.org/10.1177/0963721421995489>

Sultan. Muhammad SD. PENYULUHAN PENCEGAHAN TINDAKAN SCAM PADA TEKNOLOGI BERSAMA MASYARAKAT DURENSARI BOJONGSARI DEPOK. *Kreat Mhs Inform*. 2021;2.

<https://openjournal.unpam.ac.id/index.php/JATIMIKA/article/viewFile/14325/9305>

Canteenit. Support Vector Machine (SVM): Definition, Types and How It Works. <https://kantinit.com>. Published 2022. Accessed January 10, 2024.

[https://kantinit.com/kecerdasan-buatan/support-vector-machine-svm-pengertian-jenis-dan-cara-kerja/#Pengertian Support Vector Machine](https://kantinit.com/kecerdasan-buatan/support-vector-machine-svm-pengertian-jenis-dan-cara-kerja/#Pengertian%20Support%20Vector%20Machine)

Shahid S. Top Programming Languages in 2023: Insights from the Stack Overflow Developer Survey.

<https://medium.com>. Published 2023. Accessed January 15, 2024.

<https://medium.com/@shafaqgillani8/top-programming-languages-in-2023-insights-from-the-stack-overflow-developer-survey-8bbb8e490a15>

Kusuma DPD. *MACHINE LEARNING Teori, Program Dan Studi Kasus*. DEEPUBLISH; 2020.

<https://books.google.co.id/books?hl=id&lr=&id=4k3sDwAAQBAJ&oi>

	<p>Biodata Penulis pertama, Gayuh Puji Rahayu, merupakan Mahasiswi Prodi Teknik Informatika Universitas Putera Batam. Email : pb200210037@upbatam.ac.id</p>
	<p>Biodata Penulis kedua, Rahmat Fauzi S.Kom., M.Kom. merupakan Dosen Prodi Teknik Informatika Universitas Putera Batam. Email : fa_u_zy@yahoo.com</p>