DETEKSI MALWARE RANSOMWARE BERDASARKAN PANGGILAN API SISTEM OPERASI WINDOWS

Ransomware Detection Based On Windows Operating System API Call

HARTINAH
D082202010

PROGRAM STUDI S2 TEKNIK INFORMATIKA
DEPARTEMEN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2023

PENGAJUAN TESIS

DETEKSI MALWARE RANSOMWARE BERDASARKAN PANGGILAN API SISTEM OPERASI WINDOWS

Tesis

Sebagai Salah Satu Syarat untuk Mencapai Gelar Magister

Program Studi Teknik Informatika

Disusun dan diajukan oleh

HARTINAH
D082202010

Kepada

FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2023

TESIS

DETEKSI MALWARE RANSOMWARE BERDASARKAN PANGGILAN API SISTEM OPERASI WINDOWS

HARTINAH
D082202010

Telah dipertahankan di hadapan Panitia Ujian Tesis yang dibentuk dalam rangka penyelesaian studi pada Program Magister Teknik Informatika Fakultas Teknik Universitas Hasanuddin
pada tanggal 17 Mei 2023
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama	Pembimbing Pendamping
Dr. Eng. Ady Wahyudi Paundu, ST., MT.	**Dr.Amil Ahmad Ilham, ST., M.IT.**
NIP. 19750313 200912 1 003	NIP. 19731010 199802 1 001

Dekan Fakultas Teknik Universitas Hasanuddin	Ketua Program Studi S2 Teknik Informatika
Prof. Dr. Eng. Ir. Muhammad Isran Ramli, ST., MT	**Dr. Ir. Zahir Zainuddin, M.Sc**
NIP. 19730926 200012 1 002	NIP. 19640427 198910 1 002

PERNYATAAN KEASLIAN TESIS
DAN PELIMPAHAN HAK CIPTA

Yang bertanda tangan di bawah ini
Nama : Hartinah
Nomor Mahasiswa : D082202010
Program Studi : S2 Teknik Informatika

Dengan ini menyatakan bahwa, tesis berjudul "Deteksi Malware Ransomware Berdasarkan Panggilan API Sistem Operasi Windows" adalah karya saya dengan arahan dari komisi pembimbing (Dr. Eng. Ady Wahyudi Paundu, ST., MT., dan Dr. Amil Ahmad Ilham, ST., M.IT). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasa atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam daftar pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal (Jurnal Edukasi dan Penelitian Informatika (JEPIN) tahun 2023) dengan judul "Deteksi Malware Ransomware Berdasarkan Panggilan API dengan Metode Ekstraksi Fitur N-gram dan TF-IDF".

Dengan ini saya limpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin

Gowa, 17 Mei 2023
Yang menyatakan

Hartinah

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh

Puji Syukur Senantiasa kita panjatkan kehadirat Allah *Subhanahu Wata'ala* yang telah memberikan rahmat, hidayah dan pertolongan-Nya dalam menyelesaikan penulisan tesis dengan judul **"Deteksi Malware Ransomware Berdasarkan Panggilan API Sistem Operasi Windows"** dapat terselesaikan dengan baik. Sholawat serta Salam tidak lupa kita senantiasa kirimkan kepada junjungan Nabi Besar Muhammad *Shallallahu 'Alaihi Wasallam* yang menjadi suri tauladan bagi seluruh umat manusia.

Tesis ini disusun untuk memenuhi persyaratan untuk memperoleh gelar **Magister Komputer (M.Kom)** pada program S2 Teknik Informatika, Departemen Teknik Infomatika, Universitas Hasanuddin, Makasar. Dengan mengucapkan syukur yang sedalam-dalamnya, gelar ini penulis persembahkan kepada kedua orang tua tercinta, ayahanda **Hamzah** dan ibunda **Sapina** yang senantiasa memberikan dukungan baik dalam segi moral dan materil, motivasi dan doa yang tidak henti-hentinya dipanjatkan sehingga penulis dapat menyelesaikan penelitian ini dengan baik serta terimakasih pula kepada kelima adik dan sahabat tercinta, **Hardiayansyah**, **Harbi, Firdaus, Rahul, Ramzi,** dan **Andi Syarwani** yang senantiasa memberikan semangat dan dukungan kepada penulis.

Dalam penyusunan tesis ini, tentunya tidak lepas dari dukungan dari seluruh pihak. Dengan segala kerendahan hati, penulis menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Dewan pembimbing, Bapak Dr. Eng. Ady Wahyudi Paundu, ST., MT. dan Bapak Dr.Amil Ahmad Ilham, ST., M.IT, yang dengan sabar dan penuh tanggung jawab memberikan bimbingan, masukan serta motivasi kepada penulis dalam penyelesaian tesis ini.
2. Dewan Penguji, Bapak Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN.Eng, Bapak Dr-Eng. Ir. Muhammad Niswar, ST., M.InfoTech dan Ibu Mukarramah Yusuf, B.Sc., M.Sc., Ph.D, yang senantiasa memberikan saran yang membangun selama penelitian ini dilakukan.

3. Dosen dan staf Universitas Hasanuddin Makassar, khususnya program studi magister teknik informatika yang selalu memberikan dukungan dan motivasi untuk terus melanjutkan pendidikan.
4. Rekan-rekan mahasiswa S2 Teknik Informatika Universitas Hasanuddin yang selalu aktif dalam berdiskusi dan memberikan masukan dalam penyelesaian penelitian ini.

Penulis menyadari bahwa penyusunan tesis ini masih sangat jauh dari kata sempurna dan mempunyai banyak kekurangan. Sehingga penulis sangat mengharapkan kritik dan saran untuk kedepannya sehingga dapat memberikan manfaat bagi seluruh pembaca. Akhir kata, penulis menyampaikan permohonan yang sebesar-besarnya kepada pembaca sekiranya terdapat kesalahan-kesalahan dalam penyusunan tesis ini, wassalamualaikum warahmatullahi wabarakatuh.

Gowa, 17 Mei 2023

Hartinah

ABSTRAK

HARTINAH. Deteksi Malware Ransomware Berdasarkan Panggilan API Sistem Operasi Windows. (dibimbing oleh **Ady Wahyudi Paundu, Amil Ahmad Ilham**).

Jumlah serangan ransomware yang terus meningkat mengakibatkan kerugian yang tidak sedikit. Penanganan atas serangan ini semakin sulit dilakukan dikarenakan varian ransomware yang terus berkembang. Dibutuhkan suatu sistem yang mampu mendeteksi ransomware bahkan untuk varian ransomware terbaru. Penelitian ini dibuat untuk mendeteksi ransomware dan normalware menggunakan metode *machine learning* dengan memanfaatkan data panggilan *API* dari ransomware dan normalware. Dalam penelitian ini hanya dilakukan *binary classification* untuk semua varian ransomware yang terdeteksi. Proses ekstraksi fitur terlebih dilakukan dengan metode N-gram dan *TF-IDF* pada panggilan *API* untuk membentuk subset fitur yang digunakan dalam proses pembelajaran model. Pembuatan model deteksi dilakukan dengan melatih data panggilan *API* dari beberapa varian ransomware. Pengujian model dilakukan baik terhadap varian ransomware yang sudah dilatih sebelumnya maupun varian ransomware diluar data latih. Proses pembelajaran model dilakukan untuk mencari kesamaan fitur dari data panggilan *API* berbagai varian ransomware pada data latih, kesamaan fitur ini akan dimanfaatkan untuk mendeteksi varian lain dari ransomware diluar data latih. Hasil penelitian menunjukkan bahwa akurasi rata-rata model terhadap varian ransomware dalam data latih adalah 94% dengan skor *Error Rate* tertinggi 10%. Adapun hasil deteksi ransomware untuk varian diluar data latih menunjukkan akurasi rata-rata 83% dengan skor *Error Rate* tertinggi 30%. Berdasarkan hasil penelitian diketahui bahwa model yang dibuat mampu mendeteksi ransomware meskipun varian dari ransomware mengalami perkembangan.

Kata Kunci : Ransomware, Panggilan *API*, *Binary Classification*, N-gram, *TF-IDF*.

ABSTRACT

HARTINAH. *Ransomware Detection Based On Windows Operating System API Call. (Supervised by* **Ady Wahyudi Paundu, Amil Ahmad Ilham**)

The growing number of ransomware attacks results in significant losses. Handling this attack is increasingly difficult because of the ever-evolving ransomware variants. There is a need for a system capable of detecting ransomware, even for the latest ransomware variants. This study used machine learning methods to detect ransomware and normalware using API call data from ransomware and normalware. This study only performed binary classification for all detected ransomware variants. The feature extraction process is first carried out on API calls using the N-gram and TF-IDF methods to form a feature subset for the model learning process. Detection modelling is built by training API data calls from multiple ransomware variants. Model testing was performed on both previously trained and untrained ransomware variants. The model learning process is carried out to find similarities in the features of API call data for various ransomware variants in the training data. These feature similarities will be used to detect other ransomware variants outside of the training data. The results showed that the average accuracy of the model against the ransomware variant in the training data was 94%, with the highest Error Rate score of 10%. The ransomware detection results for variants outside the training data show an average accuracy of 83%, with the highest Error Rate score of 30%. Based on the research results, the model developed in this study can be used to detect ransomware even though the variant is still evolving.

Keywords : *Ransomware, API Calls, Binary Classification, N-gram, TF-IDF*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I

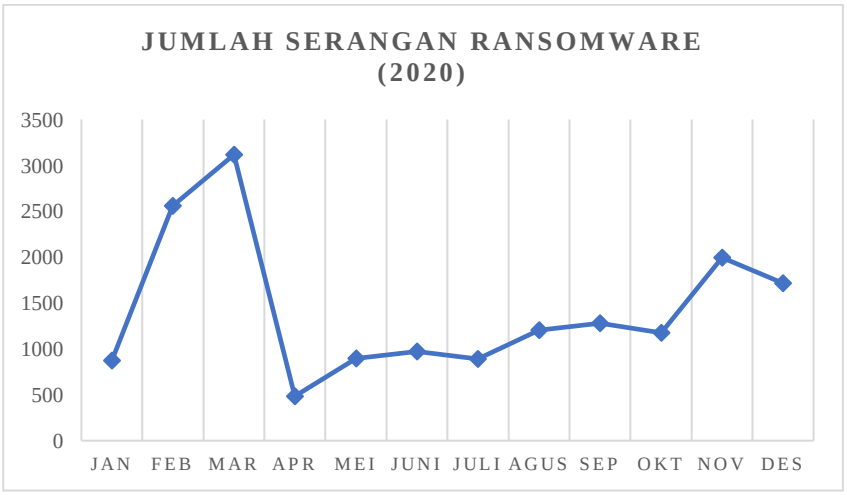
PENDAHULUAN

I.1 Latar Belakang

Malicious software (malware) merupakan program komputer yang diciptakan dengan tujuan mencari kelemahan sistem, merusak file dan sistem operasi. Malware dapat dalam bentuk virus, backdoor, worm, trojan dan ransomware semua jenis malware ini merupakan ancaman utama bagi sistem jaringan komputer (Harjono, 2013).

Pada periode januari hingga agustus tahun 2021 terdapat 888.711.736 serangan siber yang tercatat oleh sistem BSSN, dan diungkapkan oleh kepala BSSN bahwa "Salah satu tren serangan siber di Indonesia kerap kali berupa serangan ransomware yang meminta tebusan hingga berujung kebocoran data". Ransomware sendiri merupakan jenis *malicious* yang menuntut tebusan finansial dari seorang korban dengan melakukan enkripsi pada data yang bersifat pribadi. Kegiatan penyebaran ransomware dilakukan oleh penyerang atau *threat actor* dengan tujuan utama adalah finansial, oleh karenanya *threat actor* menjadikan data pribadi sebagai sandera utama.

Serangan ransomware yang masuk ke Indonesia dapat dilihat dari Gambar 1 yang tercata pada sistem BSSN sepanjang tahun 2020.



Gambar 1. Jumlah Serangan Malware Ransomware

Selain jumlah infeksi yang meningkat ancaman lain ransomware berupa varian malware ini yang juga terus berkembang dengan ancaman yang lebih berbahaya. Hal ini diungkapkan oleh salah satu perusahaan keamanan Tren Micro, dimana pada November 2019 hingga Maret 2021 terdapat 35 varian baru jenis malware ransomware yang ditemukan dengan ancaman yang lebih berbahaya yakni menggunakan teknik *double extortion* atau data yang dienkripsi akan disebarkan ke publik jika tebusan tidak dipenuhi (Agcaoili et al., n.d.)

AgeLocker	CryLock	Hades	NetWalker	REvil/Sodinokibi
Ako/MedusaLocker	DarkSide	LockBit	Pay2Key	Ryuk
AlumniLocker	DoppelPaymer	Maze	ProLock	Sekhmet
Avaddon	Egregor	Mespinoza/Pysa	RagnarLocker	Snatch
Babuk Locker	Ekans	MountLocker/AstroLocker	Ragnarok	SunCrypt
Clop	Everest	Nefilim	RansomExx	Thanos
Conti	Exx/Defray777	Nemty	RanzyLocker/ ThunderX	Xinof

Gambar 2. Varian Baru Ransomware

Berdasarkan dampak yang disebabkan dari infeksi ransomware, maka penanganan terhadap serangan ransomware sedini mungkin sudah menjadi masalah yang banyak menarik minat para peneliti maupun para praktisi bidang cyber security. Di Indonesia sendiri terkhusus BSSN telah melakukan mekanisme *Reverse Engineering* pada malware ransomware untuk mengetahui cara kerja secara detail dari ransomware, akan tetapi kendala terbesar dari mekanisme ini adalah varian ransomware yang sangat beragam, hal ini dipengaruhi karena *payload* ransomware yang diperdangankan pada situs D*ark Web* sehingga pertumbuhan varian ransomware menjadi sangat beragam dengan pola yang terus berubah-ubah tergantung dari *threat actor* ransomware itu sendiri, belum lagi dengan pengunaan *code obfuscation*, *zero day attack* yang akhirnya menjadi kendala lain dalam mendeteksi malware ransomware.

Beberapa penelitian yang dilakukan untuk mendeteksi malware ransomware dengan menggunakan beberapa metode antara lain (Hwang et al., 2020) membagun model deteksi ransomware dengan memanfaatkan perubahan yang terjadi pada komputer seperti *API* calls, registry, string, dan lainnya, penelitian ini menggunakan markov chain kemudian memanfaatkan algoritma random forest untuk proses klasifikasi. Hasil dari penelitian ini menunjukkan akurasi 97,3%, FPR 4,8%, dan FNR 1,5%, Penelitian (Maniath et al., 2017) memanfaatkan LSTM dalam melakukan deteksi ransomware dengan menggunakan analisis dinamis dari panggilan *API* ransomware, penelitian ini menghasilkan akurasi 96,67% dengan 157 *sampel* ransomware yang dilatih. Penelitian (Ashraf et al., 2019) dilakukan dengan menggabungkan analisis statis, dinamis dan deep learning (ResNet-18), penelitian ini menghasilkan akurasi yang berada pada kisaran 98%, akan tetapi akurasi turun karena adanya fitur yang berlebihan. Untuk menghindari perhitungan yang tidak perlu, maka penulis mengambil 300 fitur teratas. 300 fitur ini membawa penelitian pada akurasi 92%. (Sheen & Yadav, 2018) dengan metode yang sama yakni mengekstrak panggilan *API* yang digunakan ransomware untuk membedakan ransomware dan non ransomware, kemudian melakukan klasifikasi menggunakan random forest menghasilkan FNR 0,9653 dan FPR 0,075. Penelitian (Bajpai & Enbody, 2020) memberikan tinjauan rinci tentang panggilan *API* umum ransomware. penulis mengusulkan empat kelas panggilan *API* yang dapat digunakan untuk membuat profil dan menghasilkan fakta hubungan panggilan *API* yang efektif yang digunakan dalam pendeteksian. Hasil ekstraksi panggilan *API* ransomware pada penelitian ini menunjukkan bahwa varian ransomware lanjutan ataupun ransomware dari keluarga yang berbeda memiliki kesamaan dalam implementasi.

Berdasarkan beberapa penjelasan penelitian sebelumnya maka penelitian ini membuat sebuah hipotesis awal yakni dengan adanya kesamaan panggilan *API* yang dimiliki oleh ransomware, maka model deteksi yang dibuat tetap mampu mendeteksi ransomware meskipun

terdapat varian yang belum pernah ditemui atau dipelajari oleh model sebelumnya. Dari hipotesis inilah maka penelitian ini mengusulkan sebuah model deteksi dengan memanfaatkan panggilan *API* yang dihasilkan ransomware dan normalware kemudian memanfaatkan metode ekstraksi fitur N-gram dan TF-IDF untuk membuat subset fitur dari API sampel, serta menggunakan algortima klasifikasi yakni SVM, Random Forest dan LightGBM untuk membuat model mendeteksi ransomware yang akan digunakan dalam mendeteksi ransomware baru guna membuktikan hipotesis awal yang dibuat.

I.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas, maka rumusan masalah pada penelitian ini sebagai berikut :

1. Bagaimana membuat model berdasarkan panggilan *API* untuk mendeteksi ransomware dan normalware ?
2. Bagaimana kinerja model yang dihasilkan dalam mendeteksi varian baru ransomware yang belum pernah ditemui sebelumnya?

I.3 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut :

1. Model yang dibuat mampu mendeteksi ransomware dan normalware
2. Model yang dihasilkan mampu mendeteksi varian baru ransomware meskipun belum pernah ditemui sebelumnya

I.4 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini Manfaat dari penelitian ini adalah dapat menjadi acuan dalam pembuatan model prevention infeksi ransomware pada sistem operasi windows dengan pemanfaatakan pemanggilan API.

I.5 Batasan Masalah

Adapun batasan masalah pada penelitian sebagai berikut :

1. Penelitian ini hanya difokuskan pada serangan malware jenis ransomware

2. Dataset malware yang digunakan berasal dari situs repository malware seperti VirusShare, Any Run, dan Malware Bazaar. Sedangkan untuk dataset Normalware didapatkan dari penelitian (Fang et al., 2020)
3. Penetilian ini hanya melakukan *binary clasification* ransomware.

BAB II

TINJAUAN PUSTAKA

II.1 Kajian Pustaka

Kajian Pustaka yang tertuang pada bab ini adalah representasi dari hasil pendahuluan yang dilakukan oleh penulis, studi literatur yang dilakukan tersebut berupa review terhadap jurnal, prosiding, artikel, dan situs website yang relevan dan mendukung dalam penelitian yang akan dilakukan

II.1.1 Analisis Dinamis

Analisis dinamis merupakan proses eksekusi setiap perangkat lunak malware atau non malware dalam lingkungan yang terisolasi (misalnya, Cuckoo Sandbox) untuk mengumpulkan informasi perilaku runtime malware. Sandbox akan memeriksa proses *sampel* yang berjalan pada komputer secara keseluruhan seperti perubahan registry, komunikasi internet dan peristiwa lainnya (Syaputra, 2020)

II.1.2 *API (Application Programming Interface)*

API adalah singkatan dari *Application Programming Interface*. *API* sendiri merupakan *interface* yang dapat menghubungkan satu aplikasi dengan aplikasi lainnya. Dengan kata lain, peran *API* sebagai jalur berkomunikasi aplikasi dengan sistem operasi. Adapaun proses yang dilakukan oleh malware atau perangkat lunak jinak pada sistem yang berjalan hampir sama yakni seperti membuat, membaca, dan menulis, serta menghapus file atau memodifikasi kunci registry. Pada malware khususnya ransomware selain penggunaan Crypto *API*, panggilan *API* non-kriptografis dibuat oleh ransomware untuk memenuhi batasan lain dalam rantai pembunuhan seperti penyembunyian ancaman atau scanning dari anti virus (Bajpai & Enbody, 2020).

II.1.3 *NLP (Natural Language Processing)*

NLP termasuk dalam sub bidang kecerdasan buatan. NLP merupakan metode otomatis untuk memahami dan menganalisis bahasa

alami dan memperoleh informasi dengan menggunakan algoritma pembelajaran mesin, termasuk pengenalan suara, ekstraksi informasi, kategorisasi teks, dll. Program komputer berkomunikasi dengan sistem operasi melalui panggilan API. Komunikasi ini mirip bagaimana orang menggunakan bahasa untuk berkomunikasi dengan orang lain. Setiap jenis program memiliki pola panggilan API tertentu, sehingga NLP dapat diterapkan pada bidang pendeteksian malware (Qin et al., 2020)

II.1.4 N-Gram

N-gram adalah substring dari sampel teks atau string ucapan yang diberikan dengan panjang n. String ini dapat mencakup beberapa jenis tergantung pada aplikasinya, misalnya mencakup huruf, kata, fonetik, suku kata, dll. N-gram dibuat dengan memisahkan string teks menjadi substring dengan panjang tetap. Misalnya, “MALWARE” kemudian mengolah dengan memberi n = 3 maka akan terlihat seperti "MAL", "ALW", "LWA", "WAR", "ARE", sebagai hasil dari sifat file analisis berbasis string (Ali et al., 2020), proses ini beruntuk untuk menemukan karakteristik dari dokumen yang biasanya digunakan dalam *spelling correction*, *word prediction.*

II.1.5 TF-IDF

TF-IDF merupakan dua gabungan metode yang digunakan dalam proses pengolahan text. Pertama TF (*Term Frequency*) digunakan untuk mengukur seberapa banyak term yang ada dalam sebuah dokumen (Hakim et al., 2014). Sedangakn IDF (*Inverse Document Frequency*) memberikan nilai pada term sesuai dengan kelangkaannya. Jika sebuah term sering muncul, maka IDF memberikan bobot yang lebih sedikit dan jika sebuah term jarang muncul, bobot yang diberikan lebih banyak (Schofield et al., 2021). Formula untuk menghitung TF, IDF dan TF-IDF yakni

a. Rumus untuk *Term Frequency*

$$TF = (\frac{x}{d})$$ (Canzanese et al., 2016)

b. Rumus untuk *Inverse Document Frequency*

$$IDF = \log\left(\frac{P}{d}\right)$$ (Canzanese et al., 2016)

c. Rumus untuk TF-IDF

$$TF - IDF = TF * IDF$$ (Schofield et al., 2021)

Dimana, x adalah jumlah fitur yang diketahui pada sebuah dokumen dari proses n-gram, p adalah jumlah dokumen, dan d adalah jumlah *term* yang muncul dalam setiap dokumen.

II.2 Metode Penyelesaian Masalah

2.1. *State Of The Art* Penelitian

Tabel 1. *State Of The Art*

No	Judul, Nama, Tahun, Penerbit	Objek dan Permasalahan	Metode Penyelesaian dan Dataset	Kinerja	Korelasi < = >
1	**Judul:** Deteksi Ransomware berdasarkan Panggilan *API* Sistem Operasi Windows **Penulis :** Hartinah	**Objek :** Mendeteksi varian lain ransomware yang belum pernah dipelajari oleh model **Permasalahan :** Bagaimana model yang dibuat dapat mendeteksi ransomware varian lain meskipun belum pernah dipelajari oleh model	**Metode :** Analisis Dinamis, N-gram, TF-IDF, SVM, RF, LightGBM **Dataset :** Jumlah sampel ransomware 958 dari 16 varian berbeda dan 514 normalware, dibagi menjadi data latih (*training* dan *testing*) dan data uji (data baru)	Diharapkan model yang dibuat mempu mendeteksi varian lain ransomware meskipun belum pernah dipelajari sebelumnya berdasarkan kesamaan panggilan *API* pada proses pembelajaran.	
2	**Judul:** Ransomware *prediction using supervised learning algorithms* (Adamu & Awan, 2019)	**Objek :** Klasifikasi ransomware **Permasalahan :** Bagaimana kinerja algoritma SVM dalam melakukan	**Metode :** SVM **Dataset :** Jumlah sampel ransomware 582 dari 11 varian berbeda dan 942 normalware, dibagi menjadi data latih	Model SVM menghasilkan akurasi 88,2% dengan nilai RMSE 0,179. SVM memiliki kinerja tinggi dalam mengklasifikas	<

No	Judul, Nama, Tahun, Penerbit	Objek dan Permasalahan	Metode Penyelesaian dan Dataset	Kinerja	Korelasi < = >
	Penulis : Adamu, dkk **Tahun:** 2019 **Penerbit :** IEEE	klasifikasi ransomware Dataset :	(*training* dan *testing***)**	ikan ransomware	
3	**Judul :** MaMaDroid: *Detecting Android* Malware *by Building Markov Chains of Behavioral Models* (Mariconti et al., 2017) **Penulis :** Mariconti, dkk **Tahun:** 2017 **Penerbit :** IEEE	**Objek :** Deteksi malware pada *smart phone* android **Permasalahan :** Bagaimana kinerja model deteksi malware dengan dalam mendeteksi malware	**Metode :** Markov Chain dan Random Forest **Dataset :** Jumlah sampel malware 35.493 dan 8.447 normalware, dibagi menjadi data latih (*training* dan *testing***)**	Model ini juga mempertahankan kinerja deteksi yang baik dalam satu tahun dengan nilai F-measure adalah 87%.	<
4	**Judul :** *API Call Based* Ransomware *Dynamic Detection Approach Using* TextCNN (Qin et al., 2020) **Penulis :** Qin, dkk **Tahun :** 2020 **Penerbit :** ICBAIE 2020	**Objek :** Deteksi malware Ransomware **Permasalahan :** Bagaimana kinerja model TextCNN dalam mendeteksi Ransomware	**Metode :** TextCNN **Dataset :** Jumlah sampel ransomware 1000 dan normalware 1000, dibagi menjadi data latih (*training* 80 % dan *testing* 20%)	Hasil akurasi model sebesar 0,959 dan di klaim lebih baik dari dibanding menggunakan metode traditional dalam klasifikasi ransomware.	<
5	**Judul :** *Automatic* Ransomware *detection and analysis based on dynamic API calls flow*	**Objek :** Deteksi malware Ransomware **Permasalahan :** Bagaimana perbandingan	**Metode :** RF, SVM, Simple Logistic (SL) dan NB **Dataset :** Jumlah sampel rasomware 83	Dari ke 4 jenis algoritma diketahui Simple Logistic unggul dengan akurasi 98%.	=

No	Judul, Nama, Tahun, Penerbit	Objek dan Permasalahan	Metode Penyelesaian dan Dataset	Kinerja	Korelasi < = >
	graph (Chen et al., 2017) **Penulis :** Chen, dkk **Tahun :** 2017 **Penerbit :** RACS 2017	algoritma dalam mendeteksi Ransomware	dan 85 normalware, dibagi menjadi data latih (*training* dan *testing*)		
6	**Judul :** NLP-*based Approaches for* Malware *Classification from API Sequences* (Tran & Sato, 2017) **Penulis :** Tran, dkk **Tahun :** 2017 **Penerbit :** IES 2017	**Objek :** Klasifikasi malware **Permasalahan :** Bagaimana perbandingan kinerja metode NLP dalam proses klasifikasi malware	**Metode :** *TF-IDF, Paragraph Vector with Distributed Bag of Words and Paragraph Vector with Distributed Memory* **Dataset :** Jumlah sampel malware acak 23,080, dibagi menjadi data latih (*training* dan *testing*)	Dari hasil penelitian diketahui akurasi TF-IDF memberikan hasil paling tinggi dengan akurasi sebesar 99% menggunakan algoritma SVM	=
7	**Judul :** *Two-Stage* Ransomware *Detection Using Dynamic Analysis and Machine Learning Techniques* (Hwang et al., 2020) **Penulis :** Hwang, dkk **Tahun :** 2020 **Penerbit :** Springer	**Objek :** Klasifikasi Ransomware **Permasalahan :** Bagaimana kinerja model deteksi ransomware dengan teknik markov chain dan *machine learning*	**Metode :** Dinamic analisis, markov chain, Random Forest **Dataset :** Jumlah sampel ransomware 1176 dan 1160 normalware**,** dibagi menjadi data latih (*training* dan *testing*)	Hasil dari percobaan ini memberikan akurasi 97,3% dengan FPR 4,8%, dan FNR 1.5%.	=
8	**Judul :** Ransomware *Analysis using Fitur*	**Objek :** Klasifikasi Ransomware	**Metode :** DNN **Dataset :**	Degan menerapkan teknik PCA pada fitur	<

No	Judul, Nama, Tahun, Penerbit	Objek dan Permasalahan	Metode Penyelesaian dan Dataset	Kinerja	Korelasi < = >
	Engineering and Deep Neural Networks (Ashraf et al., 2019) **Penulis :** Ashraf, dkk **Tahun :** 2019 **Penerbit :** Cornet University	**Permasalahan :** Bagaimana kinerja model DNN dalam mendeteksi ransomware	metode statis memiliki 1700 sampel ransomware dan 1946 normalware, untuk metode dinamis terdiri 1455 ransomware dan 1989 normalware, yang dibagi menjadi data latih (*training* dan *testing*)	ekstraksi, model menghasilkan klasifikasi varian ransomware pada akurasi 92%	
9	**Judul :** *Convolutional Neural Network for* Malware *Classification Based on API Call Sequence*(Schofield et al., 2021) **Penulis :** Schofield, dkk **Tahun :** 2021 **Penerbit :** Computer Science & Information Technology (CS & IT)	**Objek :** Klasifikasi Malware **Permasalahan :** Bagaimana kinerja model dalam mendeteksi beberapa malware.	**Metode :** CNN, TF-IDF dan Categorical Vector **Dataset :** Jumlah sampel malware 7.107 dari 8 varian berbeda yang dibagi menjadi data latih (*training* dan *testing*)	Dengan metode TF-IDF dan CNN akurasi mencapai 98,17% sedangkan dengan Categorical Vector dan CNN mendapatkan hasil akurasi 95,40%	=
10	**Judul :** *Deep Learning* LSTM *Based* Ransomware *Detection* (Maniath et al., 2017) **Penulis :** Maniath, dkk	**Objek :** Klasifikasi Ransomware **Permasalahan :** Bagaimana kinerja model LSTM dalam mendeteksi ransomware	**Metode :** LSTM **Dataset :** Jumlah sampel ransomware 157 yang dibagi menjadi data latih (*training* dan *testing*)	Penelitian ini menghasilkan akurasi 96,67%	=

No	Judul, Nama, Tahun, Penerbit	Objek dan Permasalahan	Metode Penyelesaian dan Dataset	Kinerja	Korelasi < = >
	Tahun : 2017 **Penerbit :** RDCAPE				
11	**Judul :** Ransomware *detection by mining API call usage*(Sheen & Yadav, 2018) **Penulis :** Sheen, dkk **Tahun :** 2018 **Penerbit :** ICACCI 2018	**Objek :** Klasifikasi Ransomware **Permasalahan :** Bagaimana kinerja model dalam mendeteksi ransomware	**Metode :** Random Forest **Dataset :** Jumlah sampel ransomware 3.620 yang dibagi menjadi data latih (*training* dan *testing*)	Penelitian ini menghasilkan akurasi 96% dan tingkat positif palsu yang rendah 0,075.	=
12	**Judul :** *Static Analysis Based Behavioral API for* Malware *Detection using Markov Chain* (Al bakri, Abbas M, Hussein, 2014) **Penulis :** Al bakri, dkk **Tahun :** 2014 **Penerbit :** Computer Engineering and Intelligent Systems	**Objek :** Klasifikasi Malware **Permasalahan :** Bagaimana kinerja model dalam mendeteksi malware	**Metode :** Markov Chain	Hasil pemodelan *API* dengan markov chain menghasilkan akurasi deteksi 92%.	<
13	**Judul :** Ransomware *detection using process mining and classification algorithms* (Bahrani & Bidgly, 2019)	**Objek :** Klasifikasi Ransomware **Permasalahan :** Bagaimana perbandingan kinerja model dalam mendeteksi ransomware	**Metode :** CNN, TF-IDF dan Categorical Vector **Dataset :** Terdapat 21 varian berbeda dari ransomware yang dibagi menjadi data	Hasil pengujian keempat algoritma klasifikasi membawa kesimpulan bahwa algoritma J48 dan random forest memiliki	<

No	Judul, Nama, Tahun, Penerbit	Objek dan Permasalahan	Metode Penyelesaian dan Dataset	Kinerja	Korelasi < = >
	Penulis : Bahrani, dkk **Tahun :** 2019 **Penerbit :** ISCISC 2019		latih (*training* dan *testing*)	akurasi terbaik yakni 95%.	
14	**Judul :** *Automated Analysis Approach for the Detection of High Survivable* Ransomware (Ahmed et al., 2020) **Penulis :** Al Ahmed, dkk **Tahun :** 2020 **Penerbit :** KSII Transactions on Internet and Information Systems	**Objek :** Klasifikasi Ransomware **Permasalahan :** Bagaimana kinerja model dalam mendeteksi ransomware	**Metode :** TF-IDF dan ANN **Dataset :** Jumlah sampel ransomware 1,254 dari 14 varian berbeda dan 1308 normalware yang dibagi menjadi data latih (*training* dan *testing*)	Penelitian ini menghasilkan akurasi deteksi 98,7% dengan beberapa tingkat positif palsu di bawah 3%.	=
15	**Judul :** *Classifying* Ransomware *Using Machine Learning Algorithms* (Egunjobi et al., 2019) **Penulis :** Egunjobi, dkk **Tahun :** 2019 **Penerbit :** Lecture Notes in Computer Science	**Objek :** Klasifikasi Ransomware **Permasalahan :** Bagaimana kinerja model dalam mendeteksi ransomware	**Metode :** SVM, Random Forest, **Dataset :** Jumlah sampel malware 200 dan normalware 200 dibagi menjadi data latih (*training* dan *testing*)	Hasil dari penelitian menunjukkan Random Forest memberikan kemampuan deteksi yang relatif tinggi dengan akurasi 99,5%.	<
16	**Judul :** *Overview and Case Study for*	**Objek :** Klasifikasi Ransomware	**Metode :** Deep Neural Network	Penggunaan algoritma DNN menunjukkan	=

No	Judul, Nama, Tahun, Penerbit	Objek dan Permasalahan	Metode Penyelesaian dan Dataset	Kinerja	Korelasi < = >
	Ransomware *Classification Using Deep Neural Network* (Nurnoby & El-Alfy, 2019) **Penulis :** Nurnoby, dkk **Tahun :** 2019 **Penerbit :** IEEE	**Permasalahan :** Bagaimana kinerja model DNN dalam mendeteksi ransomware	**Dataset :** Jumlah sampel ransomware 582 dari 11 varian berbeda dan 360 normalware dibagi menjadi data latih (*training* dan *testing*)	kinerja yang lebih baik dengan akurasi DNN 97% sedangkan RF 90%	
17	**Judul :** *Evaluation to classify* Ransomware *varians based on correlations between API* (Zhou et al., 2020) **Penulis :** Zhou, dkk **Tahun :** 2020 **Penerbit :** ICISSP 2020	**Objek :** Klasifikasi Ransomware **Permasalahan :** Bagaimana kinerja model dalam mendeteksi ransomware	**Metode :** correlation coefficient dan SVM **Dataset :** Jumlah sampel ransomware 899 dari 9 varian berbeda dan 241 normalware yang dibagi menjadi data latih (*training* dan *testing*)	Penelitian ini menghasilkan akurasi yakni 98,2%.	=
18	**Judul :** *An Empirical Study of API Calls in* Ransomware (Bajpai & Enbody, 2020) **Penulis :** Bajpai, dkk **Tahun :** 2020 **Penerbit :** IEEE	**Objek :** Ransomware **Permasalahan :** Bagaimana pola setiap ransomware yang berasal dari varian yang berbeda	**Metode :** Analisis statis dan dinamis **Dataset :** 5 sampel varian ransowmare berbeda	Hasil visualisasi penelitian ini menunjukkan bahwa bahkan ransomware dari varian yang berbeda pada dasarnya tetap memiliki kesamaan panggilan *API*	
19	**Judul :** LightGBM-*based* Ransomware *Detection using*	**Objek :** Klasifikasi Ransomware **Permasalahan :**	**Metode :** LightGBM **Dataset:**	Hasil eksperimen menunjukkan akurasi keseluruhan	=

No	Judul, Nama, Tahun, Penerbit	Objek dan Permasalahan	Metode Penyelesaian dan Dataset	Kinerja	Korelasi < = >
	API Call Sequences (Nguyen & Lee, 2021) **Penulis :** Nguyen, dkk **Tahun :** 2021 **Penerbit :** IJACSA	Bagaimana kinerja model dalam mendeteksi berbagai varian ransomware	Jumlah sampel ransomware 1.803 dari 8 varian berbeda dan normalware 4.008 yang dibagi menjadi data latih (*training* dan *testing*)	98,7% saat melakukan klasifikasi multikelas. Secara khusus, tingkat deteksi ransomware dan normalware sebesar 99,9%.	
20	**Judul :** Malware *classification using API system calls* (Ninyesiga & Ngubiri, 2018) **Penulis :** Ninyesiga, dkk **Tahun :** 2018 **Penerbit :** IJOTM	**Objek :** Klasifikasi Malware **Permasalahan :** Bagaimana kinerja model dalam mendeteksi ransomware	**Metode :** TF-IDF, Naive Bayes, SVM, RF **Dataset :** Jumlah sampel malware dan benign sebanyak 552 yang dibagi menjadi data latih (*training* dan *testing*)	Hasil menunjukkan bahwa algoritma Random Forest memiliki akurasi terbaik yakni 96,4%.	<

II.3 Metode yang Digunakan

Berdasarkan tabel *State of The Art* diatas atas, dapat disimpulkan bahwa penelitian untuk mendeteksi malware ranosmware dapat dilakukan dengan memanfaatkan panggilan *API* . Kemudian proses ekstraksi fitur panggilan *API* dapat menggunakan teknik N-gram, dan TF-IDF. N-gram untuk membentuk subset fitur, kemudian TF-IDF berfungsi menghitung kemunculan suatu *term* dari suatu dokumen panggilan *API* yang dihasilkan oleh ransomware, bobot yang dihasilkan inilah yang akan dimasukkan ke dalam model pembelajaran untuk mendeteksi ransomware dan normalware, sebelum tahap pembelajaran dataset yang digunakan akan melalui proses *split test* atau pencarian jumlah data untuk *training* dan *testing*, kemudian untuk melakukan proses pembelajaran model akan diterapkan tiga jenis algoritma yakni SVM, Random Forest dan LightGBM, hasil deteksi ini

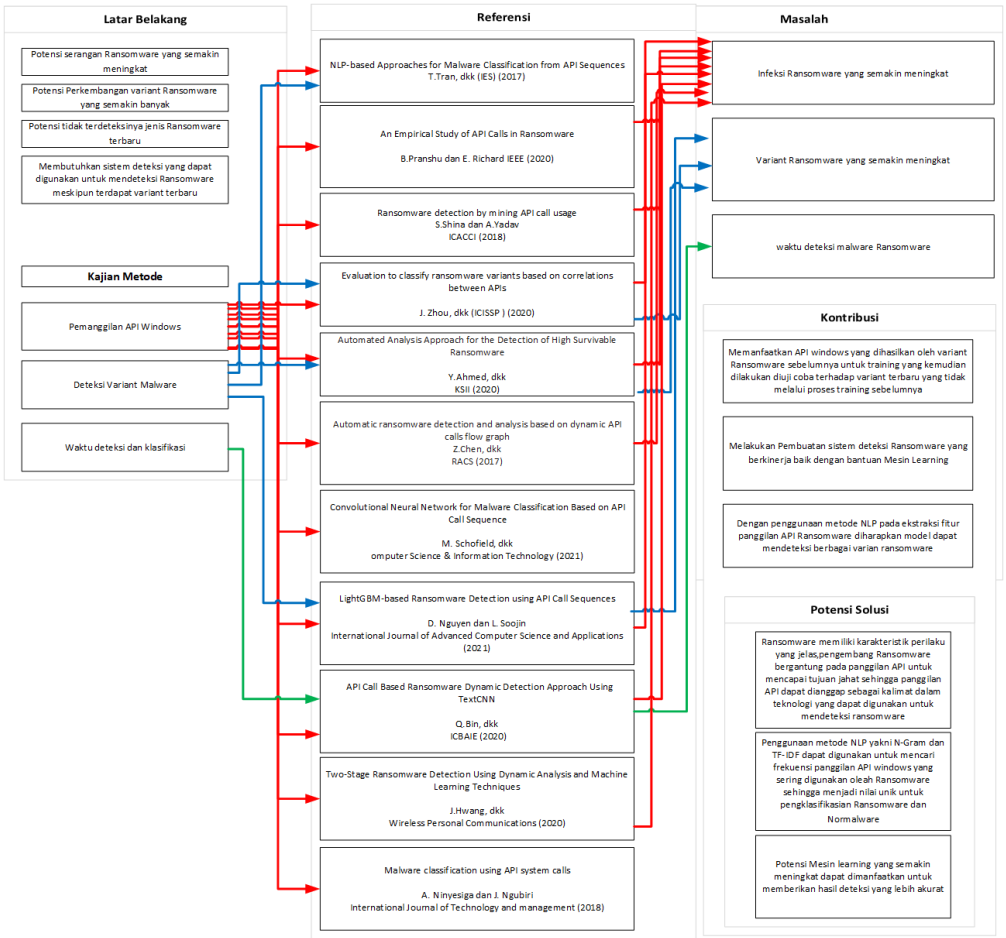
berupa kineja model dalam mendeteksi ransomware baik yang digunakan pada proses pembelajaran maupun varian ransomware lain yang tidak belum pernah melalui proses pembalajaran model.

II.4 Target Hasil Penelitian

Berdasarkan tabel dari *State of The Art*, target akurasi pada poses pembelajaran model yakni berkisar 90% sedangkan untuk target akurasi deteksi data ransomware lain yang belum pernah melalui proses pembelajaran berkisar 70%. Fokus penelitian ini untuk mendeteksi varian ransomware lain dimana varian ini belum pernah melalui proses pembelajaran model sebelumnya.

II.5 Kerangka Pikir

Kerangka pikir dapat dilihat pada Gambar 3, yang menjelaskan mengenai alur penelitian yang akan dilakukan.



Gambar 3. Kerangka Pikir