

SKRIPSI

**SISTEM PENDETEKSI *PHISING* MENGGUNAKAN
ALGORITMA *DECISION TREE* BERBASIS EKSTENSI WEB
BROWSER**

Disusun dan diajukan oleh:

**MUH. NAUFAL FEBRIANTAMA
D121181021**



**PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024**

LEMBAR PENGESAHAN SKRIPSI**SISTEM PENDETEKSI PHISING MENGGUNAKAN
ALGORITMA DECISION TREE BERBASIS EKSTENSI WEB
BROWSER**

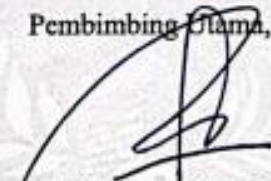
Disusun dan diajukan oleh

Muh. Naufal Febriantama
D121 18 1021


Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka
Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika
Fakultas Teknik Universitas Hasanuddin
Pada tanggal 28 November 2024
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,


Dr. Eng. Ady Wahyudi Paundu, S.T., M.T.
NIP 19750313 200912 1 003

Ketua Program Studi,


Prof. Dr. H. Idrabayu, S.T., M.T., M.Bus.Sys., IPM., ASEAN.Eng.
NIP 19750716 200212 1 004

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;
Nama : Muh. Naufal Febriantama
NIM : D121181021
Program Studi : Teknik Informatika
Jenjang : S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

{Sistem Pendeteksi Phising Menggunakan Algoritma Decision Tree Berbasis Ekstensi Web Browser}

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 28 November 2024

Yang Menyatakan



Muh. Naufal Febriantama

ABSTRAK

MUH. NAUFAL FEBRIANTAMA. *Sistem Pendeteksi Phishing Berbasis Algoritma Decision Tree pada Ekstensi Browser Google Chrome* (dibimbing oleh Ady Wahyudi Paundu).

Penggunaan internet yang semakin meluas membawa dampak positif dalam berbagai aspek kehidupan, namun di sisi lain, ancaman keamanan siber juga meningkat, termasuk serangan *phishing*. *Phishing* merupakan salah satu jenis serangan yang bertujuan untuk mencuri informasi sensitif seperti kata sandi atau data keuangan melalui situs web palsu yang meniru situs asli. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem pendeteksi *phishing* berbasis algoritma *Decision Tree*, yang diintegrasikan ke dalam ekstensi *Browser*. Sistem ini diharapkan dapat membantu pengguna mengidentifikasi situs web berbahaya secara otomatis ketika mengakses internet.

Dalam penelitian ini, algoritma *Decision Tree* digunakan untuk mengklasifikasikan URL berdasarkan fitur-fitur tertentu, seperti panjang URL, jumlah subdomain, dan penggunaan HTTPS. Dataset yang digunakan terdiri dari situs web *phishing* dan *non-phishing*, yang kemudian dilatih menggunakan algoritma *Decision Tree*. Sistem ini diuji dengan metode *black-box testing* pada ekstensi *Browser* Google Chrome untuk mengevaluasi kinerjanya.

Hasil pengujian menunjukkan bahwa sistem ini mampu mendeteksi situs web *phishing* dengan akurasi tinggi, dengan tingkat kesalahan deteksi yang rendah terhadap situs web asli (*false positives*) dan tidak ada kesalahan dalam mendeteksi *phishing* (*false negatives*). Dengan demikian, sistem yang dikembangkan dapat memberikan perlindungan yang lebih baik bagi pengguna saat berselancar di internet.

Kata kunci: *Phishing*, *Decision Tree*, Keamanan Siber, Ekstensi Browser, *Machine Learning*

ABSTRACT

MUH. NAUFAL FEBRIANTAMA. Decision Tree Algorithm-Based *Phishing* Detection System on Google Chrome Browser Extension (guided by Ady Wahyudi Paundu).

The widespread use of the internet has a positive impact on various aspects of life, but on the other hand, cybersecurity threats are also increasing, including *phishing* attacks. *Phishing* is a type of attack that aims to steal sensitive information such as passwords or financial data through fake websites that mimic the real site. This research aims to design and implement a *phishing* detection system based on the *Decision Tree* algorithm, which is integrated into the Browser extension. This system is expected to help users identify malicious websites automatically when accessing the internet.

In this study, the *Decision Tree* algorithm was used to classify URLs based on certain features, such as URL length, number of subdomains, and HTTPS usage. The dataset used consists of *phishing* and non-*phishing* websites, which are then trained using the *Decision Tree* algorithm. The system was tested using the black-box testing method on the Google Chrome Browser extension to evaluate its performance.

The test results show that this system is able to detect *phishing* websites with high accuracy, with a low detection error rate against the original website (false positives) and no errors in detecting *phishing* (false negatives). Thus, the developed system can provide better protection for users when surfing the internet.

Keywords: *Phishing*, *Decision Tree*, Cybersecurity, Browser Extensions, Machine Learning

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	i
PERNYATAAN KEASLIAN.....	ii
ABSTRAK	iii
ABSTRACT	iv
DAFTAR ISI.....	v
DAFTAR GAMBAR	vii
DAFTAR TABEL.....	viii
DAFTAR SINGKATAN	ix
DAFTAR LAMPIRAN.....	x
KATA PENGANTAR	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	4
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Kerangka Pikir.....	6
2.2 Landasan Teori	8
2.2.1 <i>Phising</i>	8
2.2.2 Algoritma <i>Machine Learning</i> dalam Deteksi <i>Phishing</i>	9
2.3 Algoritma <i>Decision Tree</i>	11
2.3.1 Prinsip Kerja <i>Decision Tree</i>	12
2.3.2 Kelebihan dan Kekurangan <i>Decision Tree</i>	13
2.3.3 <i>Decision Tree</i> dalam Deteksi <i>Phishing</i>	14
2.4 Website.....	15
2.5 Ekstensi Browser untuk Deteksi <i>Phishing</i>	15
2.6 Python.....	17
2.7 Java Script	17
2.8 Node JS.....	18

2.9 Visual Studio Code.....	19
2.10 Penelitian Terkait	19
BAB III METODOLOGI PENELITIAN.....	23
3.1 Desain Penelitian	23
3.2 Tahapan Penelitian	23
3.2.1 Observasi	24
3.2.2 Studi Literatur	24
3.2.3 Analisis	24
3.2.4 Desain Sistem	25
3.2.5 Implementasi.....	30
3.2.6 Pengujian Sistem.....	32
3.3 Alat dan Bahan Penelitian	33
3.3.1 Perangkat Keras	33
3.3.2 Perangkat Lunak	33
3.4 Metode Evaluasi	33
3.4.1 Perhitungan Rata-Rata Deteksi.....	35
3.4.2 Metrik Evaluasi.....	35
BAB IV HASIL DAN PEMBAHASAN	38
4.1 Hasil Penelitian.....	38
4.1.1 <i>System</i>	38
4.1.2 Pengujian	41
4.2 Pembahasan	57
4.2.1 Evaluasi Pengujian Website Asli dan <i>Phising</i>	57
4.2.2 Faktor yang Mempengaruhi Ketidakkonsistenan Hasil Deteksi.....	58
BAB V PENUTUP.....	62
5.1 Kesimpulan.....	62
5.2 Saran	62
DAFTAR PUSTAKA	64

DAFTAR GAMBAR

Gambar 2. 1 Kerangka Pikir.....	7
Gambar 2. 2 Cara Kerja Machine Learning.....	11
Gambar 3. 1 Tahapan Penelitian	24
Gambar 3. 2 Flowchart.....	25
Gambar 3. 3 Arsitektur Aplikasi	30
Gambar 3. 4 Pengujian Black Box Testing	30
Gambar 3. 5 Dataset Website Asli.....	31
Gambar 3. 6 Dataset Website Phising.....	32
Gambar 4. 1 Pemberitahuan safe pada Ekstensi Pendeteksi Phising	38
Gambar 4. 2 Pop up safe Ekstensi Aplikasi Pendeteksi Phising.....	39
Gambar 4. 3 Pemberitahuan susp pada Ekstensi Aplikasi Pendeteksi Phising.....	39
Gambar 4. 4 Pop up Suspicious Ekstensi Aplikasi Pendeteksi Phising.....	39
Gambar 4. 5 Pemberitahuan Phis pada Ekstensi Pendeteksi Phising	40
Gambar 4. 6 Pop up Phising Ekstensi Aplikasi Pendeteksi Phising	40
Gambar 4. 7 Grafik Perhitungan Web Asli Menggunakan Ip Address.....	47
Gambar 4. 8 Grafik Perhitungan Web Asli Tanpa Menggunakan Ip Address.....	47
Gambar 4. 9 Grafik Perhitungan Website Phising Menggunakan Ip Address.....	51
Gambar 4. 10 Grafik Perhitungan Website Phising Tanpa Menggunakan Ip Address.....	51
Gambar 4. 11 Source Code Matrik Evaluasi.....	53
Gambar 4. 12 Hasil Matrik Evaluasi Menggunakan Ip Address	53
Gambar 4. 13 Hasil Matrik Evaluasi Tanpa Menggunakan Ip Address	53
Gambar 4. 14 Source Code Confusion Matriks	55
Gambar 4. 15 Matriks kebingungan Menggunakan Ip Address	55
Gambar 4. 16 Matriks kebingungan Tanpa Menggunakan Ip Address	56

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	19
Tabel 4. 1 Pengujian Funcionality Siuitability	41
Tabel 4. 2 Hasil Uji Coba Website Asli Menggunakan Ip Address.....	44
Tabel 4. 3 Hasil Uji Coba Website Asli Tanpa Menggunakan Ip Address.....	45
Tabel 4. 4 Hasil Perhitungan Website Asli Menggunakan Ip Address.....	46
Tabel 4. 5 Hasil Perhitungan Website Asli Tidak Menggunakan Ip Address	46
Tabel 4. 6 Hasil Uji Coba Website Phising Menggunakan Ip Address	48
Tabel 4. 7 Hasil Uji Coba Website Phising Tanpa Menggunakan Ip Address	49
Tabel 4. 8 Hasil Perhitungan Website Phising Menggunakan Ip Address	50
Tabel 4. 9 Hasil Perhitungan Website Phising Tidak Menggunakan Ip Address .	50
Tabel 4. 10 Hasil perhitungan Rata-rata Menggunakan Ip Address	52
Tabel 4. 11 Hasil perhitungan Rata-rata Tanpa Menggunakan Ip Address	52
Tabel 4. 12 Matrik Evaluasi Menggunakan Ip Address.....	54
Tabel 4. 13 Matrik Evaluasi Tanpa Menggunakan Ip Address.....	54

DAFTAR SINGKATAN

Lambang/Singkatan	Arti dan Keterangan
ML	<i>Machine Learning</i>
SVM	<i>Support Vector Machine</i>
URL	<i>Uniform Resource Locator</i>
HTTPS	<i>Hypertext Transfer Protocol Secure</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
F	<i>False</i>
S	<i>Suspicious</i>
T	<i>True</i>
OOP	<i>Object Oriented Programming</i>
JS	<i>Java Script</i>
VCS	<i>Visual Studio Code</i>
IP	<i>Internet Protocol</i>
CSS	<i>Cascading Style Sheets</i>
HTML	<i>Hypertext Markup Language</i>
AI	<i>Artificial Intelligence</i>
ID3	<i>Iterative Dichotomiser 3</i>
ANN	<i>Artificial Neural Network</i>
XGBoost	<i>Extreme Gradient Boosting</i>

DAFTAR LAMPIRAN

Lampiran 1 Sistem Pendeteksi Phising	67
Lampiran 2 Kode Pengujian Sistem dengan Metode Decision Tree	68
Lampiran 3 Surat Penugasan.....	72
Lampiran 4 Berita Acara Seminar Hasil	73
Lampiran 5 Surat Izin Ujian Skripsi	74
Lampiran 6 Berita Acara Ujian Sidang.....	75
Lampiran 7 Log Book	76
Lampiran 8 Lembar Perbaikan Skripsi	77

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **“SISTEM PENDETEKSI PHISING MENGGUNAKAN ALGORITMA DECISION TREE BERBASIS EKSTENSI WEB BROWSER”** ini dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin.

Dalam proses penyusunan skripsi ini, penulis mendapatkan banyak bantuan, dukungan, serta bimbingan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis mengucapkan terima kasih kepada:

1. Tuhan Yang Maha Esa atas semua berkat, karunia serta pertolongan-Nya yang tiada batas, yang telah diberikan kepada penulis disetiap langkah dalam pengerjaan tugas akhir ini.
2. Kedua orang tua penulis serta saudara penulis, serta keluarga yang senantiasa memberikan kekuatan, motivasi, bimbingan moral, materi, kepercayaan dan kasih sayang yang tidak terbatas kepada penulis.
3. Bapak Dr.Eng. Ady Wahyudi Paundu, S.T., M.T., selaku Dosen Pembimbing I, yang telah memberikan banyak masukan, arahan, dan motivasi selama proses penyusunan skripsi ini.

4. Bapak Prof. Dr. Ir. Indrabayu.,ST, MT, M.Bus.Sys., IPM, ASEAN, selaku Ketua Program Studi Teknik Informatika yang telah memberikan dukungan akademik selama masa perkuliahan.
5. Bapak Prof. Dr. Amil Ahmad Ilham, S.T., M.IT., selaku dosen pembimbing akademik yang telah memberikan bimbingan selama masa perkuliahan penulis.
6. Bapak dan ibu dosen Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin atas bimbingan, arahan dan didikannya selama masa perkuliahan.
7. Bapak Robert, Bapak Zainuddin, Ibu Yuanita dan Ibu Arizha Tenri serta segenap staff Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah membantu kelancaran penyelesaian tugas akhir penulis.
8. Teman-teman Synchoronous 2018 dan Teknik 2018 yang telah memberikan nasihat, bantuan dan semangat selama proses penyelesaian tugas akhir ini.
9. Semua Orang yang telah membantu dan memberikan dukungan kepada penulis namun tidak sempat disebutkan dalam menyusun tugas akhir ini, penulis tidak terlepas dari kesalahan sebagai manusia, oleh karena itu penulis mengharapkan kritik dan saran yang membangun dari pembaca. Semoga tugas akhir ini bermanfaat bagi pembaca.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis dengan senang hati menerima kritik dan saran yang membangun untuk perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat, baik bagi penulis maupun pembaca yang memiliki ketertarikan pada topik yang dibahas.

Akhir kata, penulis mengucapkan terima kasih atas segala perhatian dan bantuan yang telah diberikan. Semoga Allah SWT selalu memberikan keberkahan kepada kita semua.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Gowa, 28 November 2024



Penulis,

Muh. Naufal Febriantama

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era teknologi saat ini, penggunaan internet menjadi hal yang umum dan bahkan sangat penting dalam kehidupan sehari-hari di kalangan masyarakat luas. Seiring meningkatnya kebutuhan akan informasi, internet memberikan kontribusi besar dalam memudahkan berbagai aspek kehidupan, termasuk komunikasi, pendidikan, pekerjaan, dan bisnis. Hampir semua kegiatan dan transaksi kini dapat dilakukan secara online tanpa terbatas oleh lokasi berkat dukungan internet.

Meskipun internet memberikan berbagai manfaat, namun perlu diwaspadai bahwa ada oknum yang dapat memanfaatkannya untuk melakukan kejahatan. Salah satu bentuk kejahatan tersebut merupakan pembuatan situs *web Phising* yang digunakan sebagai alat untuk melakukan penipuan. Para pelaku kejahatan ini mengincar pengguna internet yang kurang berpengalaman atau kurang waspada terhadap keamanan transaksi *online*. Data pribadi seperti akun pengguna, kata sandi, dan informasi finansial seperti detail akun bank dan data kartu kredit menjadi target utama para penjahat ini.

Menurut (Heriani, Fitri Novia 2021), *phishing* berasal dari kata dalam bahasa Inggris "fishing" (memancing), yang dianalogikan dengan "memancing" korban untuk memberikan informasi sensitif secara sukarela melalui rekayasa sosial (social engineering). *Phishing* merupakan upaya untuk memperoleh data sensitif seseorang, seperti kata sandi, informasi kartu kredit, atau data pribadi lainnya,

dengan cara penipuan yang menyamar sebagai pihak terpercaya. Fokus utama serangan *phishing* adalah manipulasi psikologis korban, di mana penyerang berusaha membuat korban secara sukarela menyerahkan informasi pribadi mereka melalui media seperti situs web palsu, email penipuan, atau pesan teks. *Phishing* berbeda dengan *spoofing*, yang merupakan manipulasi teknis di mana penyerang memalsukan identitas sumber, seperti alamat email, nomor telepon, domain situs web, atau alamat IP, agar tampak berasal dari pihak yang sah. *Spoofing* sering digunakan untuk mendukung serangan *phishing*, misalnya dengan memalsukan email agar terlihat seperti berasal dari bank.

Langkah-langkah yang dapat dilakukan dalam pencegahan terhadap serangan *Phising* yaitu dengan mengedukasi pengguna untuk dapat mengenali indikasi-indikasi serangan *Phising*, seperti melakukan pengecekan *URL* dengan teliti, menghindari mengklik atau menekan tautan yang mencurigakan, dan selalu memverifikasi keaslian komunikasi online sebelum memberikan informasi pribadi. Di samping itu, penggunaan perangkat lunak keamanan siber, seperti filter email dan perangkat lunak *anti-phishing*, juga dapat berperan dalam melindungi individu dan organisasi dari potensi serangan ini. Dari latar belakang diatas, penulis berinisiatif untuk merancang sistem pendeteksi *Phising* menggunakan algoritma klasifikasi *Decision Tree* berbasis *Web Browser*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah pada penelitian ini ialah, bagaimana implementasi dan penggunaan algoritma

klasifikasi *Decision Tree* dalam mendeteksi ancaman *Phising* melalui *Website* dengan menggunakan ekstensi *Web Browser*.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan dari penelitian ini merupakan :

1. Menjelaskan prinsip kerja dari algoritma *Decision Tree*
2. Mengimplementasikan algoritma *Decision Tree* pada sistem pendeteksi *Phising* berbasis ekstensi *Web Browser* untuk meningkatkan keamanan dan kenyamanan pengguna dalam mengakses sebuah *Website*.
3. Mengetahui pengaruh penggunaan ekstensi pada performa sistem pendeteksi *Phising* pada *Web Browser*.

1.4 Manfaat Penelitian

Berikut uraian manfaat penelitian

1. Hasil penelitian ini dapat membantu meningkatkan keamanan dan kenyamanan pengguna dalam menggunakan internet. Dengan adanya sistem ini, pengguna akan terbantu membedakan antara *Website* yang aman dan yang tidak aman untuk diakses.
2. Penelitian ini dapat memberikan kontribusi untuk pengembangan teknologi keamanan pada sistem pendeteksian *Phising* khususnya pada *Web Browser*.
3. Penelitian ini dapat memperluas cakupan penelitian terkait dengan keamanan data dan topik mengenai ekstensi *Web Browser*.
4. Hasil penelitian ini dapat menjadi dasar untuk penelitian selanjutnya terkait dengan pendeteksian *Phising* khususnya dalam penggunaan algoritma

klasifikasi dan juga pembuatan ekstensi pada *Web Browser*. Penelitian selanjutnya diharapkan mampu memperluas cakupan penelitian atau mengembangkan penelitian yang telah dilakukan sebelumnya.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini merupakan :

1. Penelitian ini hanya focus pada implementasi algoritma *Decision Tree*.
2. Penelitian ini akan memfokuskan pada implementasi algoritma *Decision Tree* sebagai metode klasifikasi utama. Penelitian ini tidak akan membahas atau membandingkan dengan algoritma klasifikasi lainnya.
3. Penelitian ini akan fokus pada perancangan *system* pendeteksi *Phising*.
4. Penelitian ini membatasi diri pada pendeteksian *Phising* berbasis *Website* lebih khusus kearah mendeteksi atribut *URL* dari sebuah *Website*. Serangan *Phising* melalui saluran komunikasi lain seperti *email* atau pesan teks tidak akan menjadi fokus utama dalam penelitian ini.
5. Pengujian performa sistem akan dilakukan dengan menggunakan lingkungan simulasi atau pengujian pada *Web Browser Chrome*. Pengujian pada aplikasi *Web Browser* lainnya tidak akan dilakukan.
6. Penelitian ini membutuhkan dataset yang mencakup contoh-contoh *Website Phising* dan non-*Phising* yang representative.
7. Penelitian ini akan melibatkan periode waktu tertentu untuk pengembangan dan pengujian implementasi algoritma *Decision Tree* pada sistem pendeteksi *Phising* berbasis *Web Browser*. Batasan waktu ini perlu dipertimbangkan agar penelitian dapat dilakukan dalam batas waktu yang realistis.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini dibagi menjadi lima bab sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini membahas landasan teori yang mendukung penelitian ini, termasuk konsep *phishing*, *algoritma Decision Tree*, dan penelitian-penelitian terdahulu terkait sistem pendeteksi *phishing*.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metodologi yang digunakan dalam penelitian, meliputi tahapan penelitian, pengumpulan data, pengolahan data, dan implementasi algoritma *Decision Tree*.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari implementasi dan pengujian sistem pendeteksi *phishing* yang telah dibangun, serta pembahasan mengenai hasil yang diperoleh.

BAB V KESIMPULAN DAN SARAN

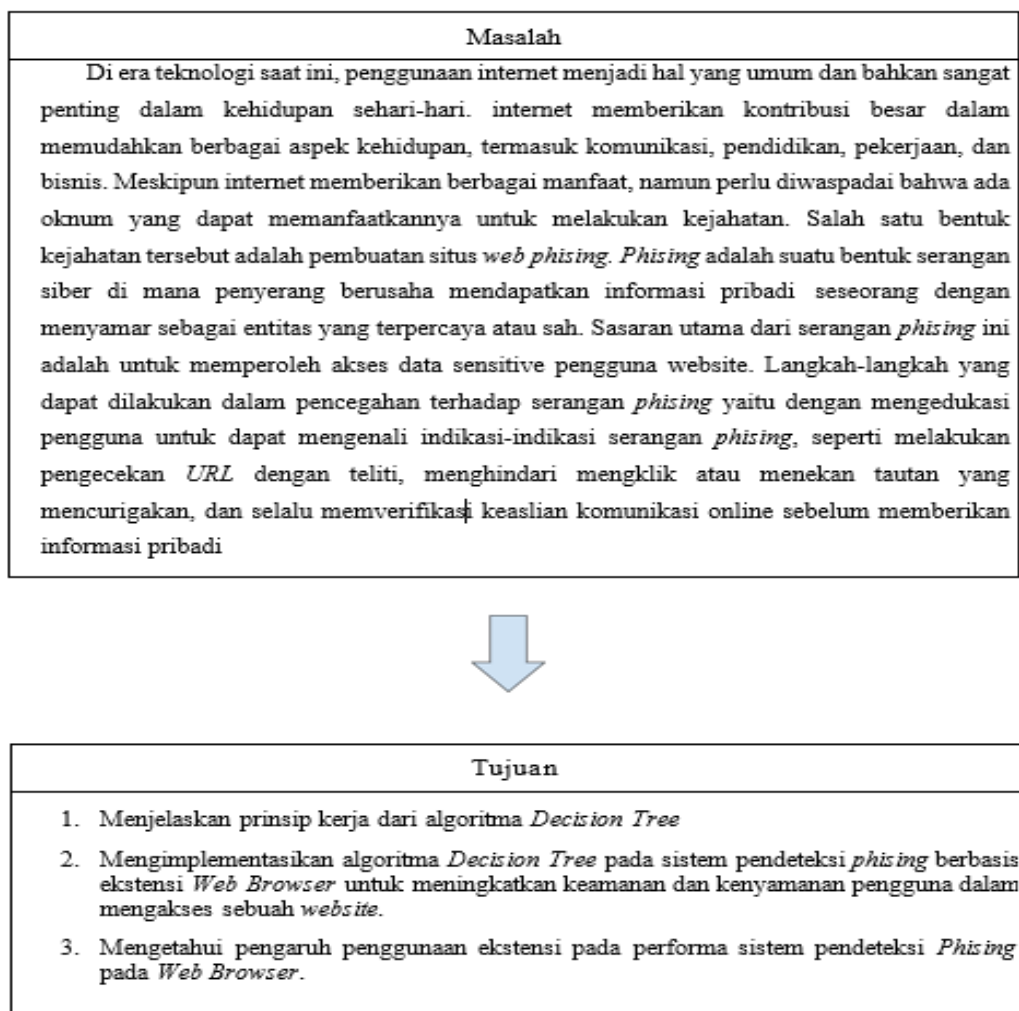
Bab ini berisi kesimpulan dari penelitian yang dilakukan serta saran untuk pengembangan lebih lanjut dalam bidang ini.

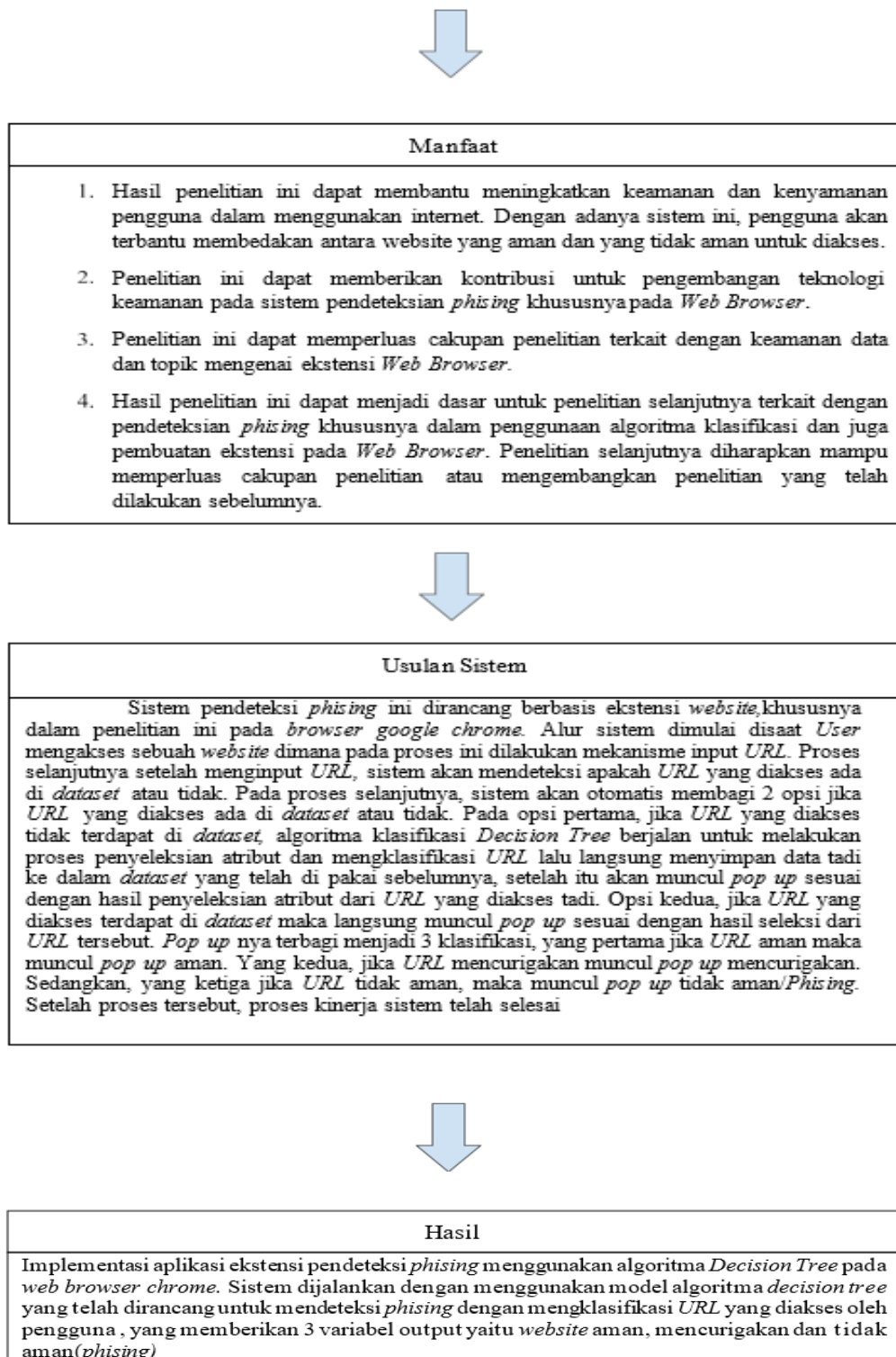
BAB II

TINJAUAN PUSTAKA

2.1 Kerangka Pikir

Agar memperjelas cara berfikir penulis dalam memberikan Solusi dari permasalahan yang di hadapi, maka penulis membuat suatu kerangka pikir. Kerangka pikir penelitian ni di buat dalam bentuk skema seperti pada Gambar 2.1





Gambar 2. 1 Kerangka Pikir

2.2 Landasan Teori

2.2.1 *Phishing*

Menurut (Heriani, Fitri Novia 2021), *phishing* berasal dari kata dalam bahasa Inggris "fishing" (memancing), yang dianalogikan dengan "memancing" korban untuk memberikan informasi sensitif secara tidak sadar melalui rekayasa sosial (social engineering). *Phishing* merupakan upaya untuk memperoleh data sensitif seseorang, seperti kata sandi, informasi kartu kredit, atau data pribadi lainnya, dengan cara penipuan yang menyamar sebagai pihak terpercaya. Fokus utama serangan *phishing* adalah manipulasi psikologis korban, di mana penyerang berusaha membuat korban secara sukarela menyerahkan informasi pribadi mereka melalui media seperti situs web palsu, email penipuan, atau pesan teks. Seiring perkembangan zaman, tindak kriminal juga semakin merebak di seluruh dunia. Sehingga ancaman yang banyak terjadi saat ini juga melalui komputer. Bagi hacker cara ini merupakan cara paling mudah untuk di jadikan serangan. Meskipun di anggap mudah dan sepele tapi tetap saja ada pengguna yang masuk ke perangkap sang hacker (Mia Haryati,2017).

Spoofing berbeda dengan *phishing*, *spoofing* adalah tindakan manipulasi teknis dengan memalsukan identitas sumber komunikasi, seperti alamat email, URL situs web, nomor telepon, atau domain, agar terlihat seolah-olah berasal dari pihak yang sah dan dapat dipercaya. Sebagai contoh analoginya, dalam sebuah serangan, korban mungkin menerima email yang tampaknya berasal dari bank resmi, yang meminta mereka untuk mengklik

tautan, serangan ini disebut *phishing*. Tautan tersebut kemudian mengarahkan korban ke situs web palsu yang dirancang menyerupai situs bank asli, dan ini disebut *spoofing*. Dalam situasi ini, *phishing* berperan dalam menarik perhatian korban dan memanipulasi mereka secara psikologis, sementara *spoofing* mendukung serangan dengan menyamarkan situs palsu agar terlihat meyakinkan. Kombinasi keduanya sering digunakan untuk mencuri informasi login atau data penting lainnya dari korban.

2.2.2 Algoritma *Machine Learning* dalam Deteksi *Phishing*

Machine Learning (ML) atau pembelajaran mesin merupakan pendekatan dalam AI yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Sesuai namanya, ML mencoba menirukan bagaimana proses manusia atau makhluk cerdas belajar dan menggeneralisasi (Atsuto Seko, Tomoya Maekawa, Koji Tsud, 2014)

Machine Learning merupakan cabang dari kecerdasan buatan (Artificial Intelligence) yang memungkinkan sistem komputer untuk belajar dari data dan meningkatkan kinerja tanpa harus diprogram secara langsung untuk setiap tugas. Dalam konteks deteksi *phishing*, *Machine Learning* mampu mengidentifikasi pola-pola atau fitur-fitur yang membedakan situs *phishing* dari situs web yang sah. Hal ini dapat dilakukan dengan menganalisis atribut-atribut tertentu dari situs web, seperti struktur URL,

protokol keamanan yang digunakan, atau keberadaan subdomain yang mencurigakan.

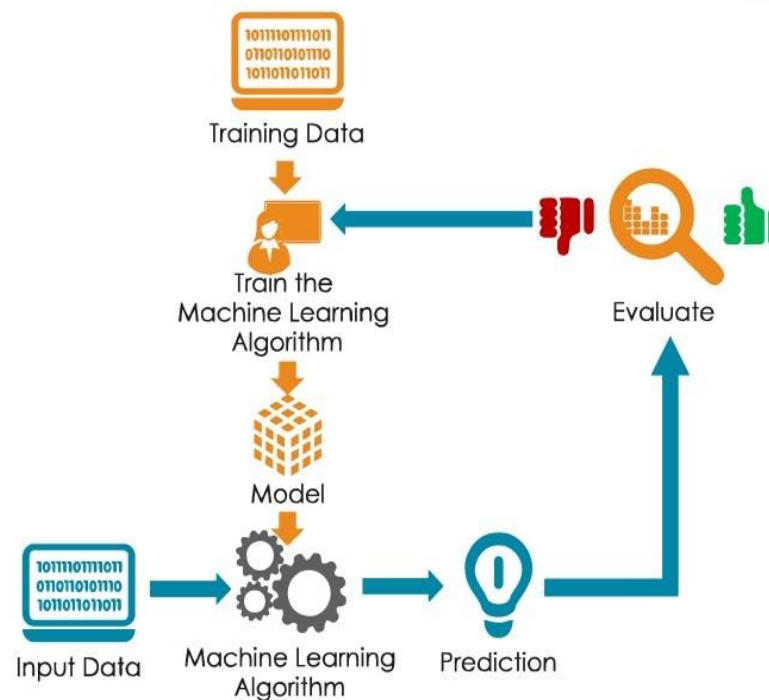
Penerapan *Machine Learning* dalam deteksi *phishing* umumnya melibatkan berbagai algoritma klasifikasi, yang bertugas untuk mengkategorikan URL sebagai *phishing* atau *non-phishing* berdasarkan fitur-fitur yang telah diekstraksi. Beberapa algoritma yang paling sering digunakan dalam deteksi *phishing* antara lain merupakan Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), dan *Decision Tree*. Setiap algoritma memiliki kelebihan dan kekurangannya sendiri, yang harus dipertimbangkan berdasarkan karakteristik data dan tujuan klasifikasi.

Misalnya, algoritma Support Vector Machine (SVM) sangat efektif untuk klasifikasi biner dan bekerja dengan baik pada data berdimensi tinggi, tetapi algoritma ini memerlukan tuning parameter yang cermat untuk mencapai performa yang optimal. Di sisi lain, algoritma Random Forest dan *Decision Tree* cenderung lebih mudah diinterpretasikan, dengan kemampuan untuk menangani data yang memiliki berbagai jenis atribut tanpa perlu preprocessing yang rumit.

Dalam hal deteksi *phishing*, algoritma *Decision Tree* sering kali dipilih karena kemampuannya yang baik dalam menangani masalah klasifikasi biner, serta struktur hasil yang mudah dipahami dan diinterpretasikan oleh pengguna. Selain itu, *Decision Tree* memungkinkan

pembagian dataset menjadi subset yang lebih kecil berdasarkan fitur-fitur yang paling signifikan, yang menjadikannya pilihan yang fleksibel dan efisien untuk mendeteksi situs *phishing*.

Cara Kerja Machine Learning



Gambar 2. 2 Cara Kerja Machine Learning

2.3 Algoritma *Decision Tree*

Metode ini merupakan salah satu metode yang ada pada teknik klasifikasi dalam data mining. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan record.

Atribut menyatakan suatu parameter yang disebut sebagai kriteria dalam pembentukan pohon. Misalkan untuk menentukan main tenis, kriteria yang diperhatikan merupakan cuaca, angin, dan suhu. Salah satu atribut merupakan atribut yang menyatakan data solusi per item data yang disebut atribut hasil. (Acmad & Slamet, 2012). D. Pohon Keputusan ID3 Algoritma ID3 (Iterative Dichotomiser 3) merupakan sebuah metode yang digunakan dalam membuat pohon keputusan, dimana algoritma ID3 ini oleh J. Ross Quinlan pada sekitar akhir 1970-an dan awal 1980-an. Algoritma pada metode ini menggunakan konsep dari entropy informasi, dimana Algoritma ID3 melakukan pencarian secara menyeluruh (greedy) pada semua kemungkinan pohon keputusan. Defiyanti & Pardede, 2010).

Decision Tree merupakan algoritma klasifikasi yang sangat populer dan banyak digunakan dalam berbagai tugas *Machine Learning*, termasuk deteksi *phishing*. Algoritma ini bekerja dengan memecah dataset menjadi subset yang lebih kecil berdasarkan atribut-atribut tertentu, yang kemudian disusun dalam bentuk pohon keputusan. Setiap cabang dari pohon keputusan ini mewakili pilihan yang dibuat berdasarkan atribut tertentu, sedangkan daun-daunnya mewakili hasil akhir klasifikasi, seperti apakah situs tersebut *phishing* atau tidak.

2.3.1 Prinsip Kerja *Decision Tree*

Algoritma *Decision Tree* memulai prosesnya dengan memecah dataset menjadi beberapa subset yang lebih kecil, berdasarkan nilai-nilai dari atribut-atribut yang paling signifikan. Proses ini diulang secara rekursif hingga setiap subset data terklasifikasikan secara penuh ke dalam kategori yang sesuai. Pada setiap simpul pohon, algoritma menentukan atribut mana

yang paling relevan untuk membagi data, berdasarkan kriteria seperti entropy atau gini index, yang digunakan untuk menghitung ketidakpastian atau impurity dalam dataset.

Setiap simpul pohon keputusan merupakan tempat di mana dataset dibagi menjadi dua atau lebih kelompok berdasarkan atribut yang terpilih. Cabang-cabang pohon akan terus berkembang hingga mencapai daun-daun pohon, yang merupakan kategori akhir dalam klasifikasi data, seperti *phishing* atau *non-phishing*. Hasil akhirnya merupakan pohon keputusan yang mampu mengklasifikasikan data baru berdasarkan atribut-atribut yang dipelajari dari dataset awal.

2.3.2 Kelebihan dan Kekurangan *Decision Tree*

Algoritma *Decision Tree* memiliki sejumlah kelebihan yang menjadikannya sangat berguna dalam tugas klasifikasi, termasuk dalam pendeteksian *phishing*. Salah satu keunggulan utama dari algoritma ini merupakan kemampuannya untuk bekerja dengan berbagai jenis data dan atribut, serta hasil klasifikasinya yang mudah dipahami dan diinterpretasikan. Struktur pohon keputusan yang intuitif memungkinkan pengguna untuk melihat secara langsung bagaimana keputusan diambil, yang sangat membantu dalam konteks di mana interpretasi model menjadi penting.

Namun, seperti halnya algoritma lainnya, *Decision Tree* juga memiliki kekurangan. Salah satu masalah utama yang sering dihadapi oleh algoritma ini merupakan *overfitting*, yaitu ketika model terlalu spesifik

terhadap data latih sehingga kehilangan kemampuan untuk memprediksi data baru dengan baik. Pohon yang terlalu dalam sering kali mempelajari detail-detail yang tidak signifikan dalam data latih, yang dapat mengurangi performa model ketika diuji pada data uji. Untuk mengatasi masalah ini, teknik pruning sering digunakan untuk memotong cabang-cabang pohon yang tidak signifikan, sehingga pohon menjadi lebih sederhana dan mampu generalisasi lebih baik.

2.3.3 Decision Tree dalam Deteksi Phishing

Dalam pendeteksian *phishing*, algoritma *Decision Tree* sangat berguna karena mampu membagi dataset URL berdasarkan fitur-fitur penting yang ada pada situs web yang dicurigai sebagai *phishing*. Beberapa atribut yang umum digunakan dalam deteksi *phishing* meliputi :

1. Panjang URL: URL *phishing* sering kali lebih panjang daripada URL asli karena penggunaan parameter tambahan untuk menyamarkan domain palsu.
2. Jumlah -: Situs *phishing* cenderung menggunakan banyak subdomain untuk meniru situs web sah, sementara situs resmi biasanya memiliki struktur domain yang lebih sederhana.
3. Penggunaan HTTPS: Banyak situs *phishing* tidak menggunakan protokol keamanan HTTPS yang sah, atau menggunakan sertifikat yang tidak valid, yang dapat menjadi tanda bahaya.

Dengan mempelajari pola-pola ini dari dataset yang ada, algoritma *Decision Tree* dapat menghasilkan model yang mampu memprediksi apakah suatu URL merupakan *phishing* atau bukan, berdasarkan atribut-atribut yang telah dianalisis.

2.4 Website

Website merupakan situs informasi yang dapat di gunakan banyak orang dengan cepat. Website ada karena perkembangan zaman dari bidang teknologi komputer. Website sudah menjadi sebuah media penyebaran informasi bagi perusahaan, organisasi, dan sekolah (Rahardja et al., 2018). Pada pembuatan *Website* biasanya digunakan bahasa pemrograman ccs dan html, dijelaskan dalam buku Pemrograman Web : Html Dan Css bahwa Hyper Text Markup Language (HTML) merupakan bahasa pemrograman yang digunakan sebagai perancangan halaman *Website*. Dalam dunia pemrograman berbasis *Website*, HTML disebut sebagai pondasi dasar halaman *Website*. Sebuah file HTML disimpan dengan ekstensi . Tahapan Proses waterfall (Yagoyamu, 2020). File itu kemudian dapat di akses menggunakan web *Browser*, sedangkan css merupakan bahasa pemrograman yang digunakan sebagai web design. Dalam mendesign *Website*, css menggunakan sebuah penanda yang biasanya kita kenal dengan id dan juga class (Herho, 2018).

2.5 Ekstensi Browser untuk Deteksi *Phishing*

Penggunaan ekstensi *Browser* merupakan salah satu solusi yang paling efektif untuk memberikan perlindungan terhadap serangan *phishing*. Ekstensi ini bekerja sebagai alat tambahan yang dipasang pada *Browser*, dan secara otomatis memantau URL yang diakses oleh pengguna. Jika suatu URL dicurigai sebagai

phishing berdasarkan analisis atribut-atribut tertentu, ekstensi akan memberikan peringatan kepada pengguna agar mereka tidak memasukkan informasi sensitif di situs tersebut.

Ekstensi *Browser* modern dapat diintegrasikan dengan algoritma *Machine Learning*, seperti *Decision Tree*, untuk meningkatkan kemampuan deteksi *phishing*. Ekstensi ini akan menganalisis URL, dan berdasarkan model klasifikasi yang telah dibangun, akan menentukan apakah URL tersebut aman atau berbahaya. Jika terdeteksi bahwa situs web tersebut *phishing*, ekstensi akan menampilkan notifikasi atau peringatan yang jelas kepada pengguna.

Langkah-langkah utama yang dilakukan oleh ekstensi *Browser* dalam mendeteksi *phishing* meliputi:

1. Pengambilan URL: Ekstensi secara otomatis mengambil URL yang diakses oleh pengguna saat berselancar di internet.
2. Klasifikasi URL: Menggunakan model *Machine Learning*, seperti *Decision Tree*, ekstensi akan menganalisis URL berdasarkan fitur-fitur yang telah dipelajari dari dataset.
3. Pemberian Notifikasi: Jika URL diklasifikasikan sebagai *phishing*, ekstensi akan memberikan notifikasi atau peringatan kepada pengguna untuk berhati-hati dan tidak memasukkan informasi sensitive

2.6 Python

Python merupakan suatu bahasa pemrograman *open source multiplatform* yang bisa dipakai kepada berbagai sistem operasi seperti MacOS, Windows, dan Linux. Tidak hanya itu, Bahasa pemrograman ini bersifat fleksibel dan gampang untuk dipelajari. Pemrograman yang ditulis pada python biasanya lebih mudah dibaca dan lebih ringkas dibandingkan dengan bahasa C. *Python* memiliki modul standar yang menyediakan sejumlah besar algoritma dan fungsi, untuk menyelesaikan pekerjaan seperti mengurai data teks dan mengunduh data dari web server. Dengan menggunakan Bahasa pemrograman *python*, Programmer dengan mudah menerapkan teknik komputasi tingkat lanjut, seperti pemrograman berorientasi objek (Herho, 2017).

2.7 Java Script

JavaScript adalah bahasa pemrograman web yang bersifat Client Side Programming Language. Client Side Programming Language adalah tipe bahasa pemrograman yang pemrosesannya dilakukan oleh client. Aplikasi client yang dimaksud merujuk kepada web browser seperti Google Chrome dan Mozilla Firefox. Bahasa pemrograman Client Side berbeda dengan bahasa pemrograman Server Side seperti PHP, dimana untuk server side seluruh kode program dijalankan di sisi server. Untuk menjalankan JavaScript, kita hanya membutuhkan aplikasi text editor dan web browser. JavaScript memiliki fitur: high-level programming language, client-side, loosely typed dan berorientasi objek (Agung, 2012).

Javascript merupakan bahasa pemrograman berbasis web dan berorientasi objek atau sering juga disebut OOP (Object Oriented Programming). Dimana

dianggap sebuah objek memiliki metode, properti dan event yang berbeda. Contohnya ketika kita mengklik tombol maka akan muncul sebuah pesan peringatan. Ketika kursor melintasi link muncul pesan. Itulah beberapa contoh OOP sederhana. Sebenarnya isi dari language tidak hanya javascript tetapi anda juga dapat menggunakan Vbscript. Yaitu bahasa pemrograman berbasis Visual Basic Script. Kita tidak membahas vbscript karena tidak kompatible dengan browser selain Internet Explorer. Perlu juga anda ketahui sedikit pengetahuan tentang javascript akan sangat membantu anda dalam memahami bahasa pemrograman lainnya seperti PHP karena syntaxnya hampir mirip. Javascript biasanya gunakan untuk event-event tertentu.

2.8 Node JS

NodeJs adalah perangkat lunak yang didesain untuk mengembangkan aplikasi berbasis web dan ditulis dalam sintaks bahasa pemrograman JavaScript. Bila selama ini kita mengenal JavaScript sebagai bahasa pemrograman yang berjalan disisi client / browser saja, maka Node.js ada untuk melengkapi peran JavaScript sehingga bisa juga berlaku sebagai bahasa pemrograman yang berjalan disisi server, seperti halnya PHP, Ruby, Perl, dan sebagainya. Node.js dapat berjalan di sistem operasi Windows, Mac OS X dan Linux tanpa perlu ada perubahan kode program. Node.js memiliki pustaka server HTTP sendiri sehingga memungkinkan untuk menjalankan server web tanpa menggunakan program server web seperti Apache atau Nginx. Untuk mengeksekusi Javascript sebagai bahasa server diperlukan engine yang cepat dan mempunyai performansi yang bagus.

Engine Javascript dari Google bernama V8-lah yang dipakai oleh Node.js yang juga merupakan engine yang dipakai oleh browser Google Chrome (Faisal, 2017).

2.9 Visual Studio Code

Visual Code Studio (VCS) merupakan *software* dengan tujuan membuat aplikasi. Selain VCS, perangkat lunak android studio juga digunakan untuk emulator sebagai testing program setelah dibuild. (Hamzan et al., 2022).

2.10 Penelitian Terkait

Sejumlah penelitian sebelumnya telah dilakukan dalam upaya untuk mendeteksi *phishing* menggunakan pendekatan *Machine Learning*, termasuk algoritma *Decision Tree*. Beberapa penelitian penting yang relevan dengan penelitian ini antara lain:

Tabel 2. 1 Penelitian Terkait

Judul	Penulis	Penerbit	Hasil
Implementasi Ekstensi Google Chrome Dalam Mendeteksi Situs Web <i>Phishing</i> Menggunakan Algoritma Random Forest	Al Ghifari, M. A. G., Hananto, B., & Wahyono, B. T. (2022, August)	Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya (Vol. 3, No. 2, pp. 640-649)	Hasil dari penelitian ini menunjukkan bahwa akurasi dari algoritma klasifikasi <i>random forest</i> untuk pendeteksian situs <i>Phising</i> sangatlah baik di angka 88%, recall 84% dan presisi 91,3%. Penulis menyarankan untuk penelitian selanjutnya menggunakan <i>dataset</i> yang mempunyai data lebih banyak daripada <i>dataset</i> yang digunakan agar proses klasifikasi nya lebih

Judul	Penulis	Penerbit	Hasil
			akurat.
Pengembangan Aplikasi Deteksi <i>Phising</i> Menggunakan Algoritma <i>Decision Tree</i> Berbasis Website	Rahul, M. (2023)	Doctoral dissertation, UIN Ar-Raniry Banda Aceh	<p>Penelitian ini berfokus pada sistem pendeteksi <i>Phising</i> menggunakan algoritma <i>Decision Tree</i> berbasis Website. Hasil penelitian ini menunjukkan tingkat akurasi yang lebih baik dari penelitian sebelumnya dengan algoritma yang sama namun dengan data penelitian yang lebih banyak dan lebih akurat. Diharapkan bagi yang ingin melanjutkan penelitian ini dapat membuat aplikasi deteksi <i>Phising</i> yang lebih kompleks seperti berbasis android dan juga membandingkan beberapa algoritma klasifikasi dan menentukan algoritma mana saja yang lebih efektif dalam mendeteksi serangan <i>Phising</i>.</p>
Detection of <i>Phishing</i> Websites using <i>Machine Learning</i>	Razaque, A., Frej, M. B. H., Sabyrov, D., Shaikhyn, A., Amsaad, F., & Oun, A.	IEEE Cloud Summit (pp. 103-107)	<p>Penelitian ini mengimplementasikan Ekstensi <i>Anti-Phising</i> menggunakan algoritma klasifikasi. Penelitian ini menggunakan 3 metode klasifikasi untuk membandingkan Teknik mana yang</p>

Judul	Penulis	Penerbit	Hasil
	(2020, October)		<p>memiliki akurasi paling baik dalam melakukan deteksi <i>phishing</i>, yaitu Regresi Logistik, Artificial Neural Network, dan XG boost.</p> <p>Tujuan dari penelitian ini merupakan untuk memprediksi apakah URL yang diberikan merupakan situs web <i>phishing</i> atau bukan. Ternyata dalam percobaan yang diberikan bahwa pengklasifikasi berbasis XG Boost merupakan pengklasifikasi terbaik dengan akurasi klasifikasi 94% untuk kumpulan data situs <i>phishing</i> tertentu.</p>
<p>Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website <i>Phishing</i></p>	<p>Putri, N. B., & Wijayanto, A. W. (2022)</p>	<p>Komputika: Jurnal Sistem Komputer, 11(1), 59-66</p>	<p>Dari hasil penelitian dengan dataset <i>Website phishing</i> sebanyak 1.353 data dan terdiri dari 10 variabel yang kemudian dievaluasi dengan confusion matrix, didapatkan bahwa algoritma Random Forest menghasilkan model dengan nilai akurasi yang lebih baik dari pada algoritma Naïve Bayes, <i>Decision Tree</i>, dan Support Vector Machine yaitu sebesar 90,77%. Sedangkan algoritma Naïve Bayes</p>

Judul	Penulis	Penerbit	Hasil
			menghasilkan model yang memiliki nilai akurasi terendah yaitu sebesar 82,31
Detection of <i>Phishing</i> Websites Using <i>Machine Learning</i>	Devan, K P K and R, Adithya Nagaraj and P, Kamaleshwaran and S, Karthick. (February 22, 2023)	International Conference on Innovative Computing & Communication (ICICC) 2022	<p>Penelitian ini bertujuan untuk membantu mengidentifikasi situs <i>Phishing</i> secara efektif dan efisien dengan tingkat akurasi yang tinggi.</p> <p>Penelitian ini menggunakan atribut URL untuk melakukan proses klasifikasi situs web <i>Phishing</i>, dengan membandingkan beberapa algoritma yaitu <i>Decision Tree</i>, <i>XGBooster</i>, dan <i>Random Forest</i>. Hasil dari penelitian ini menunjukkan bahwa algoritma <i>Random Forest</i> memiliki akurasi tertinggi dengan 98,4% , diikuti algoritma <i>Decision Tree</i> dengan 97,31% dan algoritma <i>XGBooster</i> dengan 86,6%</p>