BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi internet telah memberikan kemudahan akses informasi bagi seluruh lapisan masyarakat, termasuk anak dan remaja. Internet tidak hanya digunakan sebagai media komunikasi dan pembelajaran, tetapi juga sebagai sarana hiburan, transaksi, serta pencarian informasi sehari-hari. Namun, di sisi lain, kemudahan akses tersebut juga meningkatkan risiko paparan terhadap situs web ilegal dan berbahaya, seperti konten pornografi atau dewasa, situs *phishing*, *malware*, penipuan digital, serta bentuk konten lain yang tidak sesuai dengan usia pengguna. Permasalahan ini menjadi semakin penting karena anak-anak merupakan kelompok pengguna yang rentan terhadap paparan konten negatif dan eksploitasi di ruang digital.

Di Indonesia, perlindungan anak dari konten berbahaya di internet menjadi perhatian serius berbagai pihak. Berdasarkan pernyataan Komisi Penyiaran Indonesia (KPI), terdapat jutaan anak yang menjadi korban paparan pornografi dalam kurun waktu beberapa tahun terakhir [1]. Selain itu, laporan yang mengacu pada National Center for Missing and Exploited Children (NCMEC) menunjukkan bahwa Indonesia termasuk salah satu negara dengan jumlah kasus eksploitasi anak di ranah digital yang tinggi, bahkan berada pada peringkat atas di kawasan ASEAN dan dunia [2]. Kondisi tersebut menunjukkan bahwa risiko anak terhadap paparan konten ilegal dan aktivitas berbahaya di internet masih berada pada tingkat yang mengkhawatirkan. Pemerintah dan lembaga perlindungan anak juga terus mendorong penguatan regulasi serta mekanisme perlindungan anak di ruang digital agar penggunaan internet dapat dilakukan dengan lebih aman [3].

Selain risiko paparan konten dewasa, penggunaan internet juga memiliki risiko terhadap keamanan pengguna melalui situs *phishing* dan *malware*. Situs *phishing* umumnya dirancang untuk meniru layanan resmi agar pengguna memasukkan informasi sensitif, seperti akun, kata sandi, atau data keuangan. Sementara itu, situs

malware dapat digunakan untuk menyebarkan perangkat lunak berbahaya melalui tautan unduhan, berkas palsu, script, atau halaman yang telah dimodifikasi. Kedua jenis ancaman ini dapat membahayakan pengguna karena tidak hanya berkaitan dengan kelayakan konten, tetapi juga dengan pencurian data, penyalahgunaan akun, dan infeksi perangkat [49], [50].

Untuk memperjelas konteks permasalahan dan kebutuhan pengguna, identifikasi *stakeholder* dilakukan melalui wawancara dengan salah satu orang tua yang memiliki anak berusia 12 tahun. Berdasarkan keterangan narasumber, anak menggunakan PC berbasis Windows 11 dengan Google Chrome sebagai peramban utama. Usia 12 tahun digunakan sebagai konteks pengguna yang menjadi acuan dalam penelitian dan tidak dimaksudkan untuk mewakili seluruh kelompok usia anak. Pada kondisi penggunaan yang diamati, PC belum dikonfigurasi secara khusus untuk memisahkan hak akses orang tua dan anak karena sebagian besar pengaturan perangkat dilakukan sendiri oleh anak. Oleh karena itu, dalam skenario rancangan sistem penelitian ini, orang tua diposisikan sebagai pihak yang memiliki hak administrator untuk memasang, mengaktifkan, dan mengelola proteksi, sedangkan anak diposisikan sebagai pengguna yang dilindungi melalui akun non-administrator.

Berdasarkan hasil wawancara, orang tua mengamati bahwa anak lebih sering menggunakan PC untuk bermain gim serta menonton YouTube atau *anime*. Orang tua juga pernah menemukan anak mengakses situs pornografi. Ketika ditanyakan mengenai cara mengakses situs tersebut, anak menyampaikan bahwa akses dilakukan menggunakan aplikasi Turbo VPN. Temuan tersebut menjadi dasar penetapan situs dewasa sebagai salah satu kategori utama yang diblokir dan penggunaan VPN *desktop* sebagai kondisi pengujian *bypass* dalam penelitian.

Dalam penelitian ini, *bypass* didefinisikan sebagai upaya pengguna untuk menghindari mekanisme proteksi dengan tujuan memperoleh akses terhadap URL yang seharusnya diblokir. Percobaan *bypass* dinyatakan berhasil apabila URL kategori *phish, malware,* atau *adult* dapat dimuat melalui peramban yang diproteksi ketika proteksi aktif. Sebaliknya, percobaan *bypass* dinyatakan gagal apabila pengaya tetap mengalihkan akses menuju halaman pemblokiran. Skenario bypass

utama yang dievaluasi dalam penelitian ini adalah penggunaan aplikasi VPN desktop ketika pengguna mengakses URL berbahaya melalui Google Chrome yang telah dipasangi pengaya.

Dalam penggunaan internet secara umum, pengguna dapat mencapai URL melalui hasil mesin pencari, iklan, tautan pada media sosial atau aplikasi komunikasi, tombol unduhan, serta pengalihan dari halaman lain. Oleh karena itu, sistem dirancang untuk memeriksa URL lengkap yang dapat mencakup *hostname*, subdomain, *path*, parameter, dan URL alias. Mekanisme akses tersebut merupakan pertimbangan rancangan sistem dan tidak seluruhnya berasal dari hasil wawancara orang tua dari *stakeholder*.

Dalam skenario penggunaan sistem, orang tua memasang perangkat lunak dan *browser extension* menggunakan hak administrator, kemudian mengaktifkan serta mengelola proteksi melalui perangkat lunak desktop. Anak selanjutnya menggunakan akun non-administrator untuk mengakses internet melalui peramban yang diproteksi, sedangkan URL yang dibuka diperiksa oleh pengaya sebelum akses diizinkan atau dialihkan menuju halaman pemblokiran. Kategori klasifikasi dalam penelitian ini dibatasi pada *benign*, *phish*, *malware*, dan *adult* sesuai dengan label dataset yang digunakan. Kategori *benign* merepresentasikan URL aman, sedangkan kategori *phish* merepresentasikan URL yang berkaitan dengan aktivitas *phishing*. Perjudian daring, kekerasan, dan jenis konten negatif lainnya tidak diklasifikasikan sebagai kelas tersendiri oleh model.

Ancaman berbasis URL juga terus berkembang karena situs *phishing*, *malware*, dan konten berbahaya dapat menggunakan domain baru, subdomain berbeda, URL panjang, karakter acak, serta teknik penyamaran agar tidak mudah dikenali oleh sistem pemblokiran statis. Oleh sebab itu, sistem pembatasan akses situs web tidak cukup hanya berfokus pada konten dewasa, tetapi juga perlu mampu mengenali beberapa kategori URL, yaitu URL aman, *phishing*, *malware*, dan *adult*. Pendekatan klasifikasi multikelas berbasis URL menjadi relevan karena sistem dapat mempelajari pola struktural dan leksikal dari URL untuk membedakan beberapa jenis ancaman tersebut [51], [52].

Berbagai mekanisme pembatasan akses konten telah diterapkan untuk mengurangi risiko tersebut, seperti pemfilteran DNS, *firewall* jaringan, pencarian aman, *Blacklist* domain, serta fitur kontrol orang tua atau administrator bawaan sistem operasi atau peramban. Mekanisme tersebut umumnya bekerja dengan membatasi akses berdasarkan alamat domain, alamat IP, kategori situs, atau aturan statis yang telah ditentukan sebelumnya. Pendekatan ini bermanfaat dalam kondisi tertentu, terutama ketika sistem masih dapat mengenali alamat tujuan akses pengguna. Namun, mekanisme pemblokiran berbasis jaringan dapat memiliki keterbatasan ketika pengguna memanfaatkan Virtual Private Network (VPN). VPN mengenkripsi lalu lintas data dan mengalihkan koneksi melalui server lain sehingga sebagian sistem pemblokiran yang diterapkan pada tingkat jaringan atau router dapat mengalami kesulitan dalam mengenali tujuan akses pengguna [4].

Penelitian ini tidak menyatakan bahwa seluruh sistem kontrol orang tua yang telah tersedia tidak efektif. Efektivitas mekanisme tersebut bergantung pada lapisan tempat pembatasan diterapkan. VPN dapat melewati sebagian pemfilteran situs yang diterapkan pada tingkat jaringan atau *router* karena lalu lintas pengguna dienkripsi dan dialihkan melalui server lain, tetapi tidak selalu dapat melewati pembatasan yang diterapkan langsung pada perangkat atau akun pengguna [56]. Survei terhadap 1.000 orang tua yang memiliki anak berusia 11–18 tahun di Inggris juga menunjukkan bahwa 30% orang tua yang telah menerapkan fitur keamanan melaporkan bahwa anak mereka pernah menghindari pembatasan digital, sedangkan sebagian orang tua belum mengetahui bahwa VPN dan peramban tersembunyi dapat digunakan untuk menghindari mekanisme tertentu [57]. Meskipun temuan tersebut berasal dari konteks negara dan perangkat yang berbeda, hasilnya menunjukkan bahwa upaya menghindari pembatasan merupakan permasalahan yang perlu diperhatikan.

Selain permasalahan VPN, pendekatan berbasis *Blacklist* juga memiliki keterbatasan dalam menghadapi pertumbuhan situs berbahaya yang sangat cepat. *Blacklist* hanya dapat mengenali URL atau domain yang telah diketahui sebelumnya dan tercatat dalam daftar. Sementara itu, situs *phishing*, *malware*, dan konten ilegal sering kali menggunakan domain baru, *subdomain* berbeda, struktur

URL yang berubah, atau teknik penyamaran tertentu agar tidak mudah terdeteksi. Akibatnya, situs yang belum tercatat dalam daftar dapat tetap lolos dari pemblokiran meskipun memiliki pola yang berbahaya. Kondisi ini menunjukkan bahwa mekanisme statis tidak selalu cukup untuk menghadapi ancaman web yang terus berubah [5].

Berdasarkan kondisi tersebut, permasalahan yang dibahas dalam penelitian ini terdiri atas dua aspek. Aspek pertama adalah keterbatasan deteksi, yaitu mekanisme berbasis daftar statis belum tentu dapat mengenali URL baru atau URL yang mengalami perubahan struktur. Aspek kedua adalah keterbatasan penerapan proteksi, yaitu adanya peluang bagi pengguna untuk mencoba menghindari mekanisme perlindungan melalui perubahan jalur akses atau konfigurasi perangkat. Oleh karena itu, penelitian ini menggabungkan klasifikasi berbasis *machine learning* untuk mendukung deteksi URL dengan *browser extension* dan perangkat lunak desktop untuk menerapkan serta mengelola proteksi pada perangkat pengguna.

Machine learning memungkinkan sistem melakukan klasifikasi berdasarkan fitur-fitur yang diekstraksi dari URL, seperti panjang URL, jumlah titik, jumlah *subdomain*, penggunaan HTTPS, keberadaan karakter khusus, pola digit, kata kunci mencurigakan, struktur *path*, serta metadata lain yang relevan. Dengan pendekatan ini, sistem tidak hanya bergantung pada daftar situs yang telah diketahui, tetapi juga dapat mengenali pola yang mengindikasikan URL berbahaya. Beberapa penelitian terbaru menunjukkan bahwa *machine learning* mampu digunakan untuk klasifikasi URL berbahaya secara multikelas, seperti *benign*, *phishing*, *malware*, dan kategori berbahaya lainnya dengan performa yang baik [6]. Selain itu, pendekatan deep learning dan large language model juga menunjukkan potensi dalam klasifikasi URL karena mampu menangkap pola karakter dan hubungan semantik dalam URL [7].

Meskipun demikian, banyak penelitian mengenai klasifikasi URL masih berfokus pada evaluasi model dan belum sepenuhnya membahas bagaimana model tersebut diimplementasikan sebagai sistem proteksi yang dapat digunakan langsung oleh pengguna. Padahal, peramban merupakan salah satu aplikasi utama yang

digunakan anak dan remaja untuk mengakses internet. Oleh karena itu, integrasi model klasifikasi URL ke dalam *browser extension* menjadi pendekatan yang relevan karena pengaya dapat membaca URL yang diakses pengguna dan memberikan respons langsung ketika URL dikategorikan berbahaya. Beberapa penelitian juga menunjukkan bahwa *browser extension* dapat digunakan sebagai media implementasi deteksi URL berbasis *machine learning* secara *real-time* [8], [9].

Namun, penggunaan *browser extension* saja masih memiliki celah dari sisi penerapan. Pengaya yang dipasang secara manual masih dapat dinonaktifkan atau dihapus oleh pengguna. Selain itu, pengguna juga dapat mencoba mengakses situs melalui mode *Incognito*, memasang pengaya VPN atau *proxy*, mengganti profil peramban, atau melakukan perubahan konfigurasi peramban. Celah tersebut menjadi penting dalam konteks kontrol orang tua atau administrator, karena sistem perlindungan tidak hanya perlu mampu mendeteksi URL berbahaya, tetapi juga perlu memiliki mekanisme pendukung agar proteksi tetap aktif dan tidak mudah dinonaktifkan oleh anak atau pengguna non-administrator.

Berdasarkan permasalahan tersebut, penelitian ini tidak hanya berfokus pada model *machine learning* dan *browser extension*, tetapi juga mengembangkan perangkat lunak *desktop* sebagai komponen pendukung proteksi. Perangkat lunak ini dirancang untuk menyediakan fungsi pengelolaan proteksi bagi orang tua atau administrator, meliputi aktivasi dan deaktivasi proteksi, penerapan *browser policy*, pemeriksaan status perlindungan, pemantauan status pengaya, *event logging*, dan pengelolaan *application control*. *Browser policy* pada lingkungan Chromium mendukung pengelolaan pengaya melalui kebijakan, termasuk mekanisme *force-install*, pembatasan pengaya, dan pengaturan perilaku peramban tertentu [10], [11], [12].

Dengan adanya perangkat lunak *desktop*, sistem yang dikembangkan tidak hanya berfungsi sebagai alat klasifikasi URL, tetapi juga sebagai prototipe sistem proteksi yang lebih lengkap. *Browser extension* bertugas melakukan pemeriksaan URL dan menampilkan halaman pemblokiran ketika URL terdeteksi berbahaya, sedangkan perangkat lunak *desktop* bertugas membantu orang tua atau

administrator mengelola proteksi melalui antarmuka untuk mengelola konfigurasi proteksi. Sistem ini juga dilengkapi dengan mekanisme *local agent* untuk memantau status proteksi, mencatat aktivitas penting, serta memberikan notifikasi apabila terjadi kondisi yang memerlukan perhatian, seperti pengaya tidak aktif, proteksi dimatikan, atau konfigurasi peramban berubah.

Penelitian ini mengusulkan sistem klasifikasi dan pemblokiran situs web berbasis URL menggunakan *machine learning* yang diintegrasikan dengan *browser extension* dan perangkat lunak desktop. Pemblokiran dilakukan pada lapisan aplikasi melalui *browser extension*, bukan melalui inspeksi atau intervensi terhadap paket jaringan. Oleh karena itu, penelitian ini mengevaluasi apakah mekanisme pemeriksaan URL pada lapisan peramban tetap dapat menjalankan fungsi pemblokiran ketika jalur koneksi pengguna dialihkan melalui aplikasi VPN desktop.

Evaluasi penelitian difokuskan pada dua aspek utama. Aspek pertama adalah performa model klasifikasi dalam membedakan URL *benign*, *phish*, *malware*, dan *adult*. Evaluasi model menggunakan *macro F1-score* sebagai metrik utama serta akurasi, presisi, *recall*, *weighted F1-score*, matriks konfusi, dan ROC-AUC multikelas sebagai metrik pendukung.

Aspek kedua adalah keberhasilan mekanisme *browser extension* dalam mengizinkan URL *benign* dan memblokir URL kategori *phish*, *malware*, dan *adult* pada kondisi VPN desktop tidak aktif dan aktif. Pengujian juga mencakup variasi representasi URL berbasis *hostname*, alamat IP literal, dan alias. REST API, perangkat lunak desktop, *local agent*, *browser policy*, *application control*, log, *persistence*, dan notifikasi dijelaskan sebagai komponen pendukung dalam perancangan dan implementasi sistem, tetapi tidak diperlakukan sebagai aspek efektivitas yang berdiri sendiri.

.

1.2 Pertanyaan Penelitian

Berdasarkan latar belakang dan permasalahan yang telah dipaparkan, pertanyaan penelitian dalam penelitian ini adalah sebagai berikut:

1.2.1 Bagaimana performa model *machine learning* dalam melakukan klasifikasi multikelas URL ke dalam kategori *benign*, *phish*, *malware*, dan *adult* berdasarkan *macro F1-score* sebagai metrik utama serta akurasi, presisi, *recall*, *weighted F1-score*, matriks konfusi, dan ROC-AUC multikelas sebagai metrik pendukung?

1.2.2 Bagaimana tingkat keberhasilan mekanisme pemblokiran berbasis *browser extension* dalam mengizinkan URL *benign* dan memblokir URL kategori *phish*, *malware*, dan *adult* pada kondisi VPN desktop tidak aktif dan aktif, termasuk pada representasi URL berbasis *hostname*, alamat IP literal, dan alias?

1.3 Batasan Penelitian

Berdasarkan pertanyaan penelitian yang sudah disebutkan, berikut merupakan batasan masalah dari penelitian ini:

1.3.1. Penelitian hanya melakukan klasifikasi berdasarkan URL dan metadata terkait, tanpa melakukan analisis isi halaman web secara penuh.

1.3.2. Kategori klasifikasi yang digunakan dalam penelitian dibatasi pada kelas *benign*, *phish*, *malware*, dan *adult* sesuai dengan *dataset* dan kebutuhan sistem yang dikembangkan.

1.3.3. Sistem pemblokiran dilakukan pada lapisan aplikasi peramban dan antarmuka pengguna melalui *browser extension*, bukan melalui pemutusan koneksi jaringan, inspeksi paket jaringan, atau pemblokiran pada perangkat keras jaringan.

1.3.4. Penelitian tidak bertujuan untuk mendeteksi, mencegah, atau memblokir penggunaan VPN secara langsung. VPN *desktop* digunakan sebagai kondisi pengujian untuk mengevaluasi apakah mekanisme pemblokiran berbasis *browser extension* tetap dapat bekerja ketika VPN tidak aktif dan aktif.

1.3.5. Implementasi *browser extension* difokuskan pada peramban *desktop* berbasis Chromium, seperti Google Chrome.

1.3.6. Perangkat lunak *desktop* difokuskan pada lingkungan sistem operasi Windows dan digunakan untuk membantu konfigurasi *browser policy*,

bukan untuk melakukan modifikasi pada driver, kernel, atau sistem jaringan tingkat rendah.

1.3.7. *Stakeholder* penelitian ditentukan berdasarkan hasil wawancara dengan orang tua yang memiliki anak berusia 12 tahun dan menggunakan perangkat PC berbasis Windows 11.

1.3.8. Mekanisme perlindungan melalui *browser policy* memerlukan hak akses orang tua atau administrator pada perangkat, sehingga penelitian ini tidak membahas skenario ketika pengguna memiliki hak administrator penuh dan secara sengaja mengubah atau menghapus konfigurasi sistem di luar aplikasi.

1.3.9. *Dataset* URL yang digunakan bersumber dari *dataset* publik dan hasil kurasi peneliti, sehingga kualitas model bergantung pada kualitas, distribusi, dan representasi *dataset* yang digunakan.

1.3.10. Skenario *bypass* utama yang diuji dalam penelitian ini adalah penggunaan aplikasi VPN. Mode Incognito/InPrivate, penonaktifan pengaya, dan penggunaan peramban alternatif dibahas sebagai celah proteksi pada lingkungan desktop serta dimitigasi melalui rancangan perangkat lunak desktop dan *browser policy*.

1.3.11. Sistem yang dikembangkan merupakan prototipe penelitian dan belum ditujukan sebagai produk komersial. Implementasi siap produk, seperti publikasi pengaya resmi pada Chrome Web Store atau Microsoft Edge Add-ons untuk memperoleh ID pengaya yang stabil, berada di luar fokus utama pengujian penelitian ini.

1.4 Tujuan Penelitian

Berdasarkan pertanyaan penelitian yang telah dirumuskan, tujuan penelitian ini adalah sebagai berikut:

1.4.1. Mengembangkan dan mengevaluasi model *machine learning* untuk melakukan klasifikasi multikelas URL ke dalam kategori *benign*, *phish*, *malware*, dan *adult* menggunakan *macro F1-score* sebagai metrik utama serta akurasi, presisi, *recall*, *weighted F1-score*, matriks konfusi, dan ROC-AUC multikelas sebagai metrik pendukung.

1.4.2. Mengevaluasi tingkat keberhasilan mekanisme pemblokiran berbasis *browser extension* dalam mengizinkan URL *benign* dan memblokir URL kategori *phish*, *malware*, dan *adult* pada kondisi VPN desktop tidak aktif dan aktif, termasuk pada representasi URL berbasis *hostname*, alamat IP literal, dan alias.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari hasil penelitian ini sebagai berikut:

1.5.1. Memberikan kontribusi ilmiah dalam pengembangan sistem klasifikasi situs web berbasis URL menggunakan *machine learning*, khususnya untuk klasifikasi multikelas pada kategori *benign*, *phish*, *malware*, dan *adult*.

1.5.2. Memberikan alternatif pendekatan pembatasan akses terhadap situs berbahaya yang tidak bergantung sepenuhnya pada mekanisme pemblokiran jaringan, sehingga tetap relevan pada kondisi ketika pengguna menggunakan VPN.

1.5.3. Menunjukkan rancangan sistem proteksi yang menggabungkan model klasifikasi URL, *browser extension*, dan perangkat lunak *desktop* sebagai satu kesatuan sistem.

1.5.4. Menjadi dasar bagi penelitian lanjutan dalam pengembangan sistem keamanan dan proteksi akses situs web yang lebih adaptif, lebih sulit dinonaktifkan oleh pengguna non-administrator, dan lebih siap untuk diterapkan pada lingkungan *desktop* secara nyata.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian disusun dan dibagi atas lima bab sebagai berikut:

1.6.1. BAB I PENDAHULUAN

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

1.6.2. BAB II TINJAUAN PUSTAKA

Bab ini berisi teori dan penelitian terdahulu yang mendukung penelitian, seperti klasifikasi URL, *machine learning*, *browser extension*, VPN, dan mekanisme UI pemblokiran.

1.6.3. BAB III METODE PENELITIAN

Bab ini berisi metode penelitian, pengumpulan dan *preprocessing dataset*, ekstraksi fitur URL, perancangan model *machine learning*, perancangan arsitektur sistem, perancangan *browser extension*, perancangan perangkat lunak *desktop*, serta alur kerja sistem secara keseluruhan.

1.6.4. BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi implementasi model klasifikasi URL, implementasi API atau layanan prediksi, implementasi *browser extension*, implementasi perangkat lunak *desktop*, hasil evaluasi performa model, hasil pengujian pemblokiran URL, pengujian pada kondisi VPN aktif, serta pengujian fitur pendukung seperti status proteksi dan penerapan *browser policy*.

1.6.5. BAB V SIMPULAN DAN SARAN

Bab ini berisi simpulan penelitian dan saran untuk pengembangan penelitian selanjutnya.