

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan teknologi informasi dan komunikasi telah mendorong kita untuk menggunakan internet dalam berbagai aspek, seperti dalam bidang pendidikan, kreativitas dan jaringan sosial. Hal ini menarik lebih banyak pengguna untuk menggunakan internet, terutama anak-anak. Pertambahan pengguna internet pada usia anak mengakibatkan banyak masalah pada keamanan informasi mereka. Ketersediaan berbagai situs pada jaringan global menyebabkan anak-anak secara sengaja maupun tidak sengaja terpapar informasi yang tidak seharusnya mereka lihat seperti konten dewasa, pelecehan seksual, terorisme, konten SARA(suku, agama, ras, antar golongan), dan kebiasaan tidak bermoral yang dapat mengarahkan sikap anak pada lingkungan yang tidak sehat [1].

Sebagai bentuk tanggapan atas fenomena yang terjadi, dilakukannya berbagai metode pencegahan ancaman dari website yang berbahaya pada tingkat teknologi. Saat ini tersedia aplikasi Extension browser Dodo Kids Browser yang merupakan sistem yang dapat menjembatani antara orang tua dan anak dengan tujuan memudahkan orang tua untuk mengontrol dan mengawasi kegiatan penggunaan internet anak dengan menggunakan komputer atau laptop. Pada Dodo Kids Browser terdapat proses Konten Filter Data Pencarian yang bertujuan untuk menyaring website yang akan diakses apakah terdapat konten bersifat negatif atau tidak, jika bersifat negatif maka penelusuran website akan langsung dihentikan. Dalam melakukan klasifikasi website ke dalam kategori konten bersifat negatif, proses Konten Filter Data Pencarian menggunakan metode *string-pattern-matching*. Metode tersebut melakukan klasifikasi website berdasarkan hasil pencocokan kamus data yang berisikan kata-kata bersifat negatif terhadap informasi/kata-kunci pencarian informasi yang dilakukan melalui web browser [2]. Namun metode *string-pattern-matching* memiliki kelemahan yaitu sering terjadinya salah klasifikasi karena tidak memiliki kemampuan pemahaman semantik terhadap teks

secara keseluruhan [3]. Informasi yang seharusnya layak dikonsumsi anak, salah diklasifikasikan sebagai informasi yang berbahaya, hal tersebut dinamakan dengan *over-blocking*. Kelemahan lainnya yaitu proses *string-pattern-matching* sangat mudah dihindari dengan cara salah mengeja kata. Dalam keilmuan Teknik Informatika untuk mengekstrak pengetahuan yang berguna dari kumpulan konten halaman website dan data penggunaan dapat dilakukan dengan pendekatan Web Data Mining. Pengetahuan yang diperoleh dari hasil Web Data Mining dapat digunakan untuk mengklasifikasikan kategori website [4]. Dengan melakukan Web Data Mining menggunakan algoritma Support Vector Machine pada data penggunaan Dodo Kids Browser, kita bisa menemukan informasi atau pengetahuan yang berguna untuk memudahkan dalam menyaring informasi berbahaya.

Maka dari itu, berdasarkan pemaparan masalah dan fakta di atas diperlukan suatu analisis klasifikasi terhadap data penggunaan aplikasi Dodo Kids Browser yang bertujuan untuk memberikan rekomendasi daftar website yang berbahaya berdasarkan isi konten web tersebut agar nantinya proses filtering terhadap situs web yang dibuka oleh anak dapat dilakukan dengan tepat sasaran. Sistem filtering terhadap situs web yang baik dapat menjauhkan anak dari konten berbahaya yang ada pada internet.

1.2 Rumusan Masalah

Berdasarkan pemaparan fenomena dan fakta di atas, maka perumusan masalah dalam penelitian ini adalah apakah dengan menerapkan Web Data Mining pada data penggunaan aplikasi Dodo Kids Browser dapat menghasilkan rekomendasi daftar website yang berbahaya bagi anak dengan tepat sebagai acuan dalam penyaringan website yang akan diakses

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah melakukan analisis prediktif dengan implementasi Web Data Mining pada data penggunaan aplikasi Dodo Kids Browser pada kasus klasifikasi website yang aman atau berbahaya bagi anak.

Dari maksud yang telah dijelaskan di atas, maka tujuan yang ingin dicapai dari penelitian ini adalah untuk menghasilkan rekomendasi daftar website yang telah terklasifikasi aman atau berbahaya bagi anak sebagai acuan dalam penyaringan website yang akan diakses pada Dodo Kids Browser.

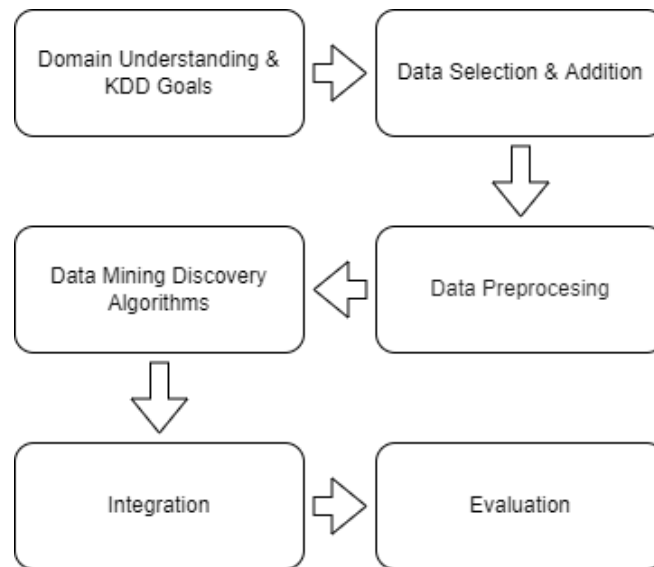
1.4 Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah:

1. Data yang digunakan adalah data penggunaan yang dianonimkan pada sistem aplikasi Dodo Kids Browser tahun 2022.
2. Fitur pada data yang digunakan untuk analisis berasal dari hasil *scrapping* data *meta-tag description*, dan *meta-tag title* pada website [5].
3. Data website yang diambil hanya pada website yang berjenis *server-side rendering* (SSR) dan tidak membutuhkan autentikasi.
4. Data teks yang digunakan adalah data teks bahasa Indonesia dan bahasa Inggris.
5. Klasifikasi yang dilakukan mengarah pada klasifikasi biner, yaitu terdiri dari dua kelas (aman dan berbahaya)
6. Algoritma web data mining yang digunakan adalah Support Vector Machine [4] [6].

1.5 Metodologi Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah Knowledge Discovery in Databases (KDD) [7] yang kemudian disesuaikan berdasarkan penelitian yang dilakukan sehingga menghasilkan sebagaimana yang dijelaskan pada Gambar berikut.



Gambar 1.1 Alur Proses Metode Penelitian

Berikut adalah penjelasan dari masing-masing tahapan :

- 1) *Domain Understanding & KDD Goals* merupakan langkah awal untuk memecahkan masalah. Pengumpulan dan pemahaman informasi dilakukan untuk menentukan fungsionalitas dari pengetahuan yang dihasilkan dan nilai apa yang diberikannya kepada pengguna.
- 2) *Data Selection & Addition* adalah tahapan untuk memilih dan membuat dataset tempat discovery akan dilakukan, termasuk menemukan data yang sudah tersedia dan memperoleh data tambahan yang diperlukan, lalu mengintegrasikan semua data tersebut pada proses KDD pada satu dataset, termasuk atribut yang nanti akan digunakan
- 3) *Data Preprocessing* adalah kegiatan untuk melakukan pembersihan dan transformasi pada data agar siap untuk dimasukkan pada proses pemodelan.
- 4) *Data Mining Discovery Algorithms* penerapan algoritma Support Vector Machine dalam menemukan pengetahuan pada data untuk digunakan dalam klasifikasi website, membuat model klasifikasi (classifier) dan melakukan evaluasi terhadap model tersebut.
- 5) *Integration* adalah tahapan pembuatan sistem secara keseluruhan, mengintegrasikan classifier sehingga bisa menghasilkan rekomendasi daftar

website yang berbahaya bagi anak sebagai acuan dalam penyaringan website yang akan diakses pada Dodo Kids Browser.

- 6) *Evaluation* dilakukan pengujian sistem secara keseluruhan pada percobaan pengguna, untuk mengetahui apakah sistem dapat menghasilkan rekomendasi daftar website yang berbahaya bagi anak sebagai acuan dalam penyaringan website yang akan diakses pada Dodo Kids Browser.

1.6 Sistematika Penulisan

Penelitian ini disusun oleh sistematika penulisan yang dibagi menjadi 5 bab secara sistematis dan sesuai dengan pokok – pokok permasalahan yang dibahas.

1. BAB 1 – Pendahuluan

Pada bab ini akan menjelaskan mengenai latar belakang masalah, perumusan inti masalah, maksud dan tujuan, batasan masalah, metodologi penelitian dan sistematika penulisan.

2. BAB 2 – Landasan Teori

Pada bagian ini akan menjelaskan teori – teori dasar yang akan digunakan pada penelitian ini. Teori – teori di sini adalah teori yang berkaitan dan dalam lingkup tema penelitian ini seperti *data mining*, *web data mining*, *web crawling*, *support vector machine* dan lain sebagainya.

3. BAB 3 – Analisis dan Perancangan

Pada bab ini akan menjelaskan tentang analisis permasalahan untuk mengetahui hal atau masalah apa yang timbul dan mencoba memecahkan masalah tersebut dengan menentukan dan merancang model klasifikasi yang dibutuhkan.

4. BAB 4 – Implementasi dan Pengujian Sistem

Bab ini akan membahas tentang pengimplementasian dari hasil analisis. Pada bagian ini juga akan ditentukan bagaimana teknik penanganan diimplementasi, diuji, dan disesuaikan dengan hasil penelitian.

5. BAB 5 – Kesimpulan dan Saran

Pada bab terakhir ini akan menjelaskan kesimpulan yang diperoleh dari hasil keseluruhan penelitian ini dan juga saran untuk pengembangan ke depannya.