

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada zaman dahulu, surat merupakan salah satu metode untuk berkomunikasi kepada orang di negara lain. Surat digunakan untuk berbagi informasi satu sama lain. Pada masa sekarang, banyak orang menggunakan email untuk memberikan informasi kepada pihak lain. Email adalah surat elektronik yang dikirim dengan menggunakan elektronik seperti komputer melalui internet. (M.Harris, 2021)

Di era digital saat ini, banyak perusahaan memanfaatkan email sebagai sarana utama untuk mentransmisikan data, sehingga memudahkan pengiriman informasi penting kepada pihak terkait dengan efisiensi dan keamanan yang lebih terjamin. Perusahaan ritel menggunakan email untuk mengirimkan iklan kepada konsumen sehingga konsumen dapat memperoleh informasi mengenai produk atau promo yang ditawarkan. Sekolah memanfaatkan email untuk mahasiswa mengirimkan tugas kepada para guru, tanpa memerlukan tatap muka. Email diperlukan untuk website seperti *Facebook*, *Twitter* dan *Youtube* untuk sistem dapat mengidentifikasi pengguna.

Email dapat digunakan secara global sebagai alat komunikasi yang efisien. Salah satu penyedia email yang populer adalah *Gmail*. *Gmail* merupakan layanan email gratis yang disediakan oleh *Google* yang dapat memberikan *pengguna* untuk mengirim atau menerima email. Meskipun email dapat menerima pesan dari

berbagai pengirim, beberapa email yang diterima dapat berupa spam. Spam merupakan masalah signifikan bagi pengguna, di mana banyak perusahaan menggunakan email untuk mengirimkan iklan. Hal ini menyebabkan *inbox* pengguna penuh dengan pesan promosi, sehingga email yang penting dapat tenggelam dan tidak terlihat oleh pengguna. Meskipun spam sering kali dianggap sebagai gangguan, email spam sebenarnya dapat membawa bahaya signifikan bagi pengguna. Para peretas memanfaatkan email untuk menyebarkan virus atau menyertakan link berbahaya yang, ketika dibuka oleh pengguna, dapat menginfeksi perangkat mereka. Selain itu, peretas dapat membuat situs palsu yang dirancang untuk menipu pengguna agar memasukkan data pribadi, yang kemudian dapat disalahgunakan oleh pelaku kejahatan siber.

Penelitian “**Analisis Perbandingan Metode *Naïve Bayes* dan Metode *Support Vector Machine* Terhadap Klasifikasi Kepada Spam Email.**” akan menggunakan aplikasi *Orange3*. *Orange3* merupakan aplikasi *open-source* yang digunakan untuk memberikan visualisasi dan mengolah data (Yunardi et al, 2022). *Orange3* dapat menghitung dataset dengan menggunakan algoritma seperti metode SVM, *Naïve Bayes* dan lain – lain. Spam email memiliki banyak kata dan simbol, maka untuk menganalisis teks, *Orange3* harus mendownload *add-on Corpus*. *Corpus* dapat memproses teks yang berada di dalam email tersebut agar *Orange3* dapat membuat email tersebut menjadi data yang dapat dianalisis oleh program tersebut.

Penelitian ini menggunakan metode *Support Vector Machine* (SVM) dan *Naïve Bayes*. SVM dan *Naïve Bayes* menghitung kata – kata yang muncul di dalam

email tetapi kedua metode tersebut memiliki proses perhitungan tersebut sehingga dapat memiliki hasil yang berbeda. SVM digunakan untuk mengklasifikasikan email spam atau *ham* karena metode ini dapat memisahkan email spam dan *ham* berdasarkan fitur dalam dataset. Sementara itu, *Naïve Bayes* dapat dilatih menggunakan dataset untuk menghitung probabilitas suatu email tergolong spam atau *ham* berdasarkan kata-kata yang terdapat di dalam email. Perbandingan hitungan kedua metode dilakukan dengan melakukan dataset yang berada di website *Kaggle*. Dalam penelitian ini, penulis akan melakukan perhitungan dengan kedua metode dan membandingkan hasil kedua metode sehingga dapat mengetahui metode mana yang lebih akurat untuk mengklasifikasi spam email.

Berdasarkan paparan diatas, penulis akan melakukan penelitian klasifikasi spam email, penulis akan membuat penelitian tentang klasifikasi spam email menggunakan *Naïve Bayes* dan SVM. Penelitian ini mengharap untuk dapat membandingkan hasil kedua metode agar dapat mengetahui hasil metode yang lebih akurat untuk mengklasifikasi spam email.

1.2 Rumusan Masalah

Berdasarkan masalah latar belakang diatas, masalah yang didapatkan oleh penulis adalah:

1. Bagaimana metode *Naïve Bayes* dan *Support Vector Machine* dapat mendeteksi spam email?
2. Bagaimana metode *Naïve Naves* atau *Support Vector Machine* yang lebih efektif untuk spam filter?

1.3 Tujuan Penelitian

Berdasar rumusan masalah yang ditulis di atas, maka tujuan penelitian adalah sebagai berikut:

1. Mengetahui nilai *best score* menggunakan *Random Search* kepada metode *Naïve Bayes* dan *Support Vector Machine*.
2. Mengetahui apakah metode *Naïve Bayes* atau *Support Vector Machine* lebih efektif.

1.4 Batasan Masalah

Batasan masalah yang dibuat dari penulis dari persoalan berikut dan masalah yang dikaji adalah:

1. Berapa spam email menggunakan gambar atau lampiran. Penelitian ini membatasi untuk menggunakan teks sehingga hasil dapat menjadi kurang optimal.
2. Aplikasi dapat menghitung kata *typo* menjadi kata baru.
3. Penelitian ini hanya menggunakan metode SVM dan *Naïve Bayes*.

1.5 Manfaat Penelitian

Berikut adalah manfaat penelitian yang telah dibuat oleh penulis:

1.5.1 Bagi Universitas

Hasil peneliti ini diharapkan oleh penulis untuk digunakan sebagai sumber informasi untuk penelitian masa depan yang dilakukan pada topik serupa.

1.5.2 Bagi Pembaca

Hasil peneliti ini diharapkan oleh penulis untuk dapat menambahkan pengetahuan para pembaca tentang spam dengan mengetahui pembahayaan spam dan mencegah bahaya tersebut.

1.5.3 Sistematika Penulisan

Agar pembaca dapat nyaman untuk membaca penelitian ini, maka sistematika penulisan adalah sebagai berikut:

Bab I Pendahuluan.

Pada bab pendahuluan akan dijelaskan tentang dasar atau sebagai landasan dalam membuat suatu penelitian. Bab ini terdiri dari beberapa sub bab yaitu latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian dan sistematika penulisan.

Bab II Landasan Teori

Bab landasan teori adalah bab yang berisi dengan teori yang mendukung penelitian penulis dan membahas tentang teori yang berisi dengan teori dan hasil penelitian dimana teori dan hasil penelitian tersebut dijadikan sebagai kerangka teori untuk dapat menyelesaikan penelitian. Bab ini terdiri dari bab landasan teori dan penelitian terdahulu.

Bab III Metodologi Penelitian

Bab metodologi penelitian adalah bab yang berisi dengan data dan metode yang digunakan untuk penelitian. Bab ini terdiri dari beberapa sub bab

yaitu metode pengumpulan data, kerangka pikir, hipotesis, tahapan penelitian, dan metode penyelesaian.

Bab IV Hasil dan Pembahasan

Bab hasil dan pembahasan adalah bab yang berisi dengan hasil penelitian yang telah dilakukan oleh penulis dan membahas tentang hasil tersebut. Bab ini terdiri dari sub bab yang merupakan hasil pengumpulan data, hasil penelitian, dan pembahasan.

Bab V Kesimpulan dan Saran

Bab ini berisi dengan kesimpulan penulis tentang hasil penelitian yang telah dilakukan oleh penulis. Bab ini terdiri dari sub bab yang merupakan kesimpulan dan saran.