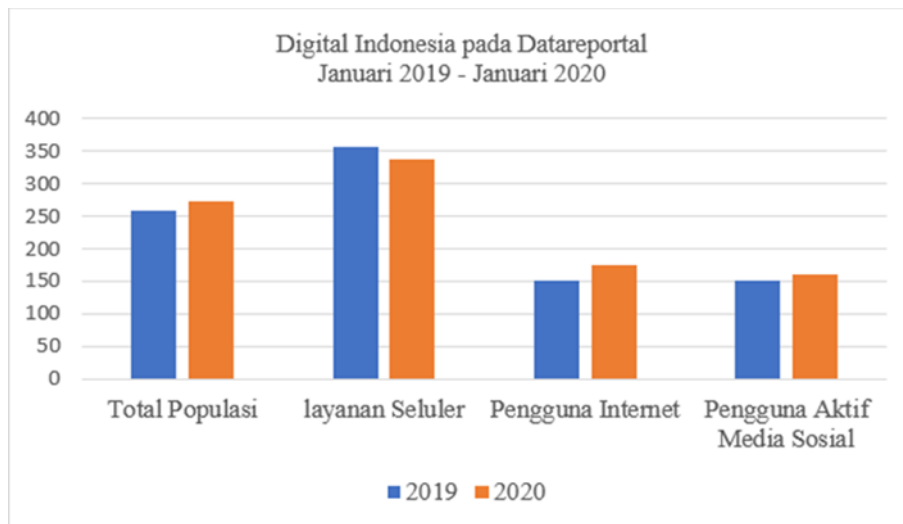


# BAB I PENDAHULUAN

## 1.1 Latar Belakang

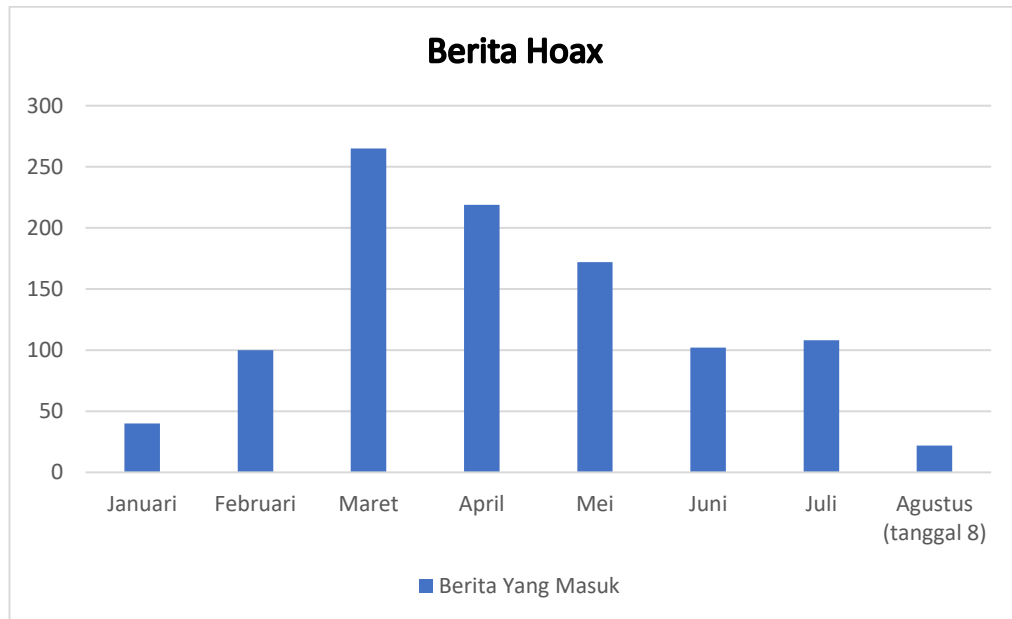
Berdasarkan Kamus Besar Bahasa Indonesia (KBBI) 2021, dijelaskan bahwa informasi tidak benar atau palsu yang didapat tanpa sumber yang jelas merupakan pengertian dari hoaks (Kemdikbud, 2016). Sedangkan dalam Oxford English Dictionary (Oxford University, 2022), hoaks didefinisikan sebagai suatu tindakan atau perbuatan yang dibuat dengan tujuan untuk membuat orang lain mempercayai sesuatu yang tidak benar. Berdasarkan data dari situs datareportal, pengguna internet dari Januari tahun 2019 (Kemp, 2019) sampai Januari 2020 (Kemp, 2020) mengalami peningkatan sekitar 25 juta pengguna. Hal ini merupakan salah satu faktor yang mempengaruhi penyebaran informasi menjadi tidak terkendali, sehingga berbagai informasi dapat diakses oleh masyarakat melalui internet tanpa validasi, dimana informasi itu dapat berupa berita palsu atau hoax. Peningkatan jumlah pengguna internet dapat dilihat pada Gambar 1.1 berikut.



Gambar 1.1 Digital Indonesia

Kata berita sendiri menurut Kamus Besar Bahasa Indonesia (KBBI), 2021 merupakan suatu informasi yang menunjukkan tempat atau keterangan waktu dan biasanya terjadi di masyarakat (Kemdikbud, 2016). Seperti telah disinggung sebelumnya, penyebaran berita dapat berupa informasi palsu atau kebohongan yang akan sulit dikendalikan apalagi dengan kondisi yang belum dapat dipastikan. Hoaks yang menyangkut kesehatan memang marak beredar di

kalangan masyarakat. Hasil survei yang dilakukan oleh Surveyor Persatuan Wartawan Indonesia (PWI) menemukan bahwa hoaks kesehatan terbanyak beredar di masyarakat (Juditha, 2019). Dapat diketahui bahwa berita hoax kesehatan adalah hoax yang populer jika dibandingkan dengan berita lainnya. Berdasarkan data yang dirilis oleh Kominfo melalui website resminya telah mencatat dari bulan Januari – Agustus 2020 bahwa sebanyak 1028 hoaks tersebar terkait berita Covid-19 (Kominfo, 2020). Dalam hal ini, berita hoaks covid merupakan salah satu berita hoaks kesehatan yang paling banyak penyebarannya seperti ditampilkan pada table berikut.



Gambar 1.2 Data Hoax terkait Covid-19

Hoaks kesehatan sangat berbahaya apabila informasi yang tidak benar dipraktikkan oleh masyarakat, seperti mengonsumsi jenis obat-obat tertentu akan membahayakan jiwa manusia. Hoaks jenis ini juga dapat menciptakan keresahan serta kepanikan bagi masyarakat yang membacanya (Juditha, 2019). Oleh karena itu cara untuk tidak terjerumus pada berita hoax adalah dengan membaca berita atau informasi dari media terpercaya. Di Indonesia sendiri menurut Ketua Komisi Penelitian, Pendataan dan Ratifikasi Dewan Pers Indonesia, Ratna Komala memaparkan bahwa terdapat sekitar 43 ribu portal media berita online yang terdata. Namun, dari jumlah tersebut hanya 500 yang resmi terdaftar oleh Dewan Pers. Kemudian dari 500 media online tersebut sudah tercatat ada 78 media yang terverifikasi faktual dan administrasi serta hanya ada tujuh media online yang lolos semua verifikasi tersebut, salah satunya adalah Detik.com (Dewan Pers, 2019). Berita mengenai hoax dapat diakses melalui situs Turnbackhoax. Turnbackhoax merupakan situs pengaduan berbagai jenis hoax. Situs ini dibuat oleh KOMINFO & MASTEL (Aji and Sarmini, 2019).

Text mining merupakan proses untuk memperoleh informasi berkualitas tinggi dari teks. Diantara proses yang dapat dilakukan dalam text mining adalah klasifikasi teks. Klasifikasi teks merupakan proses untuk menentukan suatu dokumen teks ke dalam suatu kelas tertentu (Deolika, Kusriani and Luthfi, 2019). Dalam prosesnya, klasifikasi dapat dilakukan dengan banyak cara baik secara manual ataupun dengan bantuan teknologi (Wibawa *et al.*, 2018). Klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma, diantaranya Decision Tree, K-Nearest Neighbor (KNN), Naive Bayes, Neural Network (NN), Random Forest (RF) dan Support Vektor Machine (SVM).

Penulis akan berfokus pada metode Support Vector Machine (SVM). SVM memiliki kelebihan diantara metode yang lainnya, dimana akurasi SVM lebih unggul dengan nilai akurasi pada proses k-fold validation sebesar 88,76% dibandingkan dengan Algoritma KNN dengan akurasi 88,1% (Maulina and Sagara, 2018). Kemudian perbandingan SVM dengan Multinomial Naive Bayes (MNB) mendapatkan hasil SVM lebih unggul dengan akurasi sebesar 82%, presisi 84% dan recall 82%. Sedangkan MNB memperoleh akurasi sebesar 62%, presisi 68% dan recall 62% (Nasution and Hayaty, 2019). Untuk memaksimalkan penggunaan metode SVM, penulis akan menggunakan hyperparameter tuning serta teknik optimasi Grid Search agar akurasi prediksi dapat lebih akurat. Hyperparameter tuning merupakan arsitektur dari deep learning untuk meningkatkan performa dari model prediksi (Hutter, Kotthoff and Vanschoren, 2019). Metode yang paling banyak digunakan dalam optimasi hyperparameter yaitu Grid Search (Nugraha and Sasongko, 2022). Grid search dianggap tidak efisien bahkan tidak layak untuk digunakan (Moreno *et al.*, 2018). Namun metode Grid Search masih menjadi solusi umum daripada harus mengubah satu hyperparameter pada satu waktu dan mengukur pengaruhnya terhadap kinerja model, justru hal tersebut akan menjadi tidak efisien dan tidak menjamin hasil yang optimal karena mengabaikan interaksi antara hyperparameter (Moreno *et al.*, 2018). Berdasarkan latar belakang tersebut, penelitian ini perlu mengembangkan dan menganalisis model klasifikasi berita hoax covid-19 untuk mendapatkan data prediksi antara berita faktual dan berita hoax tentang covid-19. Oleh karena itu penulis mengangkat judul “Optimasi Parameter SVM Menggunakan Grid Search Untuk Klasifikasi Berita Hoax Covid-19”. Agar dapat mengetahui apakah akurasi dari Metode SVM yang dioptimasi oleh grid search lebih baik dari metode SVM yang hanya menggunakan kernel linear. Metode SVM dengan kernel linear dipilih karena akurasi kernel linear dianggap cukup tinggi dalam kasus ini dibandingkan kernel lainnya (Ropikoh *et al.*, 2021).

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan diatas, maka rumusan masalah pada penelitian ini adalah :

1. Bagaimana menerapkan metode Support Vector Machine yang dioptimasi oleh Grid Search untuk klasifikasi berita hoax?
2. Bagaimana tingkat akurasi metode Support Vector Machine yang dioptimasi oleh Grid Search terhadap metode Support Vector Machine kernel linear?

### **1.3 Batasan Masalah**

Pembatasan masalah sangat diperlukan untuk memfokuskan suatu penelitian agar hasil yang didapat lebih maksimal. Batasan masalah pada penelitian ini diantaranya :

1. Klasifikasi berita dilakukan pada portal berita Detik dan Turnbackhoax.
2. Metode yang digunakan dalam penelitian ini adalah Support Vector Machine yang dioptimasi oleh Grid Search dengan metode pembandingan adalah metode Support Vector Machine dengan kernel linear.
3. Data yang digunakan adalah data yang diambil dari portal berita Detik.com dan Turnbackhoax, sejak 1 januari 2021 sampai 30 juni 2021.

### **1.4 Tujuan Penelitian**

Dalam sebuah penelitian ilmiah, tentu diperlukan tujuan penelitian yang jelas. Oleh karena itu tujuan dari penelitian ini diantaranya :

1. Menerapkan metode Support Vector Machine yang dioptimasi oleh Grid Search pada proses klasifikasi.
2. Menganalisa tingkat akurasi, presisi, recall, dan FI Score dari Support Vector Machine yang dioptimasi oleh Grid Search dan dibandingkan dengan metode Support Vector Machine dengan kernel linear.
3. Untuk menganalisa tingkat penyebaran berita hoax covid-19 pada situs berita online.

### **1.5 Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian ini, yaitu :

Penggunaan metode Support Vector Machine yang dioptimasi oleh Grid Search diharapkan dapat menjadi rujukan untuk pengembangan metode klasifikasi selanjutnya.