

BAB I

PENDAHULUAN

I.1. Latar Belakang

Menurut Kamus Besar Bahasa Indonesia (KBBI), kata serapan adalah kata yang terdapat dalam bahasa Indonesia, namun merupakan hasil serapan dari bahasa lain. Kata serapan dapat berasal dari bahasa daerah, seperti Jawa, Sunda, Batak, dan lain-lain, atau bahasa asing, seperti Arab, Belanda, Inggris, Portugis, dan lain-lain [1]. Pendapat lain menyebutkan kata serapan adalah kata yang berasal dari bahasa asing yang sudah diintegrasikan ke dalam suatu bahasa agar diterima pemakaiannya secara umum [2]. Contohnya adalah kata “Arsitektur” yang berasal dari Yunani dengan asal kata “*arkhitekton*”, lalu kata “Televisi” serapan “*television*” dari Inggris [3]. Kata serapan memiliki banyak unsur, berdasarkan artikel kemendikbud tercatat ada 95 penulisan unsur serapan umum ditambah ada 4 penulisan serapan khusus [4].

Ketersediaan referensi mengenai kata serapan masih terbatas. Salah satu karya yang signifikan adalah kamus kata serapan dalam bahasa Indonesia yang disusun oleh J.S. Badudu, buku senarai kata serapan bahasa Indonesia, serta Kamus Besar Bahasa Indonesia (KBBI) [5][6][7]. Namun, keberadaan referensi semacam ini semakin jarang, sehingga menimbulkan kebutuhan untuk mempelajari kata serapan secara lebih interaktif dan modern.

Indonesia juga menghadapi tantangan besar dalam meningkatkan tingkat literasi masyarakat. Berdasarkan survei *Program for International Student Assessment (PISA)* yang dirilis oleh *Organization for Economic Co-operation and Development (OECD)* pada 2019, Indonesia menempati peringkat ke-62 dari 70 negara terkait tingkat literasi [8]. Pembelajaran kosakata, termasuk kata serapan, merupakan bagian penting dari upaya peningkatan literasi. Kosakata yang kaya tidak hanya membantu individu memahami berbagai konteks bahasa tetapi juga meningkatkan kemampuan membaca dan menulis [9].

Dalam era teknologi digital, integrasi teknologi memberikan peluang untuk meningkatkan literasi bahasa secara inovatif [10]. Internet saat ini sudah

menjadi kebutuhan primer masyarakat [11], dengan hampir semua kalangan di Indonesia, dari anak-anak hingga dewasa, memiliki akses ke *smartphone*. Hal ini didukung oleh fakta bahwa pada tahun 2023 terdapat 209,3 juta pengguna aktif *smartphone* di Indonesia, yang mencakup sekitar tiga perempat dari populasi [12]. Dengan data ini, muncul peluang untuk mengembangkan aplikasi *android* yang menggabungkan fitur pengenalan kata serapan. Aplikasi ini memiliki potensi besar untuk diadopsi oleh masyarakat, sekaligus menjadi alat edukasi yang efektif dan menarik.

Penelitian serupa sebelumnya, seperti yang dilakukan oleh Fahira Khalisyah Risqullah [13], menghasilkan aplikasi kamus kata serapan yang mempunyai fitur pengenalan suara untuk mencari kata. Dalam aplikasi tersebut, pengguna dapat mencari kata serapan berdasarkan kata yang diucapkan [13]. Namun, aplikasi tersebut terbatas pada fungsi pencarian kata, tanpa adanya fitur prediksi berbasis model *machine learning*. Hal ini menunjukkan adanya peluang untuk mengembangkan aplikasi yang lebih inovatif, yang tidak hanya mencari kata serapan tetapi juga mampu mengenali dan memprediksi kata serapan dari input audio berupa kata atau kalimat percakapan pengguna.

Untuk mendukung pembuatan aplikasi tersebut, diperlukan teknologi yang mampu menangani klasifikasi kata secara akurat. *Support Vector Machine (SVM)* merupakan salah satu algoritma *machine learning* yang terbukti efektif dalam berbagai penelitian terkait klasifikasi teks dan audio. Dalam penelitian oleh Zhang et al., *SVM* menunjukkan kinerja terbaik dibandingkan algoritma lain seperti *Naïve Bayes* dan *K-Nearest Neighbors (KNN)* untuk klasifikasi teks [14]. Selain itu, dalam studi oleh Asgarnezhad et al., *SVM* mencapai akurasi 92,85% dalam klasifikasi sentimen teks, mengungguli *Naïve Bayes* [15]. Penelitian lainnya oleh Kumar et al. menunjukkan bahwa dengan bantuan ekstraksi fitur *MFCC*, *SVM* dan *CNN* efektif dalam mendeteksi audio stereo palsu, dengan akurasi pengenalan mencapai 99% dan tingkat *false acceptance rate* yang minimal, yaitu 0% [16].

Lebih lanjut, Ch A. S. Murty dan Parag H. Rughani dalam penelitian mereka mengenai klasifikasi teks pada *Dark Web* membuktikan efektivitas *SVM* dengan kernel *Gaussian* dan *Radial Basis Function (RBF)*. Model tersebut mencapai *Precision* sebesar 83%, *Recall* 90%, dan *F1-Score* 96%, menegaskan

potensi *SVM* dalam menangani berbagai jenis data teks dengan hasil yang sangat akurat [17].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengembangkan aplikasi *android* yang mampu mengenali kata serapan secara akurat dari hasil percakapan. Aplikasi ini mengintegrasikan algoritma *Support Vector Machine (SVM)* sebagai model klasifikasi kata serapan dan memanfaatkan teknologi *kotlin* untuk membuat aplikasi mobile. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi performa *SVM* dalam mengklasifikasikan kata serapan menggunakan *dataset* teks dan audio. Penggunaan dua jenis *dataset* bertujuan untuk menganalisis model mana yang terbaik untuk diimplementasikan ke dalam aplikasi *mobile* secara optimal. Dengan pendekatan ini, penelitian diharapkan dapat menjawab tantangan keterbatasan referensi kata serapan secara inovatif, sekaligus berkontribusi pada peningkatan literasi masyarakat melalui aplikasi berbasis teknologi. Selain itu, aplikasi ini juga diharapkan dapat menarik minat masyarakat untuk menggunakannya, karena menawarkan pendekatan yang berbeda dibandingkan kamus konvensional ataupun aplikasi kamus biasa. Pendekatan ini tidak hanya menghasilkan wawasan baru tentang kata serapan, tetapi juga memperkaya pengetahuan pengguna secara interaktif dan modern.

1.1. Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka dapat dirumuskan beberapa permasalahan yang muncul diantaranya yaitu:

1. Bagaimana membuat aplikasi pengenalan kata serapan berbasis Android yang terintegrasi dengan algoritma *Support Vector Machine (SVM)* dari hasil percakapan?
2. Bagaimana kinerja algoritma *Support Vector Machine (SVM)* dalam mengidentifikasi kata serapan ditinjau dari aspek akurasi, presisi, dan efisiensi?

1.2. Batasan Masalah

Ruang lingkup masalah pada penelitian yang akan dilakukan cukup luas, sehingga dilakukan pembatasan pada masalah yang akan dibahas untuk