

## 1. Pendahuluan

### Latar Belakang

Dalam era teknologi saat ini, faktor ekonomi, di mana aktivitas seperti pembelian, penjualan, transaksi, dan investasi semakin banyak yang dilakukan secara *online* dan merupakan peluang yang menjanjikan [1], [2]. Investasi adalah pengalokasian dana untuk membeli instrumen keuangan atau aset dengan tujuan menghasilkan keuntungan berupa dividen atau bunga. [3].

Investasi populer di Indonesia menurut HSBC meliputi deposito, emas, properti, saham, reksa dana, dan pinjaman antar individu (*peer to peer lending*) [4]. Survei HSBC *Media Advisory* 2017 menunjukkan bahwa 39% generasi milenial di Indonesia sangat tertarik mengambil risiko dalam investasi [3]. Pada triwulan IV tahun 2020, investasi yang terealisasi mencapai 214,7 triliun rupiah, meningkat 3,1% dibandingkan tahun sebelumnya, mencerminkan minat masyarakat yang meningkat dalam berinvestasi [4]. Juga didukung oleh munculnya berbagai aplikasi investasi online. Namun kelebihan dan kekurangan menjadi tolak ukur kualitas layanan aplikasi. Dengan menganalisis sentimen dapat mengetahui pandangan pelanggan dan mengidentifikasi potensi masalah atau area perbaikan [1], [5].

Penelitian terkait analisis sentimen menggunakan SVM telah dilakukan oleh beberapa peneliti [6], [7], [8], [9]. *Support Vector Machine* (SVM) adalah algoritma yang banyak digunakan yang dapat secara akurat mengklasifikasikan opini ke dalam kategori sentimen positif, dan negatif [10], [7], [11]. SVM mengklasifikasikan dengan membuat *hyperplane* yang membedakan antara opini positif dan negatif [10]. Untuk mempermudah dalam menangani masalah dimensionalitas, SVM mengubah informasi menjadi aspek yang lebih tinggi dengan menggunakan operasi kernel seperti *Linear*, *Radial Basis Function* (RBF), *Polynomial*, dan *Sigmoid* [12].

Penelitian [6] menganalisa sentimen ulasan aplikasi ruangguru menggunakan klasifikasi SVM, pengujian menggunakan kernel *linear* menghasilkan akurasi 89,7%. Kombinasi data *train* dan *test* menghasilkan akurasi 90%. Dengan *K-Fold*, akurasi tertinggi pada *K-Fold* 6, 9, dan 10 yaitu 90,2%. Tingkat akurasi dalam penelitian ini tinggi berada di kisaran 90%. Sedangkan pada penelitian [7] menganalisa sentimen ulasan aplikasi transportasi online menggunakan klasifikasi SVM dengan kernel RBF dan ekstraksi fitur *Word2Vec* dengan model *skip-gram*, *size*-nya 100 dimensi dan *window*-nya adalah 5. Hasil pengujian menunjukkan performa yang cukup baik yaitu untuk aplikasi Gojek akurasi sebesar 87%. Sedangkan aplikasi Grab akurasi sebesar 82%. Pada penelitian [8] klasifikasi sentimen pada ulasan aplikasi provider by.U menggunakan SVM dan TF-IDF sebagai ekstraksi fitur. TF-IDF+SVM dengan *5-Fold Validation* menghasilkan akurasi cukup baik sebesar 84,7%, presisi sebesar 84,9%, *recall* sebesar 84,7%, dan *f-measure* sebesar 84,8%. Hasil akurasi tertinggi pada *fold* 2 yaitu 86,1%. Pada penelitian [9] menganalisis sentimen pada review film menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) dan *Support Vector Machine* (SVM) sebagai algoritma klasifikasi. Pemilihan metode ini disebabkan kemampuannya dalam memberikan bobot pada kata-kata dan mengklasifikasikan data dengan dimensi tinggi. Penggunaan data *training* yang lebih banyak akan berdampak pada kinerja sistem. Berdasarkan uji skenario, terungkap bahwa kombinasi algoritma TF-IDF dan SVM efektif untuk kasus *review* film, dengan akurasi sebesar 85%, presisi 100%, *recall* 70%, dan nilai *F1-Score* mencapai 82%.

Berdasarkan hasil riset tersebut, efektifitas fitur ekstraksi diperlukan dalam klasifikasi sentimen karena teknik ekstraksi fitur yang baik dapat menangkap informasi penting dari teks dan mengubahnya menjadi representasi numerik yang berguna bagi algoritma klasifikasi. Dalam penelitian-penelitian yang telah disebutkan, TF-IDF dan *Word2Vec* telah menunjukkan kemampuan untuk meningkatkan akurasi model klasifikasi sentimen. TF-IDF memberikan bobot berdimensi tinggi pada kata-kata berdasarkan frekuensinya dalam dokumen, sehingga kata-kata yang lebih relevan mendapatkan bobot yang lebih tinggi. Di sisi lain, *Word2Vec* mengubah kata-kata menjadi vektor numerik yang menangkap makna semantik dari kata-kata tersebut, memungkinkan model untuk memahami hubungan kontekstual antara kata-kata dalam kalimat. Penelitian ini akan mengevaluasi efektifitas fitur ekstraksi *Word2Vec* *Word Embedding* dan TF-IDF serta fitur kombinasi yaitu *Word2Vec* berbobot TF-IDF dengan mengalikan vektor *Word2Vec* terhadap nilai TF-IDF untuk menghasilkan representasi vektor yang lebih kaya. Algoritma klasifikasi yang digunakan yaitu SVM dengan optimasi parameter untuk menemukan parameter terbaik, hal ini menunjukkan pendekatan yang lebih cermat dalam memilih parameter yang optimal untuk meningkatkan kinerja model dalam klasifikasi sentimen ulasan aplikasi investasi online Nanovest di *Google Play Store*. Evaluasi kinerja diperlukan agar dapat menemukan model dan hasil yang paling efektif dan optimal yang diharapkan dapat memberikan kontribusi dalam bidang analisis sentimen, khususnya dalam mengetahui tingkat kepuasan pengguna aplikasi Nanovest berdasarkan klasifikasi sentimen ulasan pengguna aplikasi tersebut.

### Topik dan Batasannya

Topik pada penelitian ini yaitu berfokus pada evaluasi efektifitas ekstraksi fitur *Word2Vec* dan TF-IDF serta kombinasi fitur TF-IDF *Weighted Word2Vec* dalam klasifikasi sentimen ulasan aplikasi investasi online

Nanovest di *Google Play store* menggunakan algoritma klasifikasi *Support Vector Machine*. Batasan pada penelitian ini yaitu menguji ekstraksi fitur yang paling efektif antara *Word2Vec* dan TF-IDF juga kombinasi fitur dalam klasifikasi sentimen menggunakan *Support Vector Machine* dengan dataset ulasan aplikasi investasi online Nanovest di *Google Play Store* yang berjumlah sekitar 10.000 data.

### **Tujuan**

Tujuan dari penelitian ini yaitu untuk mengevaluasi efektifitas ekstraksi fitur *Word2Vec* dan TF-IDF juga kombinasi fitur TF-IDF *Weighted Word2Vec* dalam klasifikasi sentimen ulasan aplikasi investasi online Nanovest di *Google Play Store* menggunakan algoritma klasifikasi *Support Vector Machine* dengan tujuan menemukan ekstraksi fitur yang paling efektif sehingga memberikan kontribusi dalam bidang analisis sentimen, khususnya dalam mengetahui tingkat kepuasan pengguna aplikasi Nanovest berdasarkan klasifikasi sentimen ulasan pengguna aplikasi tersebut.

### **Organisasi Tulisan**

Dalam penelitian ini, bagian 2 menjelaskan definisi serta arsitektur dari algoritma yang diterapkan, bagian 3 menjelaskan langkah-langkah yang akan dilakukan selama penelitian, bagian 4 membahas temuan dari penelitian, dan bagian 5 menyajikan kesimpulan berdasarkan hasil yang diperoleh.