1. Pendahuluan

1.1 Latar Belakang

Penggunaan media sosial saat ini dapat terbilang sangat tinggi, seluruh orang dapat menggunakannya secara bebas sesuai dengan keinginannya. Dengan demikian, banyak orang yang menggunakan media sosial untuk mengungkapkan pendapat, komentar, kritik, dan lain-lain. Seluruh pendapat atau komentar yang diberikan dapat dijadikan sebagai informasi penting untuk pengambilan keputusan bagi suatu organisasi atau orang. Banyaknya jumlah data yang ada tidak memungkinkan bagi manusia untuk membaca dan menganalisis secara manual sehingga dibutuhkan Analisis Sentimen. Analisis Sentimen atau yang disebut juga dengan *opinion mining* merupakan proses pengambilan informasi dengan menganalisis sekumpulan teks yang dapat diklasifikasikan berdasarkan sikap, emosi dan penilaian terhadap suatu entitas seperti produk ataupun perusahaan [1].

Telkomsel adalah operator telekomunikasi seluler terbesar di Indonesia yang sering menjadi perbincangan di media sosial terutama Twitter. Sebagai perusahaan penyedia layanan atau produk tentu analisis sentimen diperlukan sebagai bahan evaluasi dalam meningkatkan kualitas produk. Namun dengan itu saja tidak cukup, perusahaan perlu mengetahui aspek atau hal apa yang dikomentari oleh publik, sehingga perlu dilakukan pendekatan analisis sentimen berbasis aspek. Penelitian serupa mengenai analisis sentimen dilakukan oleh Joshua Acosta et al [2] tentang opini pengguna mengenai penerbangan Amerika Serikat. Metode klasifikasi yang digunakan adalah Logistic Regression dengan ekspansi fitur Word2Vec yang menghasilkan akurasi sebesar 72,00%.

1.2 Topik dan Batasannya

Sejauh ini diketahui bahwa belum banyak penelitian analisis sentimen berbasis aspek yang mencoba menggunakan ekspansi fitur. Dengan demikian pada penelitian ini dilakukan analisis sentimen berbasis aspek pada pengguna Telkomsel di Twitter dengan menggunakan data berupa tweet yang lebih banyak. Pada penelitian ini TF-IDF sebagai ekstraksi fitur dan *word embedding* Word2Vec digunakan sebagai ekspansi fitur untuk merepresentasikan sebuah kata ke dalam bentuk vektor untuk mendapatkan hubungan semantik antar kata-kata [3]. Kegunaan ekspansi fitur tersebut yaitu untuk meminimalisir ketidakcocokan kata yang diakibatkan oleh keterbatasan kata dan penyingkatan kata pada tweet. Dalam mengatasi ketidakseimbangan data yang ada maka SMOTE (Synthetic Minority Oversampling Technique) digunakan sebagai salah satu teknik dalam oversampling. Metode klasifikasi yang digunakan adalah *Logistic Regression* dikarenakan metode tersebut dapat memberikan model klasifikasi yang cukup baik dibandingkan dengan metode lain [2], [4] serta dapat digunakan untuk menghasilkan keluaran lebih dari dua kelas sehingga pada penelitian ini dapat memberikan hasil akurasi yang lebih tinggi dibandingkan penelitian yang sebelumnya. Selain itu untuk mengoptimasi proses klasifikasi maka *logistic regression* dikombinasikan dengan salah satu algoritma *boosting* yaitu *Gradient Boosting*.

Pada penelitian ini terdapat beberapa masalah yang dibahas yaitu bagaimana ekstraksi fitur TF-IDF yang dikombinasikan dengan ekspansi fitur Word2vec dengan algoritma klasifikasi *Logistic Regression* bekerja dalam analisis sentimen berbasis aspek dan juga bagaimana pengaruh beserta tingkat performansi yang dihasilkan. Adapun beberapa masalah yang menjadi batasan pada pengerjaan penelitian ini yaitu data yang digunakan diperoleh dari Twitter dalam bentuk teks cuitan/*tweet* berbahasa Indonesia yang membahas mengenai layanan dan sinyal Telkomsel. Hal ini dikarenakan Twitter menjadi media yang paling banyak digunakan orang untuk menyampaikan opini nya. Algoritma yang digunakan untuk klasifikasi hanya menggunakan *Logistic Regression* dan tidak melakukan pembandingan dengan algoritma lain nya dan ekspansi fitur *word embedding* yang digunakan adalah Word2Vec. Korpus yang digunakan pada proses *feature expansion* yaitu hanya *tweet*, IndoNews, dan IndoNews+*tweet*.

1.3 Tujuan

Tujuan pada penelitian ini yaitu untuk mengetahui seberapa besar pengaruh dan tingkat akurasi yang dihasilkan dari penerapan ekspansi fitur *word embedding* menggunakan Word2Vec dan TF-IDF sebagai ekstraksi fiturnya dengan metode klasifikasi *Logistic Regression* yang dikombinasikan dengan algoritma *boosting*.

1.4 Organisasi Tulisan

Tulisan ini disusun dengan keterangan sebagai berikut, pada bab 2 terdapat pembahasan mengenai studi literatur yang berkaitan atau mendukung penelitian ini. Bab 3 berisi teori yang terkait dengan penelitian dan membahas mengenai model sistem yang dibangun. Selanjutnya pada Bab 4 membahas tentang hasil, analisis, dan evaluasi dari penelitian. Pembahasan mengenai kesimpulan dan saran untuk penelitian selanjutnya tertulis pada bab 5.