1. Pendahuluan

1.1 Latar Belakang

Media sosial banyak digunakan seseorang untuk mengekspresikan emosi serta membagikan cerita sehari-hari mereka. Di Indonesia, salah satu media sosial dengan jumlah pengguna yang besar adalah Twitter. Sampai bulan Juli tahun 2020, pengguna Twitter yang berasal dari Indonesia ada sebanyak 11,2 Juta orang[1]. Pada Twitter, setiap unggahan pengguna disebut cuitan. Twitter merupakan tempat untuk mencurahkan isi hati pada cuitan dengan panjang karakter sepanjang 280 karakter [1].

Selain kesehatan secara fisik, kesehatan mental adalah hal yang sangat penting bagi seseorang. Bagi banyak orang, perasaan kehilangan kualitas hidup yang baik adalah suatu permasalahan yang serius. Segala hal yang menyebabkan perasaan tidak senang dan kesedihan dalam jangka waktu yang panjang dapat menurunkan rasa percaya diri, semangat, dan minat untuk menjalani kehidupan pada seseorang [2]. Depresi adalah suatu penyakit yang dapat menurunkan tingkat kualitas hidup seseorang. Prediksi lebih awal dapat memberikan pelayanan kesehatan yang tepat bagi seseorang yang menderita depresi [3]. Aktivitas sosial yang dilakukan di Twitter dapat digunakan untuk mendeteksi tanda-tanda depresi tersebut [4].

Beberapa penelitian telah berhasil memprediksi dan mengklasifikasikan tingkat depresi pada seseorang melalui media sosial Twitter. Pada penelitian berbahasa Inggris yang dilakukan oleh Choudhury et al. [5] dengan menggunakan metode *Support Vector Machine* dan fitur Linguistic Inquiry Word Count (LIWC) mampu mengklasifikasikan depresi dengan akurasi 70%.LIWC mengkuantifikasi kategori kata yang berbeda dan menentukan derajat emosi positif atau negatif, referensi diri, kata-kata kausal, dan banyak dimensi bahasa lainnya [2]. Guntuku et al. [6] menggunakan metode *Random Forest* dengan penggunaan fitur LIWC dan topik menghasilkan akurasi sebesar 78,3%. Tsugawa et al. [4] menggunakan metode *Support Vector Machine* dan fitur sosial yang menghasilkan akurasi 66%.

Nadeem et al [7] membandingkan metode *Decision Tree, Naïve Bayes*, *Logistic Regression*, dan *Support Vector Machine.* Seluruh metode menghasilkan nilai akurasi sebesar 82%, kecuali *Decision Tree* yang hanya menghasilkan nilai akurasi sebesar 67%. Kale et al. [8] membandingkan metode *Support Vector Machine, Naïve Bayes, KNN,* dan *Logistic Regression*, didapatkan metode *Logistic Regression* memiliki nilai akurasi tertinggi yaitu 96,14%. Menurut Kale et al. [8] *Support Vector Machine* memiliki kelemahan yaitu sulit menentukan data bias, sedangkan metode *Naïve Bayes* memiliki kelemahan pada setiap fitur yang berjalan secara independen, dan metode *KNN* sulit untuk menentukan nilai K. Berdasarkan pada penelitian sebelumnya, metode *Logistic Regression* mencapai nilai akurasi tertinggi karena memiliki kelebihan yaitu mampu membedakan kelas dengan jelas [8]. Oleh karena itu, pada penelitian ini menggunakan metode *Logistic Regression* untuk membangun sistem prediksi tingkat depresi pada seseorang dari media sosial Twitter dengan data berbahasa Indonesia. Selain itu pada penelitian ini juga akan melakukan pengkombinasian fitur sosial, linguistik, dan ortografik.

1.2 Topik dan Batasannya

Berdasarkan latar belakang, penelitian ini akan membangun sistem prediksi tingkat depresi pada seseorang melalui media sosial Twitter dengan menggunakan data twitter berupa cuitan dan informasi profil. Cuitan-cuitan yang didapatkan digunakan sebagai masukan untuk fitur linguistik dan ortografik. Sedangkan data informasi profil digunakan sebagai masukan untuk fitur sosial. Sistem melakukan klasifikasi dengan menggunakan metode *Logistic Regression*. Adapun rumusan masalah dalam penelitian ini adalah bagaimana pengaruh kombinasi fitur sosial, linguistik, dan ortografik terhadap akurasi prediksi depresi pada seseorang dari media sosial Twitter menggunakan metode *Logistic Regression*?

Penelitian ini memiliki beberapa Batasan masalah diantara lain, keterbatasan sumber daya dimana adanya keterbatasan waktu untuk menyebarkan kuesioner. Penelitian juga dibatasi pada pengguna dengan akun bertipe publik. Pelabelan tingkat depresi pada pengguna dilakukan secara manual berdasarkan pada penilaian kuesioner DASS-42 sehingga memungkinkan terjadinya ketidakakuratan pada label tingkat depresi pada pengguna.

1.3 Tujuan

Tujuan dari penelitian ini adalah mengetahui akurasi dari sistem prediksi tingkat depresi pada seseorang melalui media sosial Twitter dengan kombinasi fitur sosial, linguistik, dan ortografik dengan menggunakan metode *Logistic Regression*.

[1] https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/