

1. Pendahuluan

Hoax adalah informasi palsu yang sering muncul diinternet dan memiliki tujuan untuk menutupi informasi yang sebenarnya dan juga menyebarkan kepanikan atau ketakutan massal, contohnya seperti dibidang ekonomi, politik, kesehatan, teknologi, hingga keamanan. Saat ini sudah ada ratusan hingga ribuan berita yang sudah dimanipulasi isi beritanya yang sudah dan akan menyebabkan kepanikan [1].

Masalah ini tidak dapat dipisahkan dari dampak penggunaan media sosial yang cepat. Akibatnya, setiap hari ada ribuan informasi yang tersebar dimedia sosial, yang belum tentu valid, sehingga orang – orang berpotensi terkena tipuan dimedia sosial. Informasi yang tidak valid (*hoax*) dapat mempengaruhi emosi, perasaan, pikiran, atau bahkan tindakan seseorang atau kelompok [1]. Sangat disayangkan jika informasi tersebut tidak akurat atau bahkan informasi palsu (*hoax*) dengan provokatif judul yang mengarahkan pembaca dan penerima ke opini negatif.

Media sosial adalah platform yang digunakan oleh semua orang untuk berbagi informasi. Orang beralih ke media sosial digunakan untuk tempat komunikasi. Media sosial telah meledak sebagai kategori wacana online di mana orang bisa membuat dan berbagi informasi [2]. Begitu pula dengan berita *hoax* yang juga berkembang dikalangan pengguna media sosial, contohnya twitter.

Twitter merupakan media sosial bertipe *microblogging* yang didirikan oleh Jack Dorsey pada Maret 2016 dan diluncurkan pada Juli 2006. Keunikan dari twitter adalah mempunyai tweet atau post yang ada ditwitter dengan ukuran maksimum 140 karakter. Pada Twitter juga dapat ditemui berbagai macam pesan positif hingga negatif, seperti *hoax*, gosip, pornografi, penipuan, pencemaran nama baik, bahkan *self-harming*. Pengguna dapat berinteraksi dengan teman diseluruh penjuru dunia melalui pesan singkat yang ditulis. Tidak sedikit dari pesan tersebut sengaja ditulis dengan tujuan untuk menyebarkan *hoax*. *Hoax* menjadi perbincangan panas ditwitter karena dianggap meresahkan publik dengan adanya informasi yang tidak bisa sepenuhnya dipercaya [2].

Menggunakan metode *Word2vec* karena untuk mengurangi ketidaksesuaian kosakata. Penelitian pertama mengenai *Word2Vec* dilakukan oleh Mikolov menghasilkan sebuah model dari representasi kata dan frasa [3]. Dalam sistem pendeteksian *hoax* digunakan cara pengolahan yang didalamnya juga memiliki beberapa tahapan untuk mengolah setiap kata. Crawling data yang mengambil data dari tweet di twitter, memisahkannya dan membandingkannya dengan kata-kata yang sudah ada sebelumnya. Crawling data otomatis mengambil dari API twitter yang sudah ada di aplikasi web yang di sediakan oleh twitter.

Oleh karena itu, pada penelitian ini penulis akan menambahkan ekspansi fitur *Word2vec* yang dikombinasikan dengan metode klasifikasi seperti *Naïve Bayes*, *ANN*, dan *Decision tree*. Sebelum melakukan ekspansi fitur, pre-processing data merupakan langkah penting yang harus dilakukan untuk mengurangi jumlah input kata, sehingga data secara efisien dapat digunakan oleh sistem ekspansi fitur. Untuk pembobotan kata, penulis menggunakan algoritma TF-IDF (Term Frequency – Inverse Document Frequency).

Tujuan dari penelitian ini adalah mengetahui pengaruh penerapan metode klasifikasi *Naïve-bayes*, *ANN*, dan *Decision Tree* dalam mengklasifikasikan topik pada twitter, dan pengaruh penerapan ekspansi fitur pada metode *Naïve-bayes*, *ANN*, dan *Decision Tree* dalam mengklasifikasikan topik pada twitter.