

## 1. Pendahuluan

Pemilihan Umum atau yang biasa disebut pemilu adalah proses memilih untuk mengisi suatu jabatan tertentu. jabatan tersebut beraneka ragam, salah satunya adalah jabatan sebagai presiden. Pemilihan presiden diadakan setiap lima tahun sekali, di mana baru saja dilangsungkan pada April 2019. Tak sedikit media massa maupun sosial media yang membicarakan mengenai pemilihan presiden RI 2019, di mana hal tersebut menimbulkan banyak sekali opini masyarakat terkait dengan paslon presiden RI

Untuk mendapatkan dukungan dari masyarakat, calon presiden harus melakukan kampanye. Pada politik di Indonesia, sosial media mempunyai peranan penting sebagai media kampanye, Hal ini terjadi karena media sosial dapat memudahkan penggunaannya mengakses berita politik terbaru serta opini masyarakat yang tidak dimuat dalam surat kabar nasional maupun televisi[1]. Dengan keberhasilan kampanye Obama dalam pemilihan presiden AS yang memanfaatkan media sosial, para kandidat pemilu di banyak negara mencoba memanfaatkan media sosial untuk kampanye[9]. Twitter adalah salah satu situs mikroblog yang paling ringkas dan tepat[2] karena twitter merupakan area umum di mana masyarakat dan politisi mengekspresikan pandangan mereka[4]. Pemanfaatan twitter untuk kampanye dapat digunakan untuk mendapatkan informasi dan prediksi siapa yang akan memenangkan pemilu[1].

Pada pra-pelaksanaan maupun pelaksanaan pemilu presiden RI terdapat banyak sekali opini dan tanggapan dengan sentimen positif dan negatif bermunculan dari masyarakat Indonesia di twitter. Perlu dilakukan adanya pengelompokan opini berupa teks sentimen untuk memprediksi kekuatan capres lebih dini dengan cepat dan akurat. Analisis sentimen adalah proses mengenali dan mengklasifikasikan sentimen berbeda yang disampaikan secara *online* oleh masyarakat untuk mendapatkan pendekatan penulis terhadap produk, topik, atau acara tertentu yang positif, negatif atau netral[5]. Selain itu, analisis sentimen di media sosial memungkinkan untuk memantau kampanye pemilu hari demi hari[6]. Dalam penelitian[18] menunjukkan bahwa deteksi sentimen dapat mengurangi kesalahan hasil prediksi.

Pada periode sebelumnya, prediksi hasil pilpres dilakukan oleh lembaga survei dan *quick count*. Namun, hasil survei kadang tidak sesuai dengan kenyataan padahal dengan adanya survei pra-pemilu dapat membuat capres dan cawapres mengetahui parameter kemenangan dan kekalahan lebih dini sehingga bisa memperbaiki strategi kampanye[1]. Beberapa lembaga survey menggunakan *multistage random sampling* untuk mengumpulkan data. Prediksi terbaik yang dikeluarkan oleh lembaga survey dicapai oleh SSSG (Sekolah Pemerintah Soegeng Sarjadi) dengan (*Mean Absolute Error*) MAE sebesar 0.9% sedangkan prediksi twitter mempunyai MAE sebesar 0.6%.

Penelitian mengenai analisis sentimen terhadap tokoh masyarakat telah dilakukan sebelumnya seperti yang telah dilakukan oleh Nurhayatin dkk[8] di mana penelitian tersebut hanya menyimpulkan bahwa setiap tokoh masyarakat mempunyai sentimen positif dan negatif tersendiri, namun belum diadakannya prediksi untuk tokoh masyarakat yang terbaik. Selain itu, para peneliti banyak menggunakan pendekatan berbeda, salah satunya adalah dengan menggunakan data yang dipilih hanya beberapa hari atau minggu sebelum pemilihan[9]. Berbagai penelitian seperti yang menggunakan data enam bulan sebelum pemilihan dan ada juga yang satu bulan sebelum penelitian, hasilnya menggunakan data yang lebih banyak bukan berarti menghasilkan akurasi yang tinggi[9]. Selain waktu pengambilan data, kata kunci yang digunakan dalam pengambilan data juga perlu diperhatikan, hasil prediksi juga dapat ditingkatkan jika menggunakan lebih banyak kata kunci daripada hanya menggunakan nama kandidat atau pihak[9]

Berdasarkan permasalahan tersebut, maka pada penelitian ini dilakukan prediksi calon presiden Republik Indonesia pada pemilu presiden 2019. Data yang digunakan merupakan data twitter pada bulan April 2019. Hal tersebut dilakukan atas dasar penelitian[9]. Data *tweet* yang diambil merupakan data berbahasa Indonesia yang diambil dari *trending topics* di Twitter terkait dengan pemilihan presiden[10], di mana dalam *trending topics* banyak sekali *hashtag* yang akan dimanfaatkan dalam penelitian ini sebagai salah satu fitur. Data yang telah didapatkan akan diberikan label positif dan negatif. Maka dari itu, penelitian dilakukan dengan menggunakan *Supervised Learning* di mana data akan belajar dari data latih yang telah diketahui labelnya. Metode yang digunakan dalam penelitian ini adalah SVM (*Support Vector Machine*) dengan memanfaatkan pembobotan TF-IDF.

Berbagai metode untuk penelitian serupa telah dilakukan[10, 11], dan dari beberapa penelitian tersebut yang paling sering digunakan adalah SVM. Pada penelitian[10] yang membandingkan antara metode NBC (*Naive Bayes Classifier*) dan *Support Vector Machine* (SVM). Dari hasil penelitian tersebut, metode SVM menunjukkan hasil yang lebih baik dari dengan selisih 0.43. Serta penggunaan SVM sebagai *classifier* dapat bekerja dengan baik untuk menangani *input space* dengan dimensi yang tinggi[12]. Penelitian ini juga menggunakan fitur-fitur tambahan seperti *N-gram* dan TF-IDF. Pada penelitian[13] yang meneliti pemilu presiden di Mesir, penelitian tersebut hanya menggunakan unigram, sehingga pada penelitian ini dilakukan penggunaan kombinasi *N-gram*

dengan mengkombinasikan unigram, bigram dan trigram, karena kombinasi *N-gram* terbukti dapat meningkatkan performansi sistem pada klasifikasi Bahasa[12] dan belum pernah dilakukan sebelumnya pada kasus pemilihan umum. TF-IDF digunakan sebagai *feature selection*, yang mana *feature selection* digunakan untuk mengidentifikasi fitur dalam data set yang sama pentingnya, dan membuang fitur lainnya sebagai informasi yang tidak relevan[14,3] dan TF-IDF dipilih karena dimensi pada dokumen berkurang tanpa kehilangan data[15]

Maka dari itu, tujuan dari penelitian ini adalah untuk (1) memperoleh kombinasi *N-gram* yang terbaik, (2) mengetahui pengaruh fitur *hashtag* sebagai salah satu fitur bawaan pada twitter terhadap proses klasifikasi (3) mengetahui pengaruh penggunaan nilai batas ambang ukur bawah (*threshold*) DF pada proses seleksi fitur. (4) Mengetahui akurasi analisis sentimen terhadap pemilu presiden RI menggunakan metode SVM