**MASALAH**

Hoax adalah informasi atau berita yang mengandung hal-hal yang tidak pasti atau yang

tidak berdasarkan fakta atas sesuatu yang benar terjadi. Hoax juga dapat diidentifikasi

dengan beberapa ciri: berita datang dari sumber yang tidak jelas /tidak dipercaya.

Gambar, foto atau video digunakan adalah hasil rekayasa, menggunakan kalimat

provokatif, mengandung politik maupun ras. Penyebaran hoax di kalangan masyarakat

dapat menyebabkan efek negatif, seperti kerusakan,kerugian, baik materiil dan

psikologis, hilangnya kepercayaan masyarakat, dan sebagainya.

Dampak dari penyebaran berita hoax akan memiliki konsekuensi dan bahaya buruk

untuk banyak pihak, yang mana hoax dapat menyebabkan kerugian dari berbagai

aspek, baik waktu dan ekonomi, publik panik, memburuknya hubungan sosial dan

sebagainya.

**METODE PENELITIAN**

1. Sumber Data

Data yang digunakan dalam penelitian ini diambil dari artikel berita media online yang

sudah memiliki label hoax dan non-hoax. dengan 50 artikel berlabel hoax dan 100

artikel berlabel non-hoax. Jadi total keseluruhan artikel yang digunakan sebagai data

pada penelitian ini adalah 150 artikel.

2. Tahapan Analisa

Dalam penelitian ini, beberapa tahapan analisa yang digunakan yaitu:

a. Pengumpulan berita dari situs berita online diantaranya :

viva.co.id, detik.com, kompas.com, liputan6.com, metrotvnews.com,

beritasatu.com, cnnindonesia.com, idntimes.com, republika.co.id, prokal.co,

cekfakta.com, jpnn.com, okezone.com, sindonews.com, solopos.com,

tempo.co, merdeka.com, tribunnews.com

b. data berita yang telah dikumpulkan kemudian disimpan dalam format CSV

dan kemudian akan melewati tahap text preprocessing, tahap pertama adalah

tokenizing dimana semua kalimat dari data berita akan dipisahkan

berdasarkan tanda spasi.

c. Kemudian pada proses folding case, semua huruf diubah menjadi huruf kecil,

kemudian kata-kata yang memiliki arti yang sama akan dinormaisasi menjadi

satu kata yang sama.

d. Pada proses filtering, karakter selain huruf dan angka akan dihapuskan,

misalnya tanda baca (.).

e. Pada tahap stopwords removing, kata yang tidak penting dan tidak unik akan

dihapuskan.

f. Selanjutnya pada proses stemming, awalan dan atau akhiran kata dihilangkan,

sehingga didapat kata dasarnnya saja.

g. Tahap terakhir pada text preprocessinf adalah menghitung bobot tiap kata

menggunakan pembobotan TF-IDF dengan penggunaan kombinasi fitur

unigram dan bigram.

h. Data kemudian dibagi 2, training data (80%) dan test data (20%). pemilihan

training dan test data akan dilakukan secara acak oleh program.

i. Lakukan Klasifikasin teka menggunakan algoritma machine learing, seperti

Multilayer

Perceptron, Support Vector Machine, Naïve Bayes, dan Random Forest.

j. Bandingkan hasil antara algoritma berdasarkan tingkat akutasi, presisi, recall

dan score F-1 dari tiap algoritma.

**HASIL DAN PEMBAHASAN**

1. Precision

Nilai Precision adalah tingkat akurasi berdasarkan informasi yang diminta pengguna

dengan jawaban yanh diberikan sistem.

2. Recall

Recall adalah nilai kesuksesan sistem dalam menemukan kembali sebuah informasi,

dengan kata lain Recall menunjukan selengkap apa hasil relevan yang ditampulkan

sistem.

3. Skor F-1

Skor F-1 adalah perhitungan nilai performa yang dilakukan untuk melihat hasil yabg

didapat dari proses klasifikasi berdasarkan nilai Presisi dan recall yang telah didapat

sebelumnya.

4. Accuracy

Level Accuracy adalah level kedekatan antara nilai prediksi dengan nilai aktual.

**KESIMPULAN**

Pada studi ini, sistem klasifikasi berita hoax di Indonesia telah dibuat menggunakan

machine learing dengan 4 macam algoritma, antara lain : Multilayer Perceptron, Naïve

Bayes, Support Vector Machine, dan Random Forest.