

Tugas Besar IF4072 Teks

Rancangan Aplikasi Berbasis Modul NLP



Dibuat oleh:

13515009 - Gisela Supardi

13515033 - Andika Kusuma

PROGRAM STUDI INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG

2018

I. Pendahuluan

I.1. Deskripsi Singkat Aplikasi

Aplikasi yang dibangun adalah aplikasi yang dapat mendeteksi kampanye pada cuitan Twitter yang terkait pada pemilihan umum (Pemilu) presiden Indonesia tahun 2019 mendatang. Selain memisahkan cuitan kampanye dan nonkampanye, aplikasi juga dapat menentukan kubu yang didukung oleh cuitan yang telah terdeteksi sebagai kampanye. Pengguna dapat memilih untuk melakukan deteksi pada cuitan acak di aplikasi Twitter atau pada teks yang dimasukkan sendiri oleh pengguna.

I.2. Latar Belakang

Mendekati penghujung tahun 2018, cuitan masyarakat Indonesia pada aplikasi Twitter kian banyak diwarnai dengan isu-isu seputar Pemilu Presiden. Banyaknya cuitan terkait topik ini tidak mengherankan mengingat pelaksanaan Pemilu semakin dekat, yakni pada tahun 2019 mendatang. Masa-masa ini merupakan waktu yang tepat bagi kedua pasangan calon (paslon) presiden dan wakil presiden untuk menarik hati masyarakat. Para pendukung masing-masing kubu juga terlihat gencar membela pasangan calon jagoannya.

Melihat tren ini, kami memutuskan untuk membuat suatu aplikasi yang dapat mendeteksi cuitan dengan konotasi kampanye. Tujuan pembangunan aplikasi ini adalah untuk mencari tahu kepopuleran Twitter sebagai media kampanye pendukung masing-masing kubu. Di sisi lain, aplikasi ini juga dapat dimanfaatkan untuk mendeteksi kemungkinan adanya pelanggaran salah satu peraturan Pemilu, yaitu larangan kampanye pada masa tenang. Sebagai informasi, masa tenang Pemilu di Indonesia ialah sejak tiga hari sebelum hari pemungutan suara.

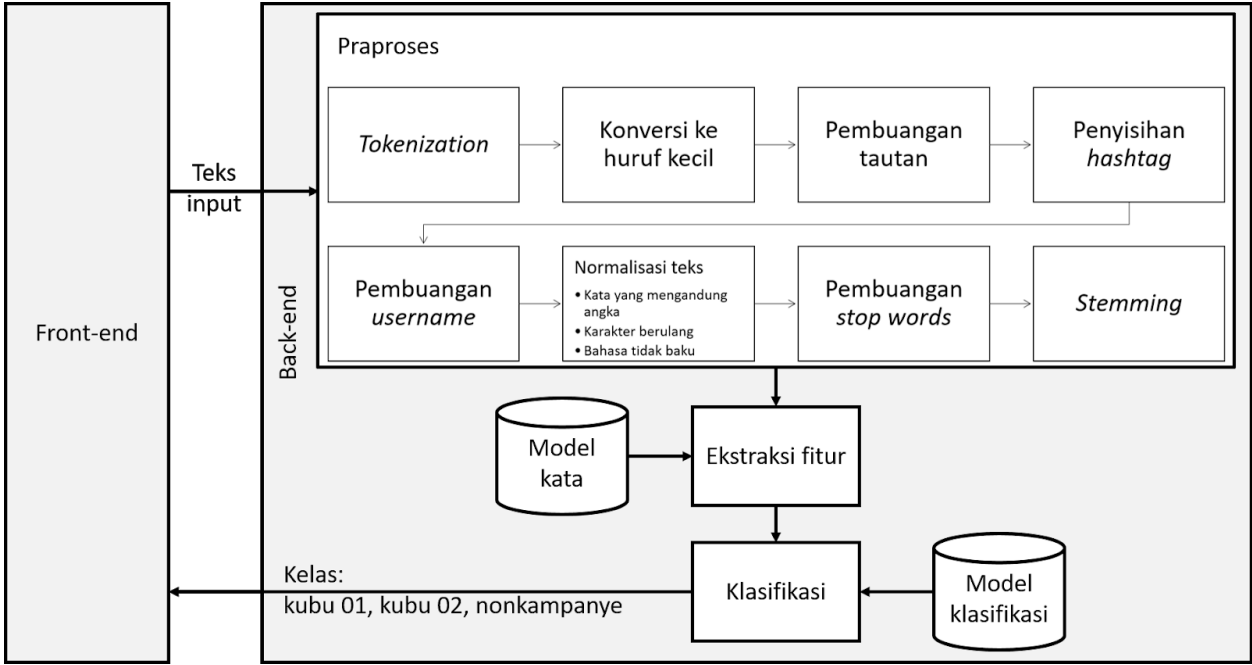
I.3. Aplikasi Lain Terkait

Pada dasarnya, aplikasi deteksi kampanye ini mirip dengan prinsip analisis sentimen dengan subjek berupa kubu masing-masing paslon. Terlebih lagi, pada Pemilu mendatang, paslon hanya terdiri dari dua kubu sehingga narasi bernada negatif terhadap salah satu paslon dapat digeneralisasi sebagai dukungan terhadap paslon lainnya. Sementara itu, narasi yang tidak bernada positif maupun negatif serta yang tidak menyinggung tentang Pemilu dapat digolongkan ke dalam kelas netral.

Penelitian yang dilakukan oleh Hidayatullah dan Ma'arif dengan judul “Pre-processing Tasks in Indonesian Twitter Messages” menjadi contoh penanganan teks yang didapat dari media sosial. Pada makalah tersebut, dirincikan arsitektur untuk menjernihkan data yang didapat dari Twitter. Arsitektur yang diajukan dalam makalah tersebut menjadi salah satu acuan dalam pembuatan struktur pra-proses data dalam aplikasi detektor kampanye.

II. Pembangunan Aplikasi

II.1. Arsitektur



Gambar 1 Arsitektur aplikasi

II.1.1. Deskripsi

Tabel 1 merincikan deskripsi setiap komponen sesuai yang terpampang pada Gambar 1. Dalam tabel juga dinyatakan kode untuk setiap komponen yang selanjutnya akan digunakan untuk mengacu masing-masing komponen tersebut. Khusus untuk komponen praproses akan dipecah lagi deskripsinya berdasarkan subkomponen.

Tabel 1 Deskripsi komponen

Kode	Komponen	Deskripsi
A	Praproses	Komponen pemroses pemformatan teks masukan. Deskripsi yang lebih detail dibahas per subkomponennya.

A-1	<i>Tokenization</i>	Teks yang diterima akan dipecah ke dalam rangkaian token. Proses ini sekaligus digunakan untuk menghapus tanda baca dari token.
A-2	Konversi ke huruf kecil	Seluruh token yang sudah diperoleh kemudian dikonversi ke dalam huruf kecil agar kata yang sama dapat dikenali sebagai satu kesatuan.
A-3	Pembuangan tautan	Token yang berupa tautan akan dihapus dari daftar karena sistem tidak dirancang untuk menelusuri tautan, jadi tautan tidak berguna.
A-4	Penyisihan <i>hashtag</i>	<i>Hashtag</i> pada Twitter digunakan sebagai penanda bahwa cuitan mengusung tema tertentu. Oleh karena itu <i>hashtag</i> dirasa cukup penting untuk diproses secara terpisah maka disimpan secara terpisah juga.
A-5	Pembuangan <i>username</i>	<i>Username</i> , kecuali milik para paslon, tidak terlalu dibutuhkan untuk tujuan ini karena dalam deteksi kampanye dianggap tidak memberi efek. Maka, selain @jokowi, @prabowo, dan @sandiono, <i>username</i> akan dibuang.
A-6	Normalisasi	Ada tiga jenis normalisasi yang dilakukan: a. Konversi angka yang dicampurkan dengan huruf Angka dapat diartikan sebagai tiga hal: digunakan secara literal (seperti t4 untuk kata tempat), pengulangan (seperti hati2 untuk kata hati-hati), dan sebagai pengganti huruf (seperti s4ya untuk kata saya). b. Karakter berulang Pengulangan karakter biasanya digunakan dengan maksud pemberian efek durasi. Fenomena ini diasumsikan terjadi dengan kasus satu karakter yang diulang lebih dari dua kali. Sementara karakter yang berjumlah tepat dua dianggap memang merepresentasikan kata aslinya karena ada kata dalam bahasa Indonesia yang memang mengakomodasi dua karakter beruntun. c. Bahasa tidak baku dan kesalahan penulisan Pada media sosial, seringkali terjadi kesalahan ketik (<i>typo</i>), juga sering digunakan kata tidak baku atau singkatan karena bahasa pada media sosial umumnya menggunakan bahasa lisan.
A-7	Pembuangan <i>stop words</i>	Komponen ini digunakan untuk membuang <i>stop words</i> atau kata yang sering digunakan secara netral dalam jumlah yang banyak.
A-8	<i>Stemming</i>	<i>Stemming</i> dilakukan untuk membuang imbuhan agar didapat kata dasarnya. Metode ini dipilih dibandingkan <i>lemmatize</i> karena <i>library</i> langkah ini untuk Bahasa Indonesia lebih mudah didapat. Selain itu, perubahan bentuk kata dalam Bahasa Indonesia tidak terlalu banyak seperti pada Bahasa Inggris sehingga <i>stemming</i> dianggap sudah cukup.

B	Ekstraksi fitur	Proses ini dilakukan untuk mendapatkan fitur dari teks masukan yang sesuai dengan model yang dimiliki.
C	Klasifikasi	Pada tahap ini, teks akan diklasifikasi sesuai dengan fitur-fitur yang dimilikinya berdasarkan model klasifikasi.
D	Model kata	Model kata adalah representasi fitur yang akan digunakan saat melakukan ekstraksi fitur.
E	Model klasifikasi	Model klasifikasi adalah komponen hasil pembelajaran mesin yang digunakan untuk mendeteksi kampanye serta kubu yang didukung kampanye tersebut.
F	Front-end	Modul ini adalah sarana interaksi antara pengguna dengan aplikasi. Sistem menerima masukan berupa teks berbahasa Indonesia. Selain teks yang dimasukkan sendiri, pengguna juga dapat memilih untuk memanen (<i>scraping</i>) data dari Twitter. Sementara luaran yang dihasilkan berupa kelas dari teks. Ada tiga jenis kelas yang dapat ditampilkan: kubu 01, kubu 02, dan nonkampanye.

II.1.2. Contoh Masukan dan Keluaran

Tabel 2 Contoh masukan dan keluaran setiap komponen

Komponen	Masukan	Keluaran
A-1	@twitter @jokowi @prabowo q sih yg penting #Pemilu2019 berjalan lancar.Siapa yg se? https://t.co/as23hbAR	['@twitter', '@jokowi', '@prabowo', 'q', 'sih', 'yg', 'penting', '#Pemilu2019', 'berjalan', 'lancar', 'Siapa', 'yg', 'se', 'https://t.co/as23hbAR']
A-2	['@twitter', '@jokowi', '@prabowo', 'q', 'sih', 'yg', 'penting', '#Pemilu2019', 'berjalan', 'lancar', 'Siapa', 'yg', 'se', 'https://t.co/as23hbAR']	['@twitter', '@jokowi', '@prabowo', 'q', 'sih', 'yg', 'penting', '#pemilu2019', 'berjalan', 'lancar', 'siapa', 'yg', 'se', 'https://t.co/as23hbAR']
A-3	['@twitter', '@jokowi', '@prabowo', 'q', 'sih', 'yg', 'penting', '#pemilu2019', 'berjalan', 'lancar', 'siapa', 'yg', 'se']	['@twitter', '@jokowi', '@prabowo', 'q', 'sih', 'yg', 'penting', '#pemilu2019', 'berjalan', 'lancar', 'siapa', 'yg', 'se']
A-4	['@twitter', '@jokowi', '@prabowo', 'q', 'sih', 'yg', 'penting', '#pemilu2019', 'berjalan', 'lancar',	['@twitter', '@jokowi', '@prabowo', 'q', 'sih', 'yg', 'penting', 'berjalan', 'lancar', 'siapa', 'yg', 'se']

	['siapa', 'yg', 'se7']	['#pemilu2019']
A-5	['@twitter', '@jokowi', '@prabowo', 'q', 'sihhh', 'yg', 'penting', 'berjalan', 'lancar', 'siapa', 'yg', 'se7'] ['#pemilu2019']	['@jokowi', '@prabowo', 'q', 'sihhh', 'yg', 'penting', 'berjalan', 'lancar', 'siapa', 'yg', 'se7'] ['#pemilu2019']
A-6a	['@jokowi', '@prabowo', 'q', 'sihhh', 'yg', 'penting', 'berjalan', 'lancar', 'siapa', 'yg', 'se7'] ['#pemilu2019']	['@jokowi', '@prabowo', 'q', 'sihhh', 'yg', 'penting', 'berjalan', 'lancar', 'siapa', 'yg', 'setuju'] ['#pemilu2019']
A-6b	['@jokowi', '@prabowo', 'q', 'sihhh', 'yg', 'penting', 'berjalan', 'lancar', 'siapa', 'yg', 'setuju'] ['#pemilu2019']	['@jokowi', '@prabowo', 'q', 'sih', 'yg', 'penting', 'berjalan', 'lancar', 'siapa', 'yg', 'setuju'] ['#pemilu2019']
A-6c	['@jokowi', '@prabowo', 'q', 'sih', 'yg', 'penting', 'berjalan', 'lancar', 'siapa', 'yg', 'setuju'] ['#pemilu2019']	['@jokowi', '@prabowo', 'saya', 'sih', 'yang', 'penting', 'berjalan', 'lancar', 'siapa', 'yang', 'setuju'] ['#pemilu2019']
A-7	['@jokowi', '@prabowo', 'saya', 'sih', 'yang', 'penting', 'berjalan', 'lancar', 'siapa', 'yang', 'setuju'] ['#pemilu2019']	['@jokowi', '@prabowo', 'sih', 'penting', 'berjalan', 'lancar', 'siapa', 'setuju'] ['#pemilu2019']
A-8	['@jokowi', '@prabowo', 'sih', 'penting', 'berjalan', 'lancar', 'siapa', 'setuju'] ['#pemilu2019']	['@jokowi', '@prabowo', 'sih', 'penting', 'jalan', 'lancar', 'siapa', 'setuju'] ['#pemilu2019']
B	['@jokowi', '@prabowo', 'sih', 'penting', 'jalan', 'lancar', 'siapa', 'setuju'] ['#pemilu2019']	[0, 0, 0, 0, 1, 0, 0, ..., 1, 0]
C	[0, 0, 0, 0, 1, 0, 0, ..., 1, 0]	non-kampanye

II.2. Teknik

Selain pembuangan *stop words* dan *stemming*, seluruh komponen dalam aplikasi deteksi kampanye diimplementasikan sendiri. Subbab ini akan membahas perrincian teknik yang digunakan pada setiap komponen. Pembahasan terbagi berdasarkan komponen.

II.2.1. Praproses

Modul ini memiliki delapan subkomponen. Setiap subkomponen memiliki tujuan yang berbeda sehingga menerapkan teknik yang berbeda pula yang dijabarkan sebagai berikut.

Komponen	Teknik
A-1	<i>Tokenization</i> dilakukan dengan memecah teks berdasarkan spasi dan <i>endline</i> . Selain itu, pemecahan juga dilakukan berdasarkan tanda baca.
A-2	Komponen ini menggunakan fungsi yang disediakan oleh Python untuk mengubah suatu string menjadi huruf kecil.
A-3	Pembuangan tautan menggunakan asumsi bahwa semua tautan ditandai dengan substring “http” dan substring tersebut hanya dimiliki oleh tautan. Sehingga setiap token yang mengandung “http” akan dibuang dari daftar.
A-4	Setiap <i>hashtag</i> yang ditemui dalam daftar token akan disimpan ke dalam daftar <i>hashtag</i> . Sementara itu, token ini akan dihapuskan dari daftar token utama.
A-5	Pengenalan <i>username</i> dilakukan dengan melihat apabila terdapat karakter ‘@’ di awal token. Setiap token yang berawalan ‘@’, kecuali yang sudah dinyatakan pada Tabel 1, akan dibuang dari daftar.
A-6a	Konversi dilakukan menggunakan model yang sudah dilatih dari sebuah korpus kecil. Model menggunakan <i>Decision Tree</i> dikarenakan model ini dapat memodelkan kasus normalisasi ini dengan cepat dan cukup baik. Fitur yang digunakan untuk model adalah <angka yang ingin diubah, karakter sebelum, boolean karakter pertama, boolean karakter terakhir, boolean karakter sebelumnya huruf hidup, boolean huruf representasi angka huruf hidup>. Model akan menghasilkan klasifikasi apakah angka akan diubah menjadi pengulangan kata sebelumnya, pelafalan angka tersebut, atau representasi huruf dari angka tersebut.
A-6b	Normalisasi terhadap huruf berulang dilakukan menggunakan aturan untuk menghapus beberapa kemunculan huruf berulang.
A-6c	Normalisasi kata informal ataupun <i>slang</i> dilakukan dengan melakukan konversi kemunculan kata informal maupun <i>slang</i> yang terdapat pada kamus yang telah dibuat menjadi kata baku.
A-7	Pembuangan <i>stop words</i> digunakan menggunakan <i>corpus</i> yang terdapat pada <i>library</i> nltk untuk Bahasa Indonesia.
A-8	<i>Stemming</i> dilakukan dengan menggunakan <i>library</i> Sastrawi (https://pypi.org/project/PySastrawi/).

II.2.2. Ekstraksi Fitur

Modul ini bertujuan untuk melakukan ekstraksi fitur terhadap masukan cuitan yang telah dilewatkan ke modul praproses. Fitur yang dikeluarkan berupa larik dengan nilai nol atau satu berdasarkan kemunculan token atau *hashtag* pada masukan sesuai dengan model kata yang ada.

II.2.3. Klasifikasi

Teknik yang digunakan untuk modul ini hanyalah pencocokan antara fitur dengan model klasifikasi yang dimiliki. Modul ini akan menghasilkan tiga jenis label klasifikasi, yakni 0 (nonkampanye), 1 (kubu 01), dan 2 (kubu 02).

II.2.4. Model Kata

Model kata didapat dengan menggunakan teknik pemilihan kata dengan frekuensi terbanyak pada korpus. Pertama, daftar token akan direpresentasikan sebagai *bag of words* dan dihitung frekuensi kemunculannya. Kata yang melebihi batas ambang akan disimpan sebagai rangkaian fitur yang akan dipertimbangkan dalam klasifikasi.

Fitur lainnya yang dijadikan model ialah *hashtag* yang terdaftar pada korpus. Semua *hashtag* yang ditemukan pada data latih dibuat menjadi fitur. Pertimbangan keputusan ini ialah bahwa *hashtag* dapat menjadi representasi yang cukup penting dari sebuah cuitan. Kedua representasi fitur ini kemudian akan digabungkan untuk menjadi fitur keseluruhan pada klasifikasi.

II.2.5. Model Klasifikasi

Model klasifikasi diperoleh dengan melakukan pelatihan terhadap korpus. Terdapat tiga buah model yang dilatih pada aplikasi ini, yaitu model *Decision Tree*, *Support Vector Machine*, dan *Multilayer Perceptron*. Klasifikasi cuitan pada aplikasi dilakukan dengan mengambil hasil *voting* dari ketiga model ini. Model klasifikasi akan mengeluarkan hasil prediksi berupa nilai label 0, 1, atau 2.

II.3. Eksperimen

II.3.1. Informasi Data

Data yang digunakan adalah data hasil panen (*scraping*) pada Twitter menggunakan *library* “tweepy”. Cuitan dipanen menggunakan filter berupa *geocode* agar dapat memperoleh

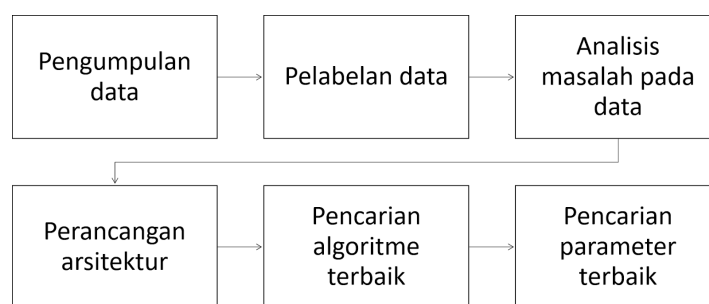
cuitan yang berasal dari Indonesia, terkhusus Pulau Jawa. Adapun koordinat yang digunakan adalah koordinat untuk Gunung Tidar dengan radius 600 kilometer. Gunung Tidar digunakan sebagai titik tengah karena beberapa sumber mengatakan bahwa gunung tersebut merupakan titik tengah Pulau Jawa. Cuitan juga dipanen menggunakan filter untuk menghilangkan cuitan dengan media gambar maupun video.

Hasil panen kemudian diberi label untuk keperluan pembelajaran. Total data yang berhasil dikumpulkan dan diberi label ialah sebanyak 2411 data cuitan. Awalnya data dipanen tanpa menggunakan *query* apa pun. Namun ternyata data kampanye yang diperoleh menjadi sangat sedikit sehingga akhirnya digunakan beberapa *query* untuk memperoleh cuitan seputar Pemilu seperti contohnya “paslon” dan “pemilu2019”.

Data yang dipilih sebagai korpus adalah data yang cuitan yang mayoritas merupakan Bahasa Indonesia. Cuitan dengan Bahasa Inggris atau bahasa daerah tidak dimasukkan sebagai korpus. Oleh karena itu, pada proses deteksi juga diasumsikan bahwa masukan berupa Bahasa Indonesia. Selain cuitan dengan bahasa yang berbeda, cuitan yang hanya berupa *mention* atau *link* juga dihapus dari korpus.

II.3.2. Skenario Eksperimen

Eksperimen bertujuan untuk menemukan teknik pemrosesan yang optimal untuk kasus klasifikasi ini. Selain itu, dicari pula algoritme terbaik beserta parameternya sesuai dengan fitur yang ingin diproses. Skenario eksperimen secara umum dapat dilihat pada Gambar 2.



Gambar 2 Alur skenario eksperimen

II.3.3. Hasil

Hasil dari langkah pengumpulan dan pelabelan data telah dijelaskan sebelumnya pada subbab II.3.1 sehingga tidak akan dibahas lagi. Setelah terbentuk korpus, langkah selanjutnya adalah menganalisis masalah yang ada pada data agar dapat merancang arsitektur yang mampu

menangani masalah tersebut. Hasil analisis menunjukkan bahwa korpus memiliki beberapa masalah:

- mengandung tautan yang tidak dapat dibuka sistem,
- banyak kata yang tidak baku,
- tidak memiliki struktur kalimat yang jelas,
- banyak kesalahan ketik,
- memuat fenomena alih-kode ke Bahasa Inggris atau bahasa daerah, serta
- memuat kata alay.

Oleh karena itu, perlu dilakukan beberapa normalisasi untuk memperbaiki data cuitan yang dituangkan melalui rancangan arsitektur sistem dan pelatihan data, khususnya pada bagian praproses. Hasil rancangan arsitektur akhir dapat dilihat melalui Gambar 1.

Dalam rancangan arsitektur akhir, terdapat tiga model yang dihasilkan. Model pertama digunakan untuk melakukan normalisasi pada komponen A-6a. Model yang bertujuan untuk melakukan klasifikasi kasus pemrosesan ini didapat melalui eksperimen singkat dengan korpus sederhana. Model kedua adalah model kata, seperti yang sudah dijelaskan pada II.2.4. Teknik representasi fitur dipilih mengingat masalah ketidakjelasan struktur kalimat. Model ketiga adalah model klasifikasi yang juga telah dibahas pada II.2.5. Eksperimen untuk model ini lebih kompleks karena melibatkan korpus yang cukup besar dan pemilihan algoritme beserta parameter yang paling cocok.

Eksperimen untuk model klasifikasi memberikan parameter terbaik untuk tiga jenis model klasifikasi yang digunakan, yaitu model *Decision Tree*, *Support Vector Machine*, dan *Multilayer Perceptron*. Parameter terbaik untuk setiap model serta akurasi terhadap data latih dan data tes dapat dilihat pada Tabel 1. Setelah itu, ketiga model dengan parameter terbaik tersebut akan dipergunakan untuk menghasilkan sebuah model baru dengan bobot masing-masing model sama yaitu 0.33.

Tabel 1 Hasil Eksperimen Model Klasifikasi

Model	Parameter Terbaik	Akurasi Data Latih	Akurasi Data Tes
<i>Decision Tree</i>	criterion='gini' max_depth=9 min_samples_leaf=5	88.8428%	89.6694%

	min_samples_split=2		
<i>Support Vector Machine</i>	C=10 gamma=0.01	91.2863%	92.1488%
<i>Multilayer Perceptron</i>	activation='logistic' hidden_layer_sizes=(10,) learning_rate='constant'	93.2688%	91.7355%

Selain parameter terbaik untuk setiap model, terdapat juga parameter pengambilan fitur yang dapat diubah yaitu *threshold*. Parameter *threshold* adalah batas jumlah kemunculan minimal yang harus dipenuhi oleh suatu token agar dapat ditetapkan sebagai fitur. *Threshold* yang digunakan adalah 10 untuk jumlah data latih sekitar 2000 cuitan (0.5%).

II.3.4. Analisis Hasil

Arsitektur praproses yang dihasilkan masih memiliki banyak kekurangan, dilihat dari rangkaian fitur hasil ekstraksi yang masih kurang menggambarkan teks. Komponen yang dirasa masih perlu banyak pengembangan adalah komponen A-6c. Ada banyak singkatan yang dipergunakan dalam *tweet* masyarakat Indonesia yang ambigu karena dapat berarti beberapa hal. Selain itu, kesalahan ketik belum semuanya terpetakan dalam kamus. Diperlukan juga konteks kalimat selain kamus kata untuk memperbaiki kesalahan ketik. Oleh sebab itu, kami mengusulkan untuk menambahkan subkomponen pemberian etiket POS (*part-of-speech*) agar dapat membantu proses normalisasi.

Di sisi lain, komponen A-7 juga memerlukan pengembangan. Hal ini mengingat daftar kata yang menjadi fitur masih mengandung kata kurang bermakna seperti “haha” dan “wkwk”. Sebagai informasi, kedua kata tersebut sering digunakan dalam cuitan media sosial Twitter untuk mengekspresikan emosi tertawa.

Kami juga mendapati bahwa masih ada *username* lain di luar yang didaftarkan yang dapat digunakan untuk mengindikasi kampanye. Contohnya adalah akun-akun resmi partai politik di Indonesia. Ada beberapa *tweet* dalam korpus yang menyatakan dukungan ataupun serangan dengan melibatkan *username* parpol. Oleh karena itu, sebaiknya *username* tidak dibuang atau dilakukan penambahan daftar putih.

Model kata dirasa masih sangat kurang merepresentasikan fitur yang signifikan untuk klasifikasi. Hipotesis kami ialah bahwa fitur yang dipilih kurang tepat. Fitur yang digunakan mengasumsikan jarak setiap fitur sama, yaitu 1.

Untuk model klasifikasi, hasil eksperimen menunjukkan nilai akurasi ketiga model yang cukup baik pada data latih maupun data tes. Namun, model klasifikasi yang dihasilkan masih belum dapat mengenali cuitan di luar data tes secara sempurna. Hal ini kemungkinan besar disebabkan oleh kurang meratanya pengambilan cuitan pada saat pemanenan data. Pemanenan data sendiri dilakukan dengan menggunakan kakas Tweepy yang melakukan pemanenan pada waktu yang berdekatan. Selain itu, jumlah korpus latih juga masih sangat sedikit sehingga belum dapat mencakup semua jenis cuitan yang mungkin. Ditambah lagi, tingginya kreativitas pengguna Twitter dalam menghasilkan cuitan kampanye secara tidak langsung menurunkan kinerja aplikasi ini.

Simpulannya, aplikasi detektor kampanye ini masih memiliki banyak kekurangan karena keterbatasan waktu dalam implementasi. Saran kami, apabila aplikasi ingin dikembangkan lebih lanjut, sebaiknya korpus diperbesar agar dapat lebih merepresentasikan dunia nyata. Kemudian, perlu dilakukan pengembangan untuk komponen A-5, A-6c, A-7, dan D. Selain itu, perlu adanya penambahan praproses berupa pembubuhan etiket POS untuk dapat digunakan sebagai fitur model klasifikasi serta pengembangan komponen A-6c.

II.4. Pembagian Peran

Anggota	Peran
13515009	Pelabelan data, eksperimen, <i>frontend</i> aplikasi, makalah
13515033	Modul preprocess, eksperimen, model normalisasi angka, <i>backend</i> aplikasi, makalah