

UJIAN PRAKTIK NATURAL LANGUAGE PROCESSING: ANALISIS SENTIMEN TWEET PILKADA DKI JAKARTA 2017

M. Andrew Arjunanda Yasin



Kelas Ular

PENDAHULUAN



Latar Belakang

Pada sistem politik demokrasi di Indonesia, media sosial memainkan peranan penting sebagai media kampanye pemilu (pemilihan umum). Hal ini terjadi karena media sosial dapat memudahkan penggunanya mengakses berita politik paling mutakhir serta opini masyarakat yang tidak dimuat dalam surat kabar nasional maupun televisi (Alexa, 2013). Twitter dilirik karena mudah diakses, jumlah pengikut tidak terbatas dan jumlah karakter pesan singkat (tweet) hanya 140 karakter sehingga mampu menyampaikan maksud dan tujuan dengan singkat, padat dan jelas. Tahun 2017 merupakan periode akhir masa jabatan Gubernur DKI Jakarta, sehingga perlu diselenggarakan pemilihan kepala daerah (pilkada) periode 2017-2022. Pada Pilkada ini terdapat tiga pasangan cagub (calon gubernur) beserta cawagub (calon wakil gubernur). Dalam pra-pelaksanaan maupun pelaksanaan Pilkada DKI Jakarta terdapat berbagai opini dan tanggapan dengan sentimen positif dan negatif pada Twitter.



PENDAHULUAN



Tujuan

Untuk tingkat akurasi dari sentimen negatif dan positif pada dataset dengan binary_crossentropy

Urgensi

Permasalahan yang muncul ketika menganalisa seluruh hasil sentimen dan pengklasifikasian tweet pada Twitter secara manual maka akan dibutuhkan waktu dan usaha yang banyak dalam pelaksanaannya. Pada periode sebelumnya, prediksi hasil pilkada dilakukan oleh lembaga survei dan quick count (Kompas, 2014). Hasil survei yang dilakukan oleh lembaga survei terkadang tidak sesuai dengan kenyataan, namun dengan adanya survei yang dilaksanakan pada pra-pemilu membuat cagub dan cawagub mengetahui parameter kemenangan dan kekalahan lebih dini sehingga bisa memperbaiki strategi kampanye. Dengan adanya analisa sentimen ini, diharapkan hasil sentimen menjadi benar – benar memiliki batas yang jelas antara sentimen positif dan negatif (Dayalani & Patil, 2014) sehingga akurasi yang dihasilkan oleh sistem ini mengalami peningkatan.



DATASET



AutoSave ☐ dataset_tweet_sentiment_pilkada_DKI_2017 Search (Alt+Q)

File Home Insert Page Layout Formulas Data Review View Help

Undo Paste Font Alignment Number Styles Cells Editing

POSSIBLE DATA LOSS Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format. Don't show again Save As...

A1 Id,Sentiment,Pasangan Calon,Text Tweet

	A
1	Id,Sentiment,Pasangan Calon,Text Tweet
2	1,negative,Agus-Sylvi,Banyak akun kloning seolah2 pendukung #agussilvy mulai menyerang paslon #aniessandi dengan opini dan argumen pmbenaran..jangan terkecoh
3	2,negative,Agus-Sylvi,#agussilvy bicara apa kasihan yaa...lap itu air matanya wkwkwkwk
4	3,negative,Agus-Sylvi,Kalau aku sih gak nunggu hasil akhir QC tp lagi nunggu motif cuitan pak @SBYudhoyono kayak apa.. pasca #AgusSylvy Nyungsep..
5	4,negative,Agus-Sylvi,Kasian oh kasian dengan peluru 1milyar untuk tiap RW #agussilvy tidak mempan menangin pilkada #QuickCount #PilkadaSerentak2017
6	5,negative,Agus-Sylvi,Maaf ya pendukung #AgusSylvy..hayo dukung #AniesSandi diputar 2 @ronavioleta @NetizenTofa
7	6,negative,Agus-Sylvi,"aneh deh lebay masa ini di sangkut pautkan sama kandidat Calgub, anda lebay seperti yg anda dukung #agussilvy"
8	7,negative,Agus-Sylvi,Kepada Allah SWT kami ucapkan rasa syukur dan kepada pak @SBYudhoyono @presidenSBY #AHY kami sampaikan terimakasih. Makna 17% suara anda.
9	8,negative,Agus-Sylvi,Terima Kasih teruntuk pendukung #AHY #SYLVI kalianlah pembeda kali ini
10	9,negative,Agus-Sylvi,Trima Kasih atas Kelstiqomahan Relawan #AHY @AgusYudhoyono lah yg tih mengantarkan #AniesSandiDKI1 #JakartaGubernurBaru dg cukup signifikan
11	10,negative,Agus-Sylvi,Mengenang pidato kekalahan #ahy
12	11,negative,Agus-Sylvi,"dammnn,!! politik cantik #SBY ngorbanin #AHY"
13	12,negative,Agus-Sylvi,"Batal nyoblos, baru tau ternyata ga ada no.1 di kertas suara. #AHY #kangen #mosing"
14	13,negative,Agus-Sylvi,sudah boleh Ngakak? survey mu jauh panggang dari api ! #Ahy
15	14,negative,Agus-Sylvi,aku juga kemarin #AHY 19april2017 suaraku utk paslon 3 #DKI
16	15,negative,Agus-Sylvi,Karna yang berlebihan itu tidak baik ya bah #AHY
17	16,negative,Agus-Sylvi,"om AA kemana ? lepas #AHY kalah kok dia tutup mulut tidak koar"" mau bongkar SBY"
18	17,negative,Agus-Sylvi,kalau pencatutan nama #ahy dan mpo #silvi melanggar hukum ga ya.. wkwkwkwkwk
19	18,negative,Agus-Sylvi,ya pastilah #FPI kan ga suka ama Ahok jd pas #AHY udah kalah otomatis dukungannya lari ke mas #Anisa AniesSandi
20	19,negative,Agus-Sylvi,dia lah Banteng.. dia lah Garuda!! ksatria seperti bpk @prabowo gentle spt #AHY @AgusYudhoyono ..jadi cinta sama #HajiDjarot
21	20,negative,Agus-Sylvi,"Gara2 e-KTP dikorupsi, fotoku yg semula mirip #AHY, pelan tapi pasti skrang malah mirip - Stickmoticons""

dataset_tweet_sentiment_pilkada

Ready Accessibility: Unavailable

PREPROCESSING DATA

Preprocess: Convert Label

```
▶ #Mengubah sentimen menjadi menjadi 0 dan 1
#x menghimpun dataset
#y menghimpun sentimen
y = [ 0 if i=='negative' else 1 for i in data['Sentiment'] ] #0 untuk sentimen negatif dan 1 untuk sentimen positif
x = data['Text Tweet']
print(x[:5], y[:5])
```

```
↳ Id
1    Banyak akun kloning seolah2 pendukung #agussil...
2    #agussilvy bicara apa kasihan yaa...lap itu ai...
3    Kalau aku sih gak nunggu hasil akhir QC tp lag...
4    Kasian oh kasian dengan peluru 1milyar untuk t...
5    Maaf ya pendukung #AgusSilvy..haya dukung #Ani...
Name: Text Tweet, dtype: object [0, 0, 0, 0, 0]
```



EKSTRAKSI FITUR

Split Data

```
[14] X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.30) #Split 30% untuk data testing dan 70% untuk data training
```

```
[15] (unique, counts) = np.unique(y_test, return_counts=True)
      frequencies = np.asarray((unique, counts)).T
      print(frequencies)
      #Terdapat 142 data yang sentimennya negatif dan 128 data yang sentimennya positif
```

```
[[ 0 142]
 [ 1 128]]
```



MODEL

Model

```
▶ model = tf.keras.Sequential([  
    tf.keras.layers.Embedding(vocab_size, embedding_dim, input_length=max_length), #  
    tf.keras.layers.GlobalAveragePooling1D(), #Pooling 1 Dimensi dikarenakan bentuk  
    tf.keras.layers.Dense(24, activation='sigmoid'), #Multi Perceptron 1 layer beris  
    tf.keras.layers.Dropout(0.5), #Tambahkan Dropout untuk menghandle Overfit  
    tf.keras.layers.Dense(1, activation='sigmoid') #1 layer untuk output layer singl  
])  
model.compile(loss='binary_crossentropy',optimizer='nadam',metrics=['accuracy']) #Mo
```





PERFORMA MODEL



Training

```
▶ num_epochs = 90
  history = model.fit(training_padded, training_labels, batch_size = 10, epochs=num_epochs, validation_data=(testing

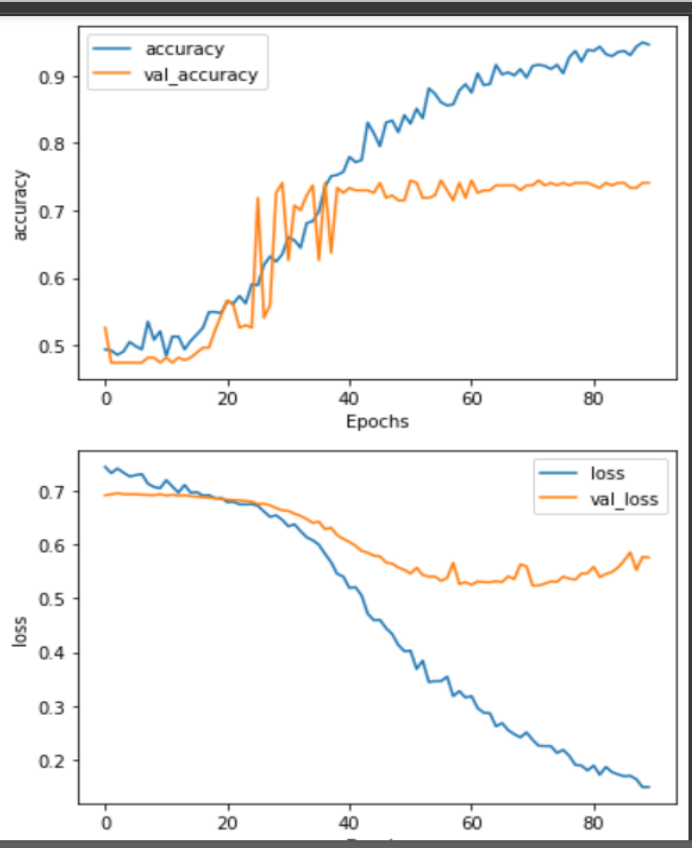
Epoch 59/90
↳ 63/63 - 0s - loss: 0.3274 - accuracy: 0.8778 - val_loss: 0.5263 - val_accuracy: 0.7407 - 240ms/epoch - 4ms/step
Epoch 60/90
63/63 - 0s - loss: 0.3153 - accuracy: 0.8873 - val_loss: 0.5302 - val_accuracy: 0.7185 - 254ms/epoch - 4ms/step
Epoch 61/90
63/63 - 0s - loss: 0.3180 - accuracy: 0.8746 - val_loss: 0.5249 - val_accuracy: 0.7444 - 267ms/epoch - 4ms/step
Epoch 62/90
63/63 - 0s - loss: 0.2957 - accuracy: 0.9032 - val_loss: 0.5313 - val_accuracy: 0.7259 - 240ms/epoch - 4ms/step
Epoch 63/90
63/63 - 0s - loss: 0.2870 - accuracy: 0.8857 - val_loss: 0.5302 - val_accuracy: 0.7296 - 242ms/epoch - 4ms/step
Epoch 64/90
63/63 - 0s - loss: 0.2862 - accuracy: 0.8873 - val_loss: 0.5297 - val_accuracy: 0.7296 - 252ms/epoch - 4ms/step
Epoch 65/90
63/63 - 0s - loss: 0.2618 - accuracy: 0.9159 - val_loss: 0.5318 - val_accuracy: 0.7370 - 263ms/epoch - 4ms/step
Epoch 66/90
63/63 - 0s - loss: 0.2679 - accuracy: 0.9016 - val_loss: 0.5301 - val_accuracy: 0.7370 - 235ms/epoch - 4ms/step
Epoch 67/90
```


PERFORMA MODEL

Display History

```
[20] #Matplotlib digunakan untuk visualiasi data
def plot_graphs(history, string):
    plt.plot(history.history[string])
    plt.plot(history.history['val_'+string])
    plt.xlabel("Epochs")
    plt.ylabel(string)
    plt.legend([string, 'val_'+string])
    plt.show()

plot_graphs(history, "accuracy")
plot_graphs(history, "loss")
```



KESIMPULAN



Kesimpulan nya adalah tingkat sentimen analisis opini berdasarkan dataset accuracy mencapai angka 0.9492 tetapi ada suatu kondisi dimana jika ia tidak dapat mendeteksi nama orang, tempat, dan makanan maka hasilnya akan selalu negatif



TERIMA KASIH