**Nama : I Gede Laksmana Yudha**

**Email : laksmanayudha22@gmail.com**

**SENTIMENT ANALISIS MENGENAI OMNIBUS LAW PADA TWITTER**

Omibus berasal dari bahasa latin yang artinya ‘untuk semuanya’. Maka secara sederhana Omnibus Law dapat diartikan suatu undang – undang yang mengatur banyak hal. Akhir – akhir ini di Indonesia Omnibus Law RUU Cipta Kerja menjadi sorotan karena terdapat beberapa pasalnya yang dianggap kontroversial, RUU Cipta Kerja dinilai serikat buruh hanya mementingkan kepentingan investor. RUU Cipta Kerja hanya salah satu bagian dari omnibus law. Dalam omnibus law, terdapat tiga RUU yang siap diundangkan, antara lain RUU tentang Cipta Kerja, RUU tentang Ketentuan dan Fasilitas Perpajakan untuk Penguatan Perekonomian, dan RUU tentang Pengembangan dan Penguatan Sektor Keuangan (Kompas.com - 05/10/2020, 10:22 WIB) . Maka dari itu analisis sentiment ini bertujuan untuk mengetahui tanggapan masyarakat Indonesia mengenai Pengesahan dari Omnibus Law tersebut.

**Proses pengambilan data dan pembersihan** dilakukan dengan menjalankan file laksmanayudha22\_tugas13.py. Selanjutnya untuk **proses analisis sentiment** dilakukan pada file final\_project.ipynb

Pada topik kali ini data yang digunakan merupakan data tweet yang diambil menggunakan twitter API dimulai dari tanggal 18-11-2020 sampai 28-11-2020. Dari pengambilan tersebut diperoleh 1094 data tweet.

**TOP & Most Data**

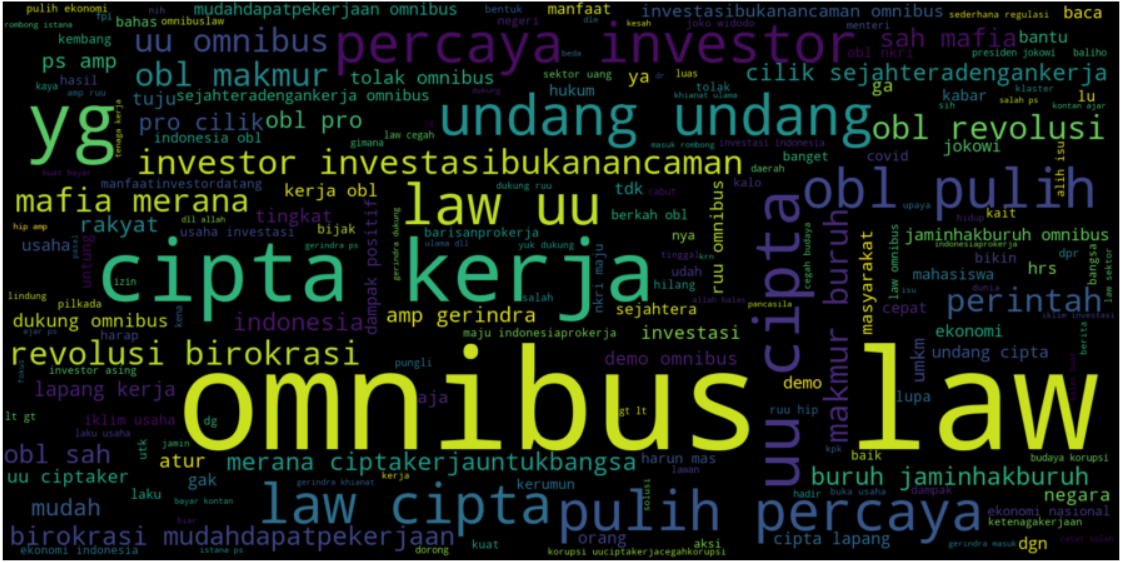
Kali ini akan dicoba mencari data yang sering muncul sehingga bisa didapatkan gambaran awal mengenai kecenderungan opini masyarakat. Langkah awal yang dilakukan adalah mengambil semua data yang diperlukan dari database. Setelah itu menggabungkan clean tweet menjadi 1 paragraf untuk, kemudian melakukan wordcloud dan plot data.

|  |
| --- |
| import sqlite3  import pandas as pd  import numpy as np  from scipy import stats  import itertools  from nltk.tokenize import word\_tokenize  from nltk.probability import FreqDist  from wordcloud import WordCloud  import matplotlib.pyplot as plt  import tweepy  # ambil data dari database  conn = sqlite3.connect('laksmanayudha22\_final.db')  cursor = conn.cursor()  query = """ SELECT tweetid, cleantweet FROM tweet; """  cursor.execute(query)  conn.commit()  all\_data = cursor.fetchall()  cursor.close()  conn.close()  # pisahkan tweetid dan clean tweet  clean\_tweet = []  tweet\_id = []  for data in all\_data:  clean\_tweet.append(data[1])  tweet\_id.append(data[0])  # word tokenize  data = [word\_tokenize(tweet) for tweet in clean\_tweet]  # multidimensional ke single dimension  data = list(itertools.chain(\*data))  fqdist = FreqDist(data)  print(fqdist.most\_common(20))  paragraf = ' '.join(data)  print(paragraf)  wordcloud = WordCloud(width=1600, height=800, max\_font\_size=200).generate(paragraf)  plt.figure(figsize=(16,10))  plt.imshow(wordcloud, interpolation='bilinear')  plt.axis("off")  plt.show() |

Output :

[('omnibus', 1073), ('law', 1073), ('obl', 423), ('kerja', 360), ('uu', 330), ('cipta', 279), ('yg', 191), ('undang', 180), ('amp', 153), ('buruh', 143), ('indonesia', 139), ('investor', 125), ('usaha', 124), ('pulih', 112), ('birokrasi', 106), ('investasi', 96), ('percaya', 96), ('revolusi', 95), ('sah', 93), ('investasibukanancaman', 90)]

20 kata yang sering muncul adalah omnibus, law, obl, kerja, revolusi, investasi bukan ancaman, dan seterusnya. Berikut gambar dari Word Cloud :



**Gambar 1. Word Cloud**

Dilihat dari WordCloud dan kata yang sering muncul dapat disimpulkan sementara omnibus law UU merupakan suatu investasi dan bukan ancaman. Kemudian UU ini dianggap lebih percaya terhadap investor daripada jaminan hak buruh.

**Sentimen Analisis - Polarity**

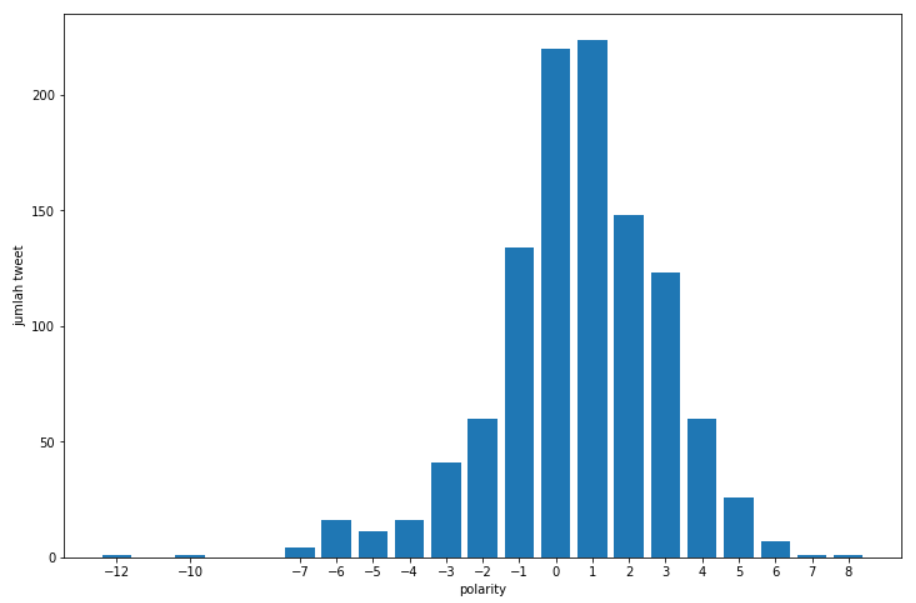
Untuk sentiment analisi awal menggunakan pendekatan polarity. Untuk melakukan hal tersebut pertama yang dibutuhkan adalah kumpulan kata positif dan negatif.

|  |
| --- |
| pos\_list= open("./kata\_positif.txt","r")  pos\_kata = pos\_list.readlines()  neg\_list= open("./kata\_negatif.txt","r")  neg\_kata = neg\_list.readlines() |

Kemudian mencari kecenderungan setiap tweet dan melakukan plot.

|  |
| --- |
| # loop kata pos neg  hasil = []  for tweet in clean\_tweet:  count\_pos = 0  count\_neg = 0  for kata\_pos in pos\_kata:  if kata\_pos.strip() in tweet:  count\_pos+=1  for kata\_neg in neg\_kata:  if kata\_neg.strip() in tweet:  count\_neg+=1  hasil.append(count\_pos - count\_neg)  labels, counts = np.unique(hasil, return\_counts=True)  fig, ax = plt.subplots(figsize=[12,8])  ax.bar(labels, counts, align='center')  ax.set\_xticks(labels)  ax.set\_xlabel("polarity")  ax.set\_ylabel("jumlah tweet")  plt.show() |

Output :

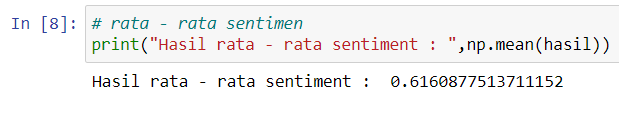


**Gambar 2. Perbandingan polarity dan jumalah tweet**

Rules :

* Total Nilai > 0, maka sentimen positif
* Total Nilai < 0, maka sentimen negatif
* Total Nilai = 0, maka sentimen netral

Dari plot data diatas polarity paling banyak terjadi pada 0 dan 1. Sehingga sentiment yang dominan adalah positif.



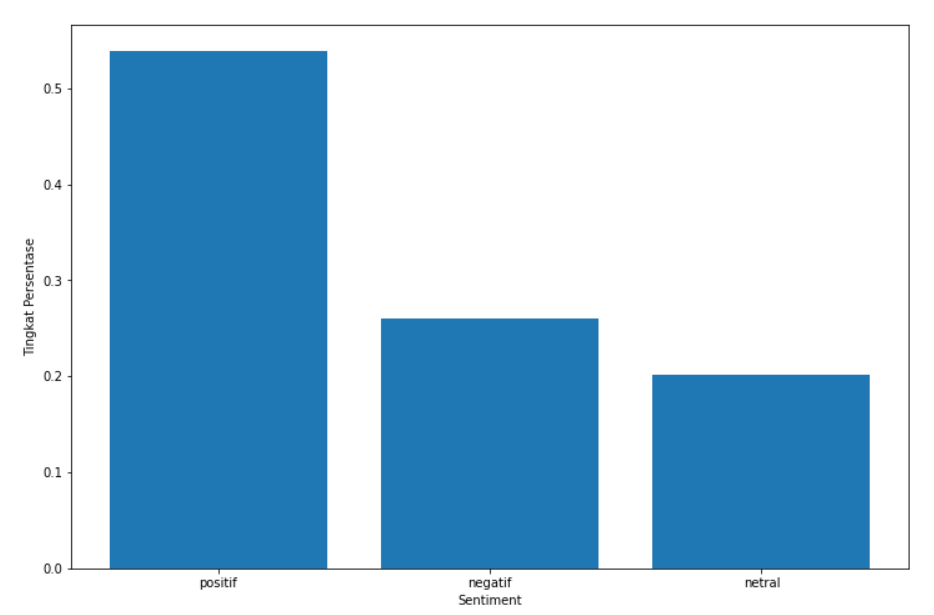
Dilihat dari gambar hasil diatas, rata – rata polarity adalah 0.6160877513711152 , ini berarti duntuk 1094 tweet kecenderungan tanggapan masyarakat positif terhadap bahasan Omnibus Law.

Kemudian langkah selanjutnya mengubah setiap sentiment sesuai dengan rules. Kemudian simpan ke dalam database.

|  |
| --- |
| # ubah sentiment  sentiments = []  for h in hasil:  if h > 0:  sentiments.append(1)  elif h < 0 :  sentiments.append(0)  else :  sentiments.append(2) |

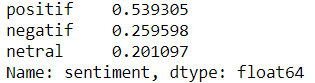
**Tingkat Persentase Sentiment**

Untuk mengetahui ini ambil sentiment dari database, ubah nilai sentiment menjadi nilai stringnya (misal 1 menjadi “positif”) kemudian hitung setiap nilainya, lalu plot. Maka akan dihasilkan :



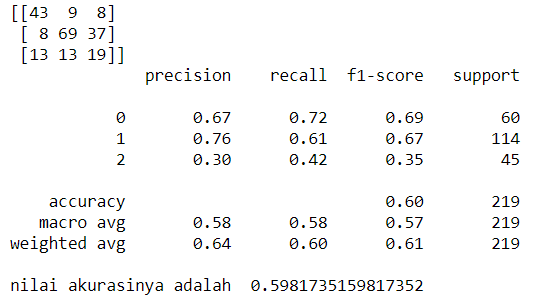
**Gambar 3. Tingkat Persentase Sentiment**

Dengan nilai presentase :

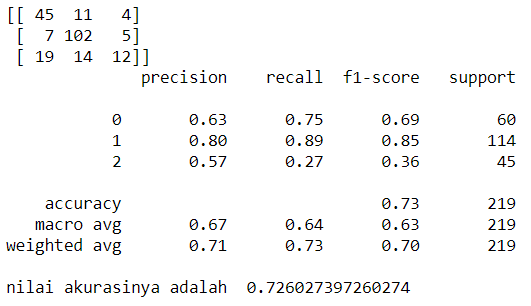


**Sentimen Analisis – Perbandingan Metode SVM, KNN, Naïve-Bayes**

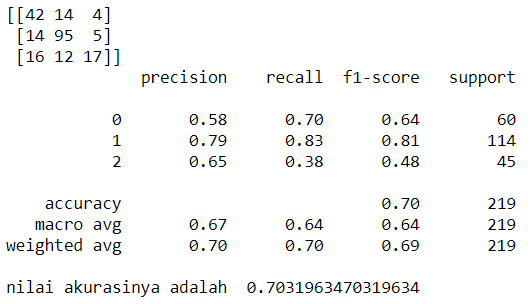
Data yang digunakan disini merupakan data yang telah diolah melalui pendekatan polarity. Dengan membagi data train sebesar 80 % dan test sebesar 20 %. Setelah dilakukan prediksi maka didapatkan akurasi dari setiap metode sebagai berikut :

**Naïve-Bayes** :  


**SVM :**



**KNN** :

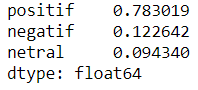


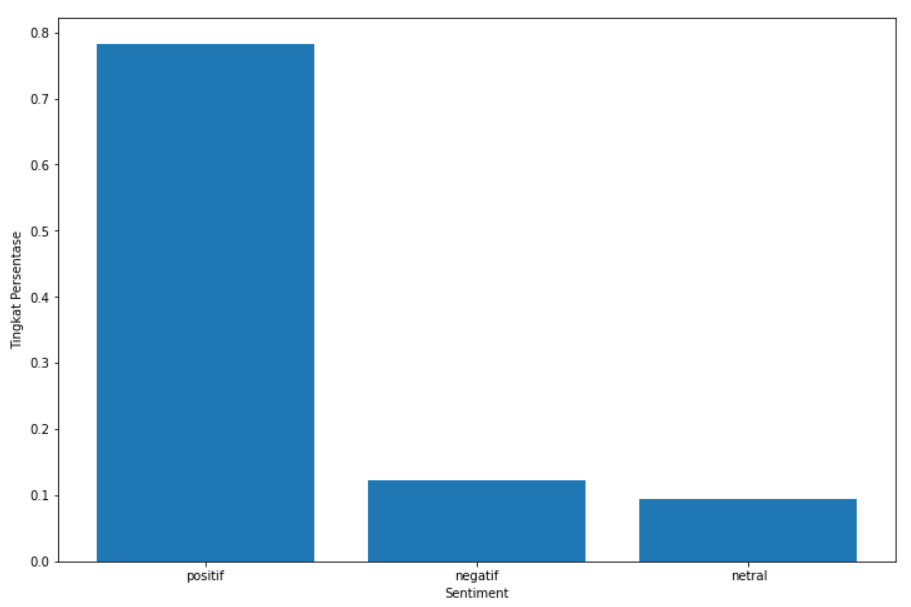
Nilai akurasi teringgi diperoleh oleh metode SVM yaitu 0.726027397260274. Kemudian selanjutnya akan dicoba pada data yang diambil pada 29 November 2020 dari pukul 00:00 hingga 12:20 WITA menggunakan metode SVM.

|  |
| --- |
| from laksmanayudha22\_tugas13 import DataHandler, query, database, jumlah\_tweet, twitter\_api  data\_handler = DataHandler(database)  DataHandler.date\_since = "2020-11-29"  data\_handler.get\_data(twitter\_api, query, jumlah\_tweet)  data\_handler.save\_sql()  data\_handler.clean\_data()  # ambil data  conn = sqlite3.connect('laksmanayudha22\_final.db')  cursor = conn.cursor()  query = """ SELECT cleantweet FROM tweet WHERE createddate > '2020-11-29'; """  cursor.execute(query)  conn.commit()  data29 = cursor.fetchall()  cursor.close()  conn.close()  # multidimensional ke single dimension  data29 = list(itertools.chain(\*data29))  # ML  vectorizer = TfidfVectorizer (max\_features=320)  x = vectorizer.fit\_transform(df['cleantweet']).toarray()  # pisahkan data feature dan target  y = df.iloc[:,-1]  # data test  x\_test = vectorizer.fit\_transform(pd.Series(data29)).toarray()  # prediksi  model\_s = svm.SVC(kernel='linear')  model\_s.fit(x,y)  y\_preds = model\_s.predict(x\_test)  dff = dff.replace({1:'positif', 0:'negatif', 2:'netral'})  dff.value\_counts(normalize=True)  # bar plot  data\_plot = dff.value\_counts(normalize=True)  label = data\_plot.index  freq = data\_plot.values  # plot data  fig, ax = plt.subplots(figsize=[12, 8])  ax.bar(label, freq)  #set label  ax.set\_xlabel('Sentiment')  ax.set\_ylabel('Tingkat Persentase') |

Setelah dilakukan pengambilan data dan dilakukan prediksi terhadap data tersebut dan dilakukan plot, maka diperoleh :

Persentase sentiment :





**Gambar 4. Tingkat Persentase Sentiment Pada 29 November 2020**

Dilihat dari gambar diatas, sampai saata ini sentiment masyarakat mengenai Omnibus Law masih cenderung ke arah yang positif.