

Antusiasme Warganet Indonesia Terhadap Produk Marvel Cinematic Universe pada Tahun 2018

Nama : Budi Gunawan
NIM : 1608107010009

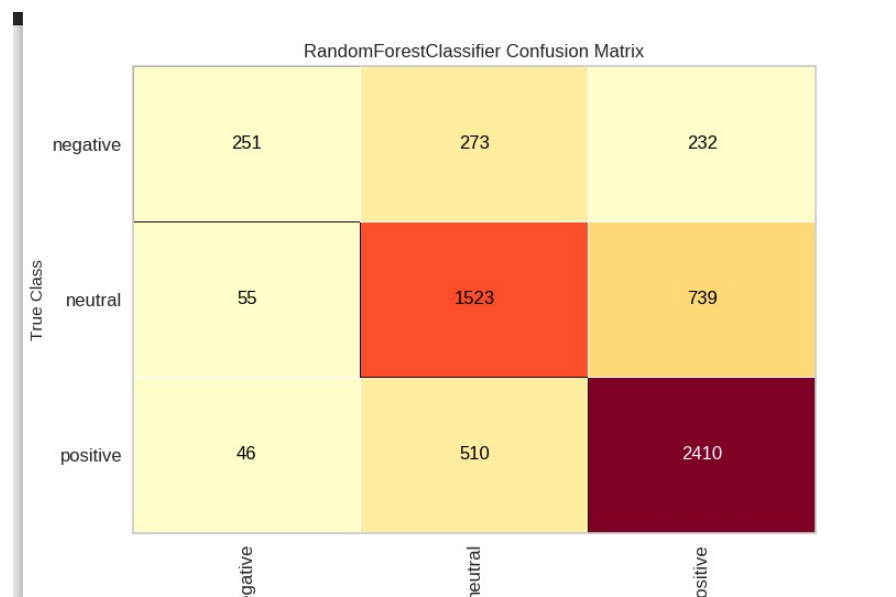
1. Preprocessing

Data twitter dengan kata kunci “marvel” dan bahasa yang difilter merupakan bahasa Indonesia berhasil di-crawling sejumlah 30191 tweet diolah dengan terlebih dahulu menghilangkan link dan gambar, kemudian juga karakter-karakter khusus yang tidak berguna untuk analisis. Kemudian barulah dinormalisasi dengan mengubah tulisan tweet menjadi huruf kecil semua dan menormalisasi kata-kata alay dalam bahasa inggris dan juga kata-kata alay dalam bahasa indonesia agar mudah dideteksi bahasanya dan terutama sentimennya. Dalam hal cleaning ini, saya menggunakan Optimus Psyspark. Setelah benar-benar bersih, barulah ditentukan label sentimen untuk tiap tweet berdasarkan polaritynya di mana jika di atas 0 maka positif, di bawah 0 maka negatif dan 0 maka netral, dengan menggunakan Polyglot karena mampu mendeteksi banyak bahasa. Berikut ditampilkan 10 data teratas :

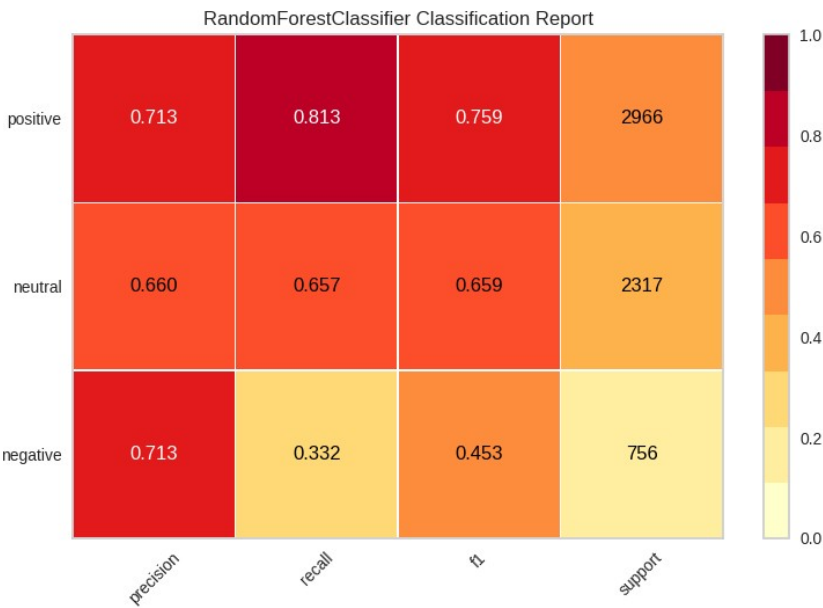
| | date | username | tweet | hashtags | likes_count | sentiment |
|---|------------|----------------|---|----------|-------------|-----------|
| 0 | 2018-12-30 | pipimerahku | bucin gampang banget cry boroboro sekuat capta... | [] | 0 | neutral |
| 1 | 2018-12-30 | iqmalhakimie | haahahhah pasal marvel dc sebagainye haihh | [] | 0 | neutral |
| 2 | 2018-12-30 | skaisahh | hiiiiiii wish closer back coll eeee menyesal r... | [] | 1 | negative |
| 3 | 2018-12-30 | hendryhash5 | penonton indonesia mah suka film superhero act... | [] | 0 | positive |
| 4 | 2018-12-30 | ahluksyorga | layan cerita2 marvel kat cinema marvel nih nak... | [] | 0 | negative |
| 5 | 2018-12-30 | shmxnlight6 | teman menonton aquaman pas filmnya kelar credi... | [] | 0 | positive |
| 6 | 2018-12-30 | tiakomalasari | pemeran munpuni film cerita bagus dipaksain di... | [] | 0 | positive |
| 7 | 2018-12-30 | tiakomalasari | seleru suka film ant man venom dibandingkan fi... | [] | 0 | positive |
| 8 | 2018-12-30 | uzer26 | shahmerizzat game marvel champion kat phone me... | [] | 0 | positive |
| 9 | 2018-12-30 | supermarcheeee | spiderman versi india diterima marvel jepang i... | [] | 0 | neutral |

2. Modelling

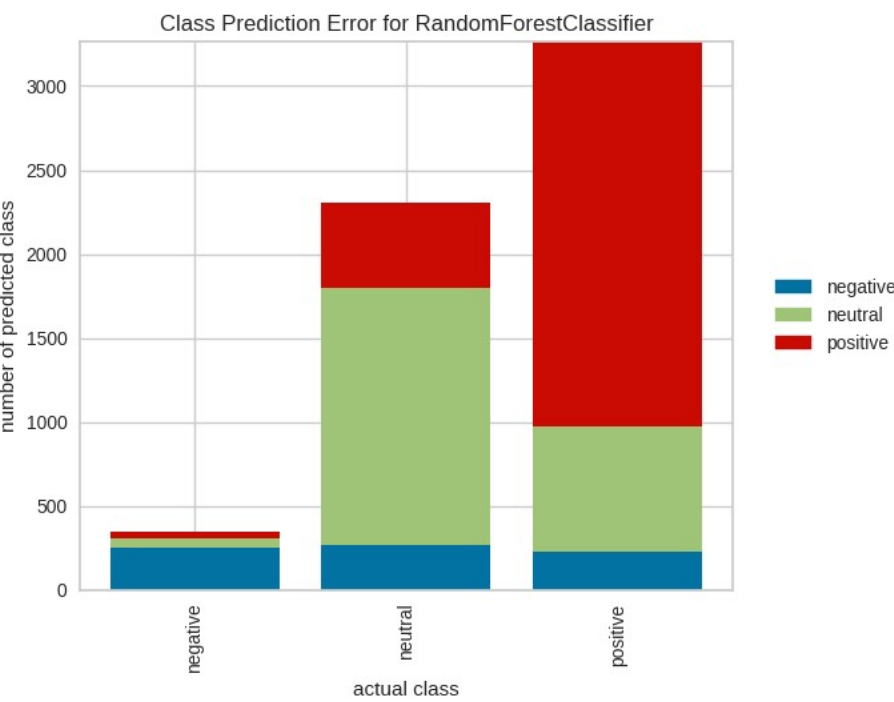
Setelah mendapatkan data bersih dan labelnya, kemudia data tweet diekstrak fiturnya, dalam project ini menggunakan metode TF-IDF kemudian dibangun modelnya dengan RandomForest Classifier dan data train 80% yang diuji dengan data test 20% sehingga mendapatkan hasil confusion matrix sebagai berikut :



dan data klasifikasinya sebagai berikut :



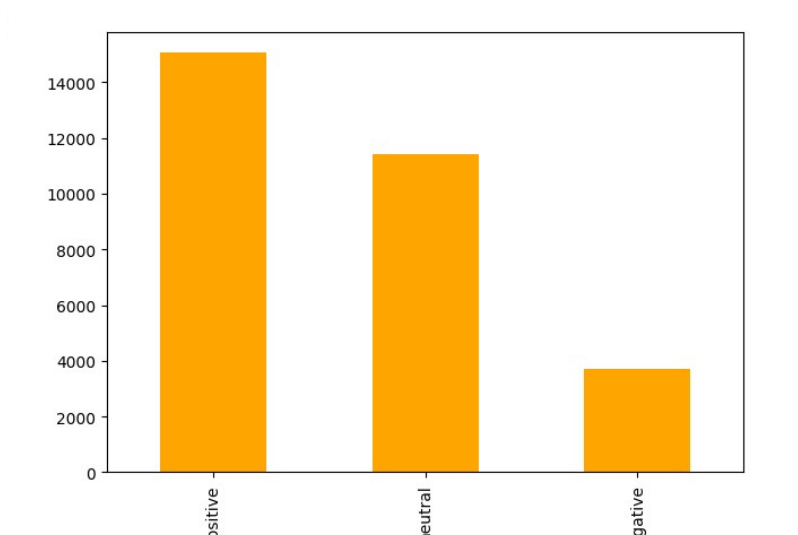
Serta prediksi error pada tiap kelas positif, netral, dan negatif adalah sebagai berikut :



3. Visualisasi

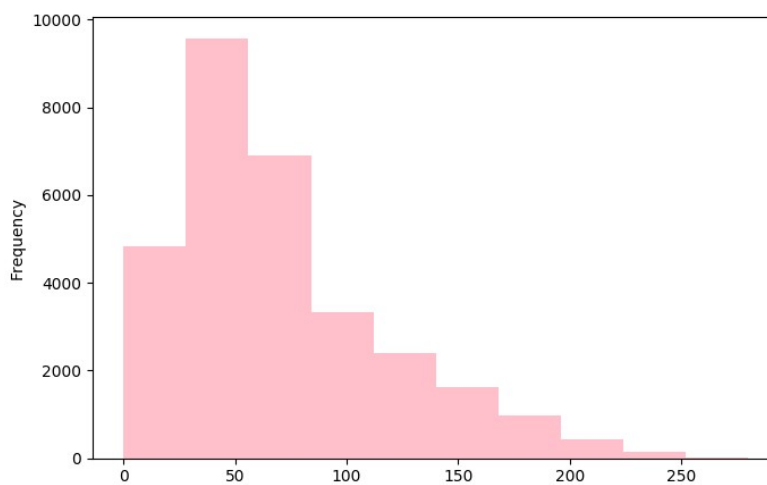
Dari data yang didapatkan, ada banyak hal yang dapat divisualisasikan dan dianalisis, berikut hasil yang didapatkan :

a. Persebaran sentimen



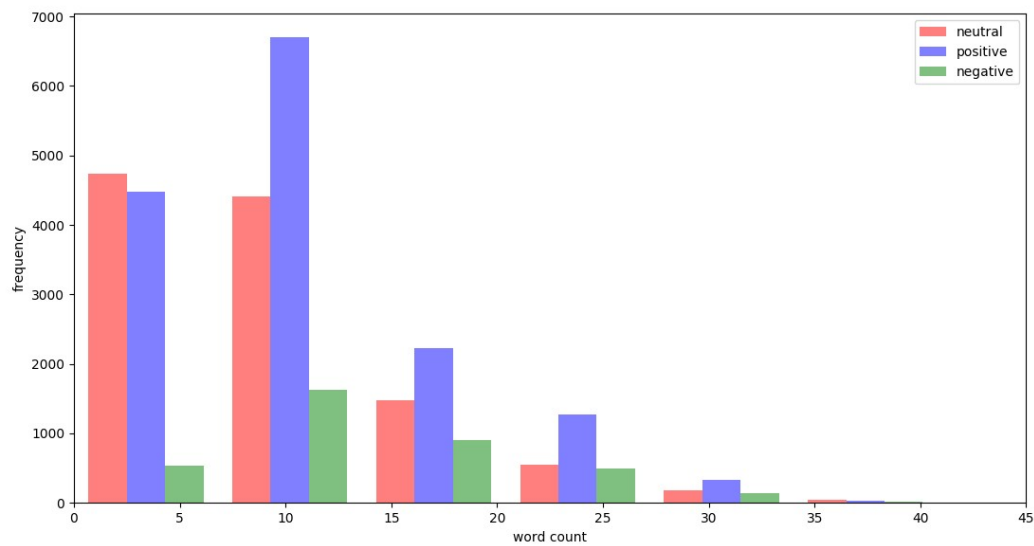
Dari data di atas dapat diketahui bahwa sentimen positif warganet twitter Indonesia lebih banyak dibandingkan dengan sentimen negatifnya

b. Frekuensi jumlah karakter teks twitter



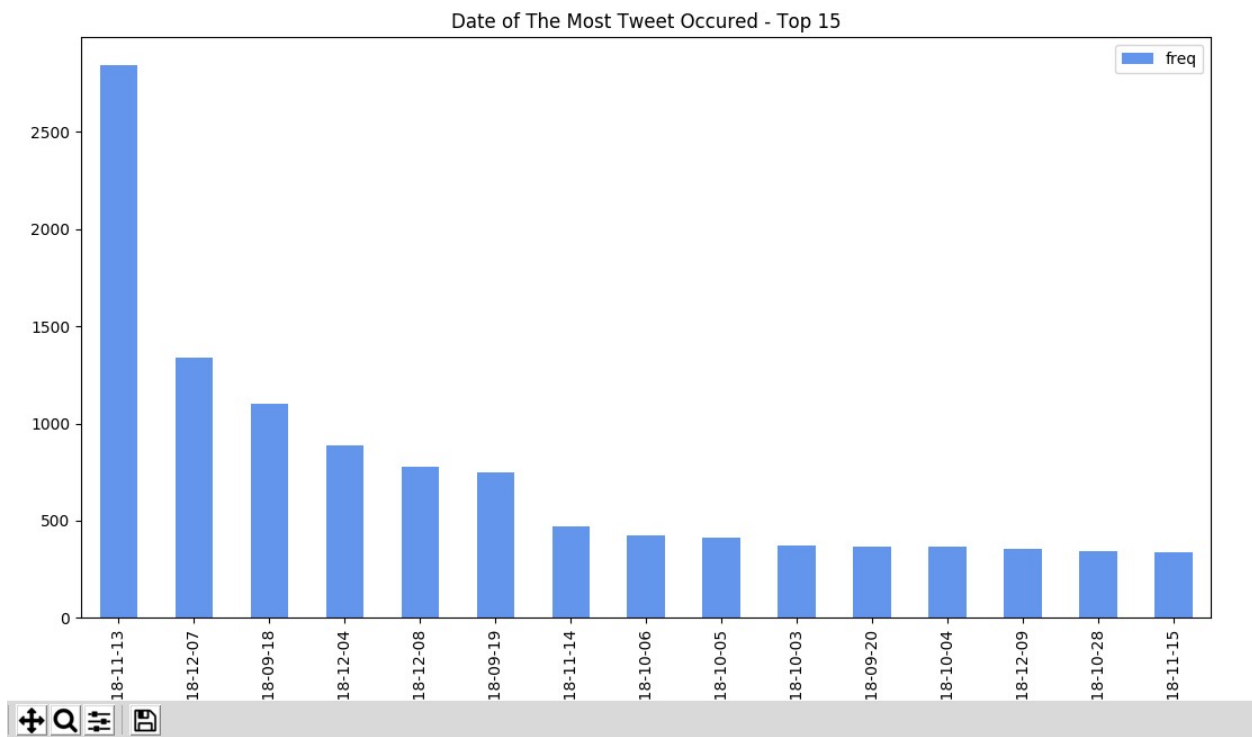
Dari persebaran data di atas dapat diketahui bahwa jumlah karakter teks yang sering digunakan seringkali singkat di bawah 100 karakter

c. Frekuensi jumlah kata yang digunakan per sentimen



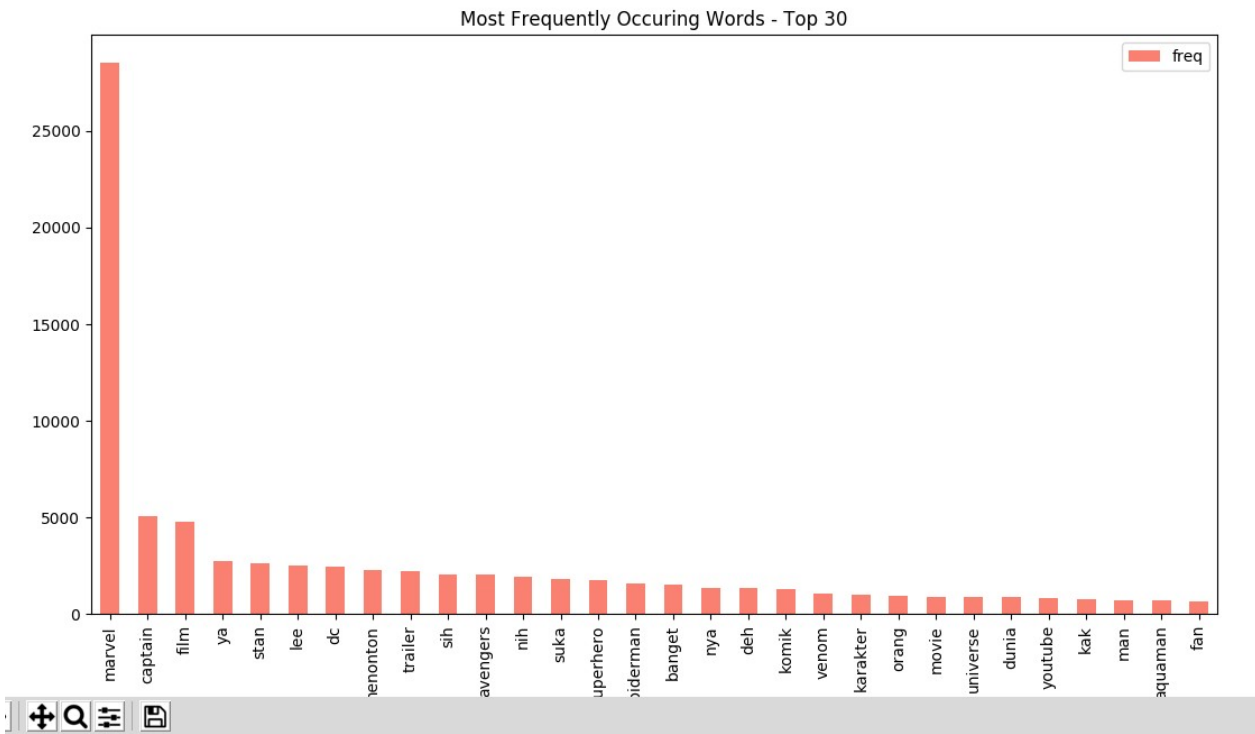
Sentimen positif banyak ditemukan di tweet dengan 10 kata, negatif banyak di tweet dengan 5 kata ke bawah

d. Tanggal dengan jumlah tweet marvel terbanyak

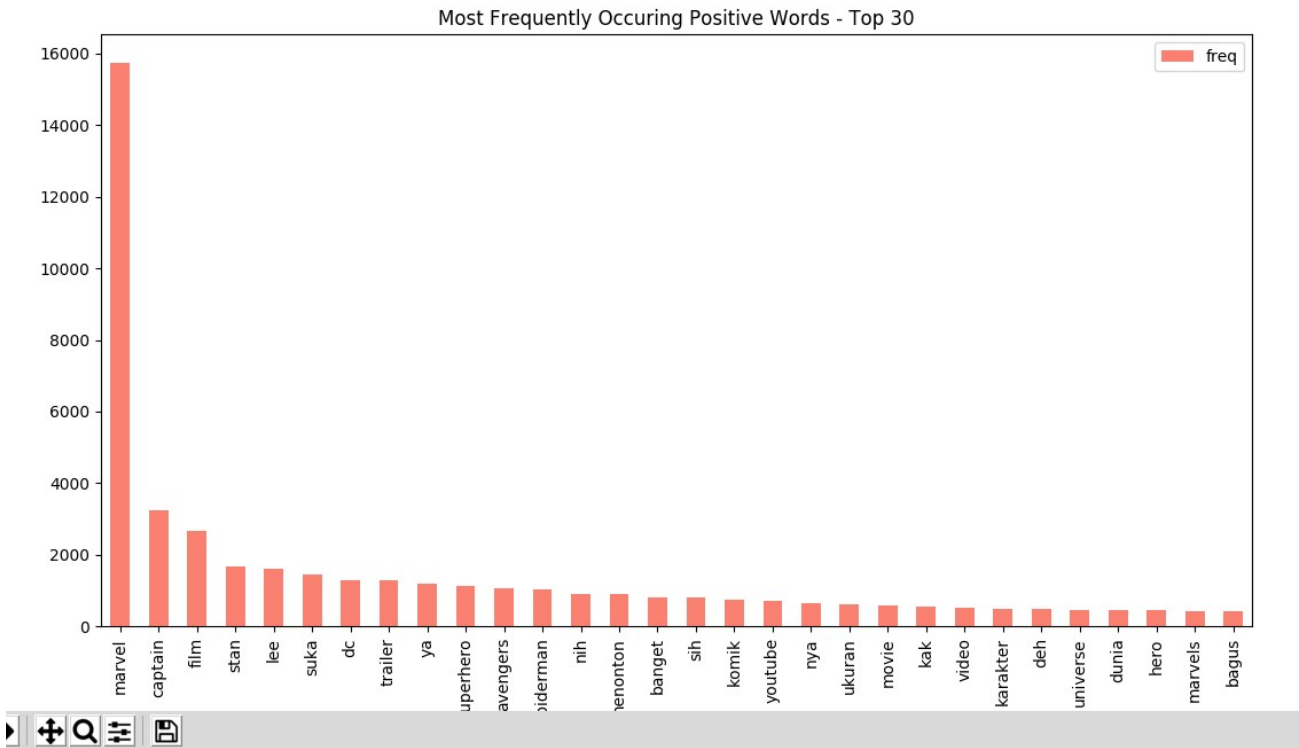


Dapat diketahui bahwa tanggal 13 November 2018 merupakan tweet tentang marvel terbanyak

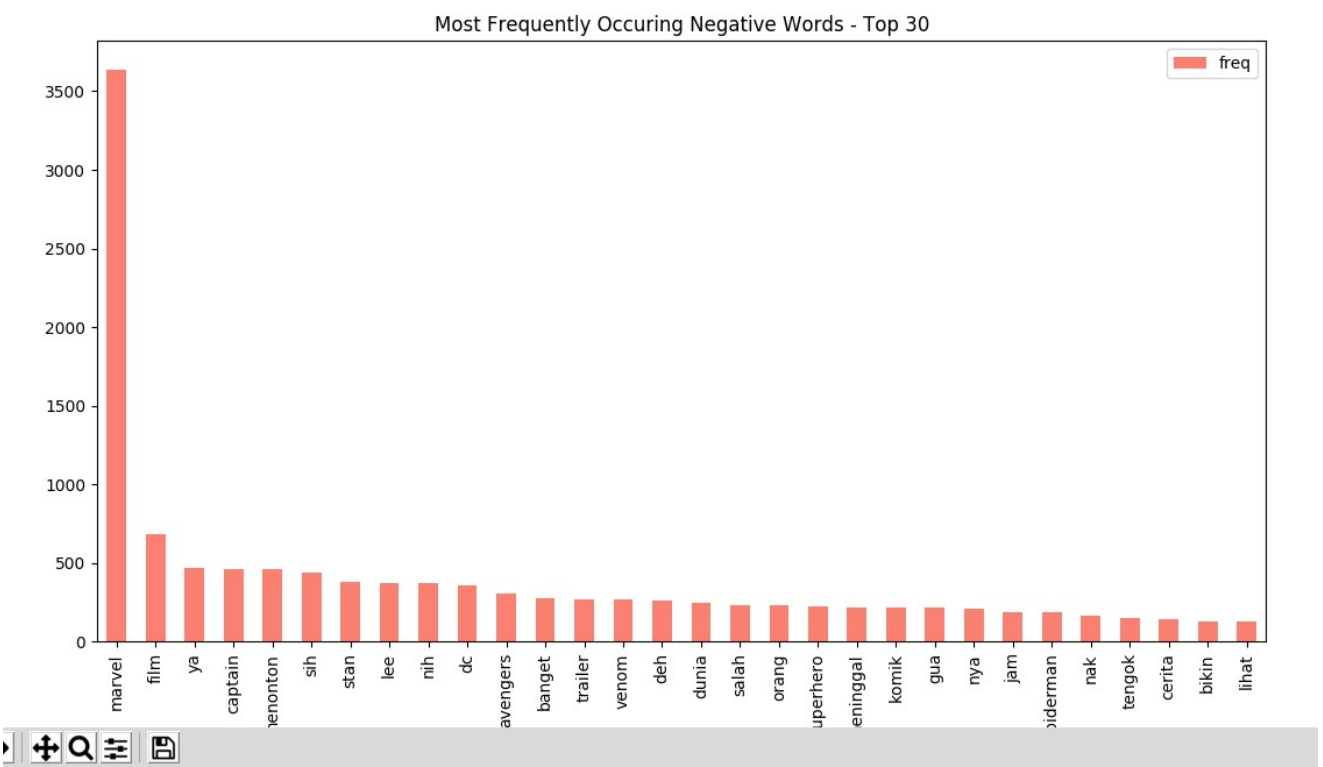
e. Kata yang paling sering ditemukan di tweet marvel



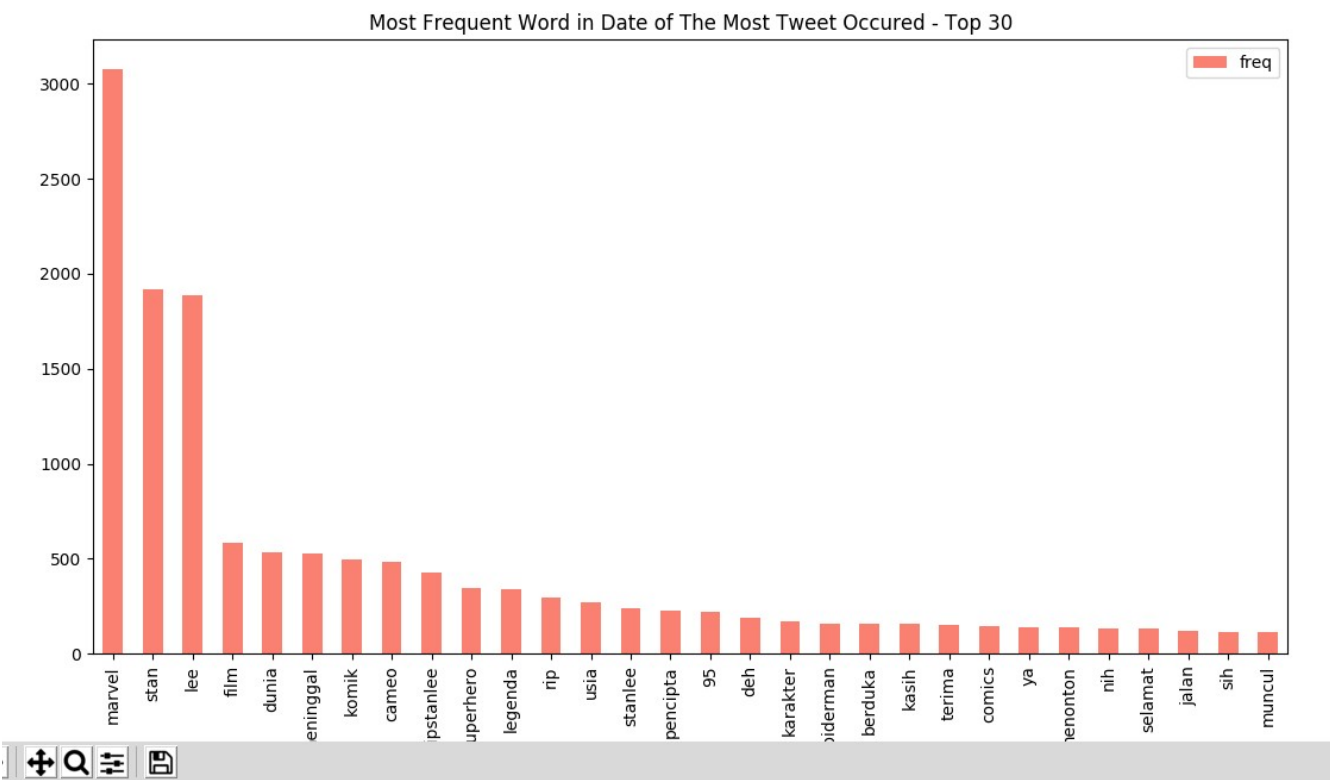
f. Kata positif yang paling sering ditemukan di tweet marvel



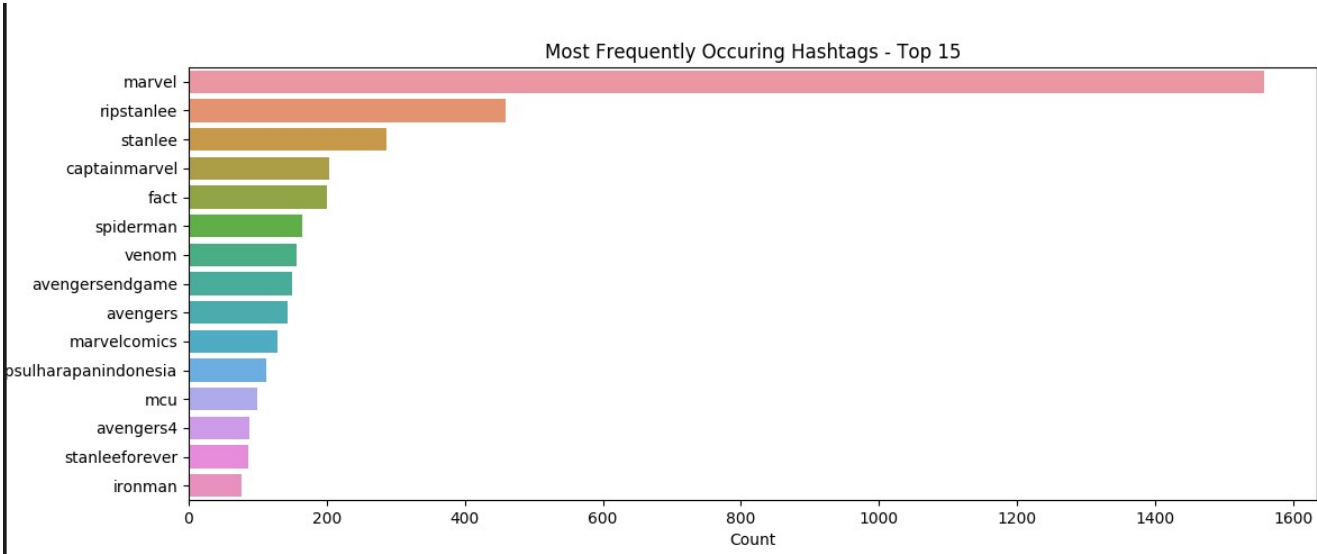
g. Kata negatif yang paling sering ditemukan di tweet marvel



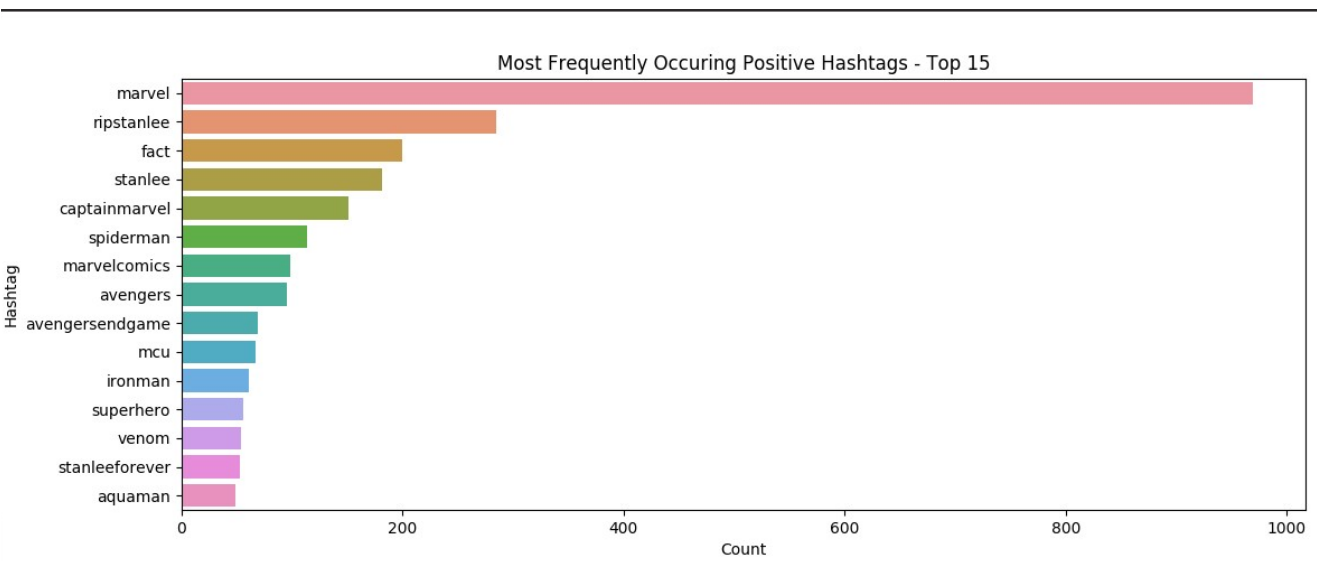
h. Kata yang sering muncul di tanggal 13 November 2018 (tweet terbanyak)



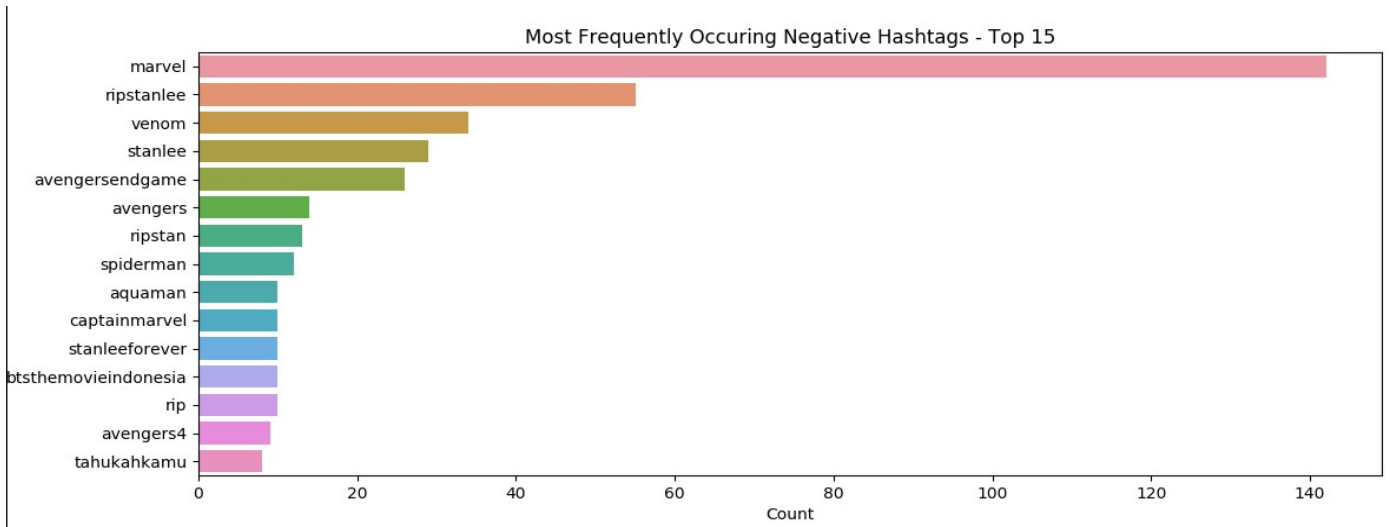
I. Hashtag twitter tentang marvel terbanyak



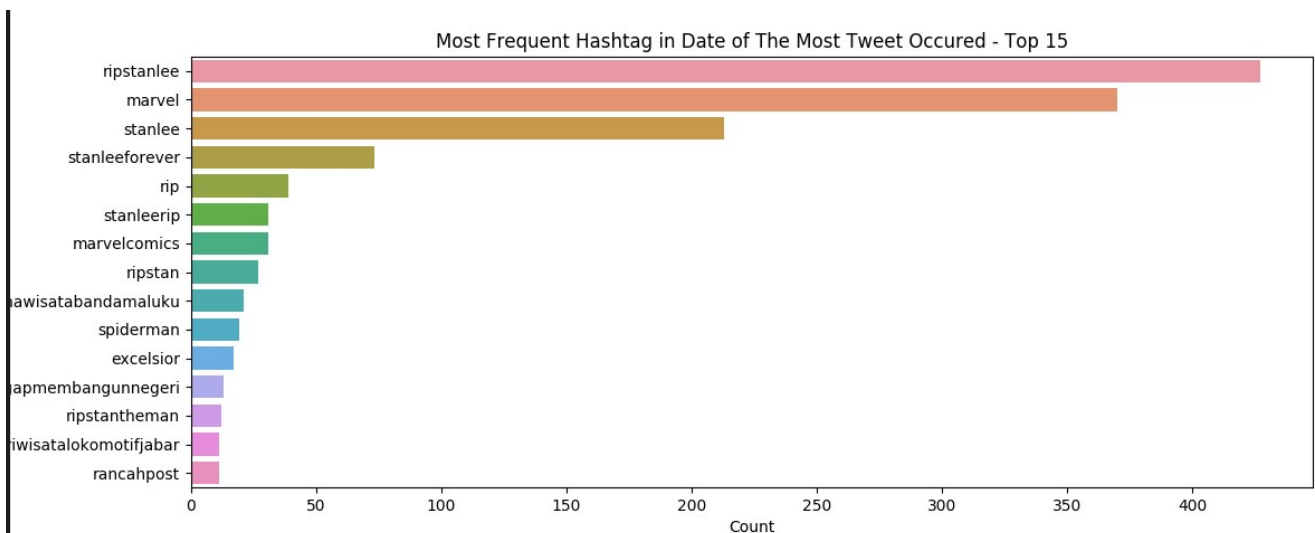
j. Hashtag positif tentang marvel terbanyak



k. Hashtag negatif tentang marvel terbanyak

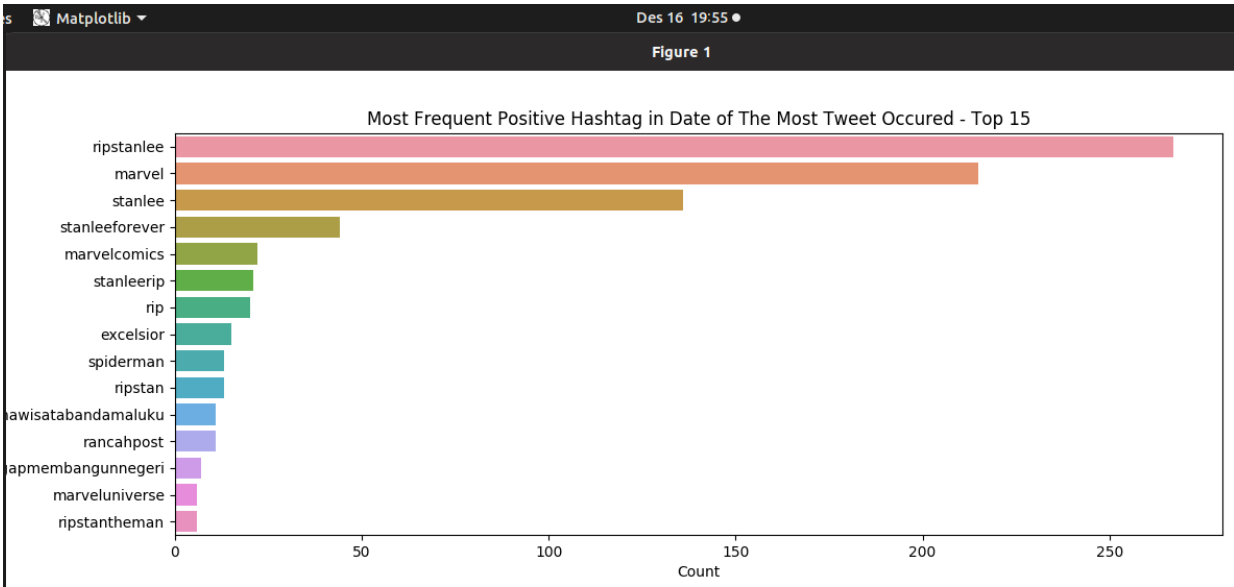


l. Hashtag terbanyak di tanggal 13 November 2018 (tweet marvel terbanyak)



Dari data di atas dapat diketahui bahwa mengapa tweet terbanyak pada tanggal 13 November 2018, yaitu karena bertepatan dengan wafatnya pencipta marvel yaitu Stan Lee.

j. Hashtag Positif terbanyak di tanggal 13 November 2018 (tweet marvel terbanyak)



k. Hashtag Negatif terbanyak di tanggal 13 November 2018 (tweet marvel terbanyak)

