

ALGORITMA ANALISIS SENTIMEN YANG BERFOKUSKAN EMOJI

Vishnupriya A/P Ravishankar

Prof. Dr. Shahrul Azman Mohd Noah

Fakulti Teknologi dan Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Kajian ini adalah berdasarkan analisis sentimen iaitu untuk mengenal pasti sentimen sesebuah teks media sosial yang mengandungi emoji. Selain itu, algoritma ini berfokuskan kepada penggunaan emoji yang digunakan semasa meluahkan perasaan atau pendapat. Hal ini kerana, emoji memainkan peranan penting dalam konteks sesebuah ayat / teks. Malah, emoji ini sukar untuk dikenalpasti dengan pengaturcaraan analisis sentimen yang biasa. Kesukaran ini dapat ditangani dengan melalui beberapa teknik. Teknik - teknik yang boleh digunakan adalah pengelasan pembelajaran mesin. Pembangunan algoritma ini adalah berasaskan kaedah pembelajaran mesin. Sentimen ini terbahagi kepada tiga, iaitu positif, neutral dan negatif. Selepas menganalisis teks, algoritma ini menghasilkan output yang menunjukkan sentimen teks tersebut. Algoritma ini dibangunkan untuk kemudahan pengurusan perniagaan, individu terkenal dan pelbagai lagi bidang. Dengan menggunakan algoritma ini, mereka akan mendapat pendekatan yang terjamin daripada para masyarakat. Maklumat seperti ini sangat berguna untuk pembangunan sesuatu organisasi mahupun negara.

1 PENGENALAN

Teknologi Web semakin meningkat dari masa ke semasa. Oleh kerana arus modenisasi ini, bilangan orang yang meluahkan perasaan dan pendapat melalui web semakin meningkat secara drastik. Maklumat ini berguna untuk pihak awam dan juga swasta. Maklumat dan pendapat sangat berguna dalam pengurusan perniagaan, kerajaan dan juga individu. Melalui maklumat-maklumat yang diperolehi pihak tertentu dapat mengetahui sentimen pendapat masyarakat terhadap mereka. Hal ini sangat bermanfaat untuk kejayaan mereka dalam masa akan datang.

Pada zaman dahulu, pihak awam dan swasta perlu bersusah payah untuk mengetahui pendapat masyarakat serata. Hal ini disebabkan, pihak awam dan swasta tidak mempunyai komunikasi yang efektif dengan masyarakat. Pendapat - pendapat masyarakat diambil secara manual. Contohnya, jika pengurusan jenama X ingin mengetahui pendapat masyarakat tentang produk mereka, mereka harus menjalankan kajian atau sesi soal selidik. Sebuah kumpulan harus membuat kajian dan soal-selidik. Kumpulan penyelidik tersebut harus pergi ke tempat-tempat tertentu untuk mengetahui pendapat masyarakat. Selain itu, data yang diperoleh daripada pengumpulan soal selidik hanya sedikit. Pendapat keseluruhan tidak dapat diketahui kerana sesi soal selidik hanya dijalankan kepada segelintir masyarakat.

Pengetahuan tentang pendapat masyarakat tidak lagi menjadi masalah dengan pembangunan web pada masa kini. Penggunaan web oleh pihak masyarakat meningkat sejak 20 tahun yang lepas. Pada zaman sekarang, semakin ramai orang meluahkan perasaan dan pendapat mereka terhadap produk, servis, individu, atau peristiwa dalam blog atau media sosial mereka. Data ini sangat berguna untuk pembangunan organisasi. Data ini semua dikumpulkan dan dianalisis dengan cara analisis sentimen.

Peningkatan pendapat telah memberi peluang untuk bidang sains data dan analisis menjadi popular. Kaedah kecerdasan pengkomputeraan terbukti untuk menjadi alat persaingan yang penting dalam industri hari ini. Sebagai contoh, dalam membuat analisis sentimen untuk sesuatu perniagaan, corak pendapat masyarakat boleh ditimbang untuk memahami pelanggan, meningkatkan jualan dan pemasaran. Proses analisis sentimen boleh mengenal pasti corak dalam sesebuah data.

Data yang terdapat dalam internet boleh diproses kepada bentuk yang lebih senang untuk dibaca oleh proses analisis sentimen. Proses ini dapat mengenal pasti sentimen sesuatu teks. Terdapat tiga jenis sentimen utama iaitu, positif, neutral dan juga negatif.

2 PENYATAAN MASALAH

Analisis sentimen adalah topik kajian yang hangat sekarang. Teknik ini digunakan untuk mengenal pasti prestasi sesebuah perniagaan, produk atau sentimen sesebuah maklumat disebarkan, positif dan negatif. Analisis sentimen juga digunakan sebagai domain untuk kewartawanan untuk mengetahui topik yang digemari oleh masyarakat. Walaupun dunia aplikasi analisis sentimen ini besar, penggunaan analisis sentimen ini terhad kepada teks sahaja. Hal ini mungkin memberi konklusi yang tidak tepat dan salah kerana kegunaan alat atau ekspresi yang lain seperti emoji.

Pengguna media sosial seperti *Twitter* dan *Facebook* banyak menggunakan emoji. Terdapat impak yang besar kepada *tweets* dan pendapat yang diluahkan di media sosial dengan penggunaan emoji. Bukan sahaja teks, emoji juga harus diambil kira dalam analisis sentimen. Pada masa sekarang, tidak banyak model atau proses sentimen yang mengambil kira emoji untuk membuat analisis sentimen.

3 OBJEKTIF KAJIAN

Terdapat banyak model yang boleh menentukan sentimen sesuatu teks, sama ada positif, neutral atau negatif. Tetapi tidak banyak model berfungsi untuk mengecam dan mengenal pasti konteks sesuatu emoji. Dalam sesuatu teks yang mengandungi penggunaan emoji, emoji juga harus diambil kira untuk memahami konteks teks tersebut dengan keseluruhan. Oleh itu, model analisis sentimen yang berfungsi untuk mengenal pasti sentimen emoji harus dibangunkan. Sehubungan dengan itu **matlamat utama kajian ini ialah untuk mengelaskan polariti sentimen teks yang mengandungi emoji.**

Bagi mencapai matlamat ini, objektif berikut digariskan:

- a. Mengenalpasti kekutuban emoji yang digunakan dalam teks media sosial.
- b. Membina set data ujian melibatkan teks yang mengandungi emoji.

- c. Menggunakan algoritma pembelajaran mesin yang sesuai untuk mengelaskan teks yang mengandungi emoji.

4 METOD KAJIAN

Terdapat dua jenis metodologi yang boleh digunakan untuk proses analisis sentimen. Pendekatan tak terselia dan pendekatan terselia adalah dua metod yang paling digunakan untuk proses ini. Algoritma analisis sentimen ini menggunakan pendekatan terselia.

Algoritma analisis sentimen yang berfokuskan emoji ini menggunakan pendekatan terselia. Pendekatan terselia adalah algoritma pembelajaran mesin yang digunakan untuk mendapat konklusi selepas melatih mesin dengan set data. Pendekatan terselia menggunakan teknik pembelajaran mesin khususnya melibatkan penghasilan pengelas yang berciri untuk mentafsir dan mengenali sentimen teks. Penghasilan pengelas yang tepat memerlukan data untuk melatih mesin untuk menimba ilmu. Pengelas yang boleh digunakan untuk pendekatan pembelajaran mesin ini adalah *Support Vector Machine (SVM)*, *Neural Network*, *Maximum Entropy* dan *Naïve Bayes*.

4.1 Fasa Perancangan

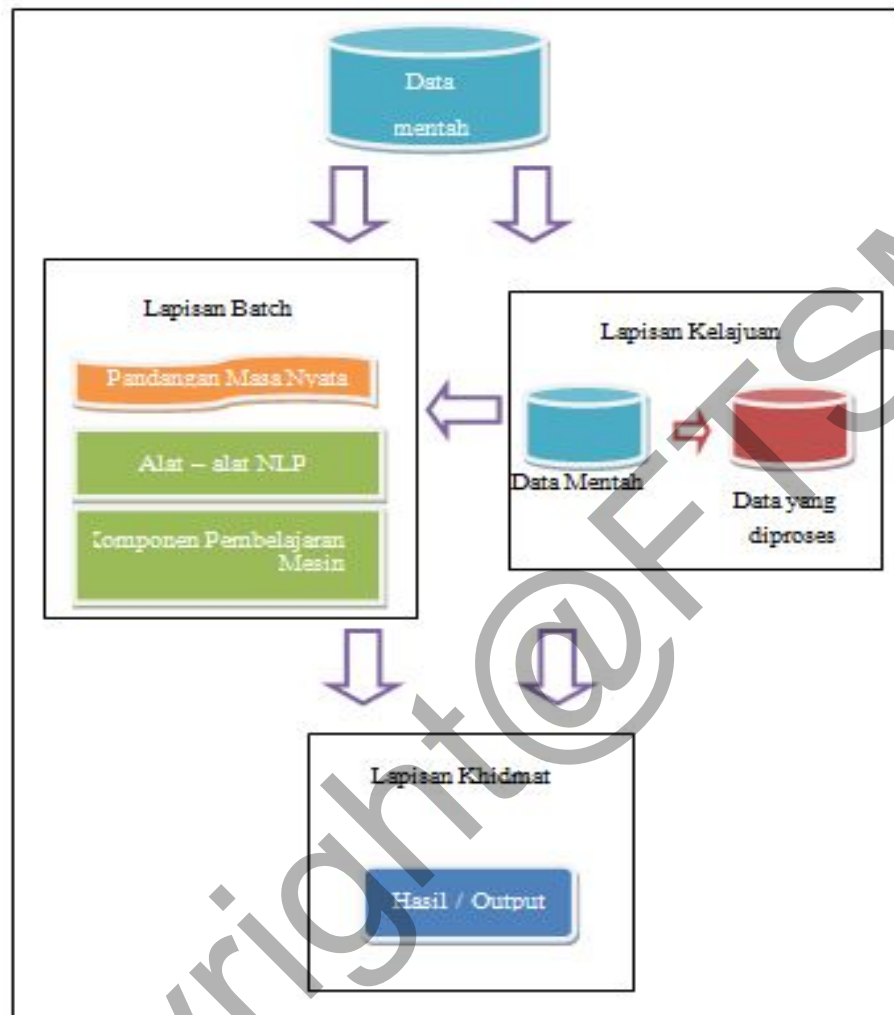
Fasa perancangan ini melibatkan mengenalpastian masalah, cadangan penyelesaian masalah, penetapan matlamat dan objektif kajian, penentuan skop dan rancangan awal pembangunan kajian. Langkah seterusnya adalah untuk mengkaji kesusasteraan kajian dengan terperinci. Kajian – kajian yang telah dibuat oleh pengkaji yang lain diteliti untuk pencetusan idea. Terdapat pelbagai jenis emoji yang digunakan dalam media sosial. Semua data dan jenis emoji dikenalpasti. Polariti dan sentimen setiap emoji yang digunakan dalam media sosial dikenalpasti untuk melancarkan fasa analisis.

4.2 Fasa Analisis

Fasa analisis ini melibatkan analisis dan tafsiran maklumat yang dikumpul. Analisis tentang kepentingan kajian dan kesesuaian kajian ini untuk penggunaan masa depan dinilai. Algoritma analisis sentimen yang berfokuskan emoji ini dipastikan berguna pada masa depan. Selain itu, keperluan perisian dan perkakasan juga dikenalpasti untuk pembangunan projek ini. Bahasa pengaturcaraan yang digunakan untuk algoritma analisis sentimen yang berfokuskan emoji adalah *Python 3*.

4.3 Fasa Reka Bentuk

Fasa ini merupakan fasa yang penting dalam kajian ini. Senibina tunggal tidak sesuai untuk merealisasikan kajian ini. Oleh itu, senibina Lambda (Marz dan Warren, 2015) diambil sebagai inspirasi. Senibina perisian ini dibahagi kepada 3 tahap yang berbeza iaitu, lapisan kelajuan, lapisan *batch* dan lapisan khidmat. Fungsi setiap lapisan ini diubahsuai untuk menepati objektif kajian ini. Rajah 1 menunjukkan senibina kajian ini.



Rajah 1

A Lapisan Kelajuan

- (i) Data mentah disimpan dalam fail jenis .txt.
- (ii) Emoji dan ikon emosi dalam data mentah ditukar kepada simbol dan perkataan yang boleh dibaca oleh mesin
- (iii) Jenis fail ditukar pengekodan utf-8 dan disimpan dalam fail .csv.

B Lapisan *Batch*

- **Bahagian Pertama – Penentuan Sentimen Teks**

- (i) Data daripada lapisan kelajuan dibaca
- (ii) Setiap perkataan dan simbol dikenal pasti (*Tokenized*)
- (iii) Emoji / Ikon emosi dikenal pasti
- (iv) Polariti emoji / ikon emosi dalam setiap teks didapati daripada pangkalan data emoji sentimen yang disediakan terlebih dahulu
- (v) Teks melalui fasa pra pemprosesan
- (vi) Polariti teks didapati melalui *Vader Lexicon*, *Sentimen Intensity Analyzer* dan NLTK.
- (vii) Polariti emoji dan polariti teks ditambah untuk mendapatkan polariti akhir teks.
- (viii) Polariti akhir teks digunakan untuk menentukan sentimen setiap teks tersebut

Jadual 1 – Skor polariti dan sentimen

Polariti	Sentimen
Kurang atau bersamaan dengan -0.1	Negatif
Lebih daripada -0.1 dan kurang daripada 0.1	Neutral
Lebih atau bersamaan dengan 0.1	Positif

- **Bahagian Kedua – Penghasilan set data ujian dan set data latihan**

- (i) Simbol emoji dan ikon emosi dikenal pasti

- (ii) Teks melalui pra-pemrosesan data
- (iii) Perkataan – perkataan yang tinggal ditukarkan kepada katar dasar.
(*word stemming and word lemmatization*)
- (iv) Setiap perkataan dibahagi kepada tiga bahagian iaitu, kata sifat (*adjective*), kata kerja (*verb*) dan kata keterangan (*adverb*)

Jadual 2 Jenis perkataan dan peta tag

Jenis Perkataan	Peta Tag (<i>Tag Map</i>)
Kata sifat (<i>adjective</i>)	J
Kata kerja (<i>verb</i>)	V
Kata keterangan (<i>adverb</i>)	R

- (v) Perkataan – perkataan yang diproses dan emoji dimasukkan ke dalam satu *string* untuk setiap teks untuk melatih dan menguji.

- **Bahagian Ketiga – Penghasilan Model Pengelas**

- (i) Data set daripada bahagian kedua dibahagi untuk melatih dan menguji mesin

Jadual 3 Jenis data dan peratusan

Jenis data set	Peratusan
Data set untuk melatih	70%
Data set untuk menguji	30%

- (ii) Pengekod label(*label encoder*) dilakukan pada data

- (iii) Seterusnya, vektorisasi perkataan (*word vectorization*) dilakukan menggunakan vektor TF-IDF
- (iv) Pengelas yang sesuai digunakan untuk melatih dan menguji data
- (v) Ketepatan (*accuracy*), ketepatan (*precision*) dan dapatan balik (*recall*) digunakan untuk mengetahui keberkesanan model pengelas
- (vi) Model pengelas disimpan menggunakan *Pickle*.

B Lapisan Khidmat

- (i) Dataset daripada lapisan Kelajuan digunakan untuk menganalisis sentimen
- (ii) Model pengelas daripada lapisan *Batch* digunakan untuk menentukan sentimen.
- (iii) Data dibuat pra-pemprosesan dan pemprosesan seperti *Word Stemming and Word Lemmatization*
- (iv) Seterusnya, data melalui proses pengekod label (*label encoder*) dan vektorisasi perkataan (*word vectorization*)
- (v) Model pengelas digunakan atas data yang diproses untuk menentukan sentimen
- (vi) Sentimen ditentukan dalam bentuk label. Label ini ditukar kepada perkataan untuk memudahkan kefahaman.

Jadual 4 Label dan sentimen

Label yang diberi oleh model	Sentimen
pengelas	
0	Negatif
1	Neutral
2	Positif

(vii) Teks dan juga sentimen setiap teks yang dijangka oleh model pengelas boleh dimuat turun

4.4 Fasa Pengujian

Algoritma analisis sentimen ini menggunakan *Naïve Bayes* untuk mengelaskan sentimen teks yang menggunakan emoji. Hal ini kerana, *Naïve Bayes* menunjukkan ketepatan yang tinggi berbanding dengan model pengelas yang lain dalam klasifikasi teks dan emoji. *Naïve Bayes* adalah model yang menggunakan kebarangkalian tetapi dalam masa yang sama mempunyai kebebasan yang tinggi untuk merumuskan andaian. Rekabentuk senibina kajian ini memerlukan pandangan masa nyata semasa menjalankan analisis sentimen. Ketepatan *Naïve Bayes* ini tepat dan juga model pengelas ini mengambil beberapa saat sahaja untuk mengelas data set yang besar.

5 HASIL KAJIAN

(0, 3167)	0.5089294070981548	(1559, 3597)	0.25623084161682674
(0, 2939)	0.4987153261541142	(1559, 3553)	0.2757577973169548
(0, 1685)	0.45676071703474996	(1559, 3143)	0.1446615061292319
(0, 1676)	0.34142891831113825	(1559, 2393)	0.17162405388748297
(0, 1249)	0.4087417560813958	(1559, 2056)	0.2458085938344187
(1, 4755)	0.2429305846160643	(1559, 1960)	0.2960317558403455
(1, 4404)	0.3353183237748329	(1559, 1804)	0.11077833603057861
(1, 4332)	0.30909833947742216	(1559, 682)	0.2960317558403455
(1, 3118)	0.18407662135461134	(1559, 356)	0.15935083934978728
(1, 2873)	0.3834893583961743	(1559, 317)	0.2395827419977168
(1, 2366)	0.28792950609863455	(1560, 4778)	0.44457537375176126
(1, 2155)	0.26170952180122387	(1560, 4170)	0.24860607435453763
(1, 1804)	0.11566799521194154	(1560, 4020)	0.3198591918018586
(1, 1314)	0.30909833947742216	(1560, 3859)	0.2876683219380392
(1, 719)	0.16862341004377485	(1560, 2941)	0.31181118490702237
(1, 476)	0.28792950609863455	(1560, 2483)	0.28126851028227323
(1, 412)	0.2708320653536076	(1560, 950)	0.34874391160709284
(1, 366)	0.2862035966900892	(1560, 101)	0.5074685064397358
(1, 175)	0.15855716524988223	(1561, 3936)	0.4102807750538735
(2, 4728)	0.29154831032206	(1561, 2542)	0.22588002394501155
(2, 3151)	0.3010463707608327	(1561, 2481)	0.3657702979769018
(2, 1959)	0.36005371177650813	(1561, 2395)	0.34505857379495275
(2, 1673)	0.24574265152871763	(1561, 1044)	0.4168462460361467
(2, 1650)	0.0995444815249653	(1561, 753)	0.49412117234535247
(2, 825)	0.22029583967926464	(1561, 545)	0.3314785730298448
:	:		

[2 2 2 ... 2 2 0]

Rajah 2 Sentimen teks

Rajah 2 menunjukkan hasil ujian menggunakan klasifier *Naïve Bayes*.

A) Skor *Naïve Bayes*

```
#Testing Naive Bayes Model
Naive.score(Test_X_Tfidf,test_y)

0.7272727272727273
```

Rajah 3 Skor *Naïve Bayes*

B) Matriks *Confusion*

```
#Formulation of confusion matrix
from sklearn.metrics import confusion_matrix
y_pred = Naive.fit(Train_X_Tfidf,train_y).predict(Test_X_Tfidf)
cm = confusion_matrix(test_y, y_pred)
print (cm)

[[30  6  4]
 [ 6 34  6]
 [ 6 14 48]]
```

Rajah 4 Matriks *Confusion*

Matriks *Confusion* adalah jadual yang sering digunakan untuk menggambarkan prestasi model klasifikasi. Dalam kajian ini, model pengelas *Naïve Bayes* digunakan untuk sekumpulan data ujian yang mana nilai yang dikenal pasti.

C) Ketepatan dan Dapatan Balik

```
#Precision and recall for each label
print("label precision recall")
for label in range(3):
    print(f"{label:5d} {precision(label, cm):9.3f} {recall(label, cm):6.3f}")

label precision recall
0      0.714  0.750
1      0.630  0.739
2      0.828  0.706
```

Rajah 5 Ketepatan dan Dapatan Balik

Label 0 menunjukkan ketepatan dan dapatan semula untuk sentimen yang negatif. Label 1 menunjukkan ketepatan dan dapatan semula untuk sentimen yang neutral. Label 2 menunjukkan ketepatan dan dapatan semula untuk sentimen yang positif.

```
#Precision and Recall of the Naive Bayes Model
print("precision total:", precision_macro_average(cm))

print("recall total:", recall_macro_average(cm))

precision total: 0.723833850270632
recall total: 0.7316709292412616
```

Rajah 6 Ketepatan dan Dapatan Balik

Cara mengira ketepatan (*precision*) dan dapatan semula (*recall*) adalah seperti berikut:-

- (i) Ketepatan (*precision*) =
- (ii) Dapatan semula (*recall*) =
 - a. Positif Benar (TP) - klasifikasi kelas adalah positif dan ramalan model pengelas adalah positif.
 - b. Negatif Benar (TN) - klasifikasi kelas adalah negatif dan ramalan model pengelas adalah negatif.
 - c. Positif Palsu (FP) - klasifikasi kelas adalah negatif tetapi ramalan model pengelas adalah positif.
 - d. Negatif Palsu (FN) - klasifikasi kelas adalah positif tetapi ramalan model pengelas adalah negatif.

D) Ketepatan Model

```
#Get the accuracy of the model
print("Naive Bayes Accuracy Score -> ", accuracy_score(predictions_NB, test_y)*100)

Naive Bayes Accuracy Score -> 72.72727272727273
```

Rajah 7 Ketepatan Model

Cara mengira ketepatan (*accuracy*)

(i) Ketepatan (*accuracy*) =

Selepas penghasilan model, periksa ketepatan menggunakan nilai sebenar dan nilai ramalan. Ketepatan *Naïve Bayes* dalam algoritma analisis sentimen yang berfokuskan emoji adalah 72.73%.

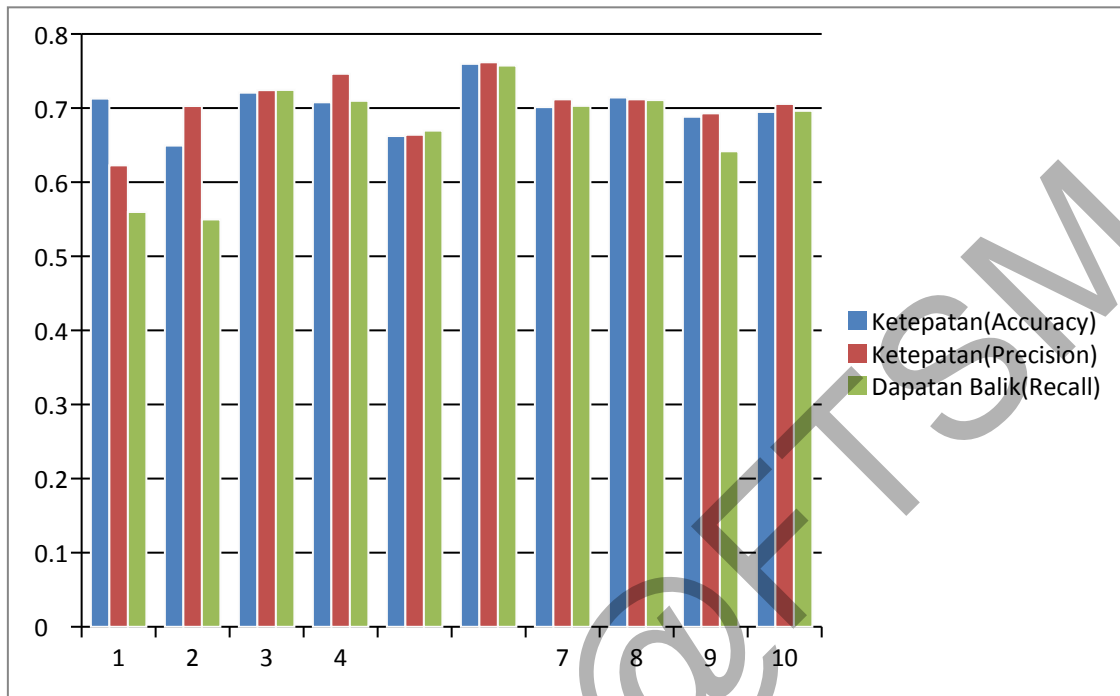
LIPATAN K RAWAK (*K- FOLD RANDOM TESTING*)

Jadual 5 Lipatan K-Rawak

	1	2	3	4	5
Iterasi					
Ketepatan (Accuracy)	0.7143	0.6494	0.7208	0.7078	0.6523
Ketepatan (Precision)	0.6227	0.7028	0.7242	0.7463	0.6614
Dapatan semula (Recall)	0.5598	0.6497	0.7245	0.7098	0.6696
	6	7	8	9	10
Iterasi					
Ketepatan (Accuracy)	0.7597	0.7013	0.7143	0.6883	0.6948
Ketepatan (Precision)	0.7618	0.7118	0.7118	0.6929	0.7056
Dapatan semula (Recall)	0.7574	0.7030	0.7108	0.6417	0.6963

Jadual 6 Purata Lipatan K-Rawak Naive Bayes

	Purata
Ketepatan (Accuracy)	0.7005
Ketepatan (Precision)	0.7041
Dapatan semula (Recall)	0.6849



Rajah 8 Visualisasi Lipatan K-Rawak Naive Bayes

Ketepatan algoritma ini diperoleh daripada perbandingan hasil lapisan *batch*, bahagian pertama dengan model yang dihasilkan, iaitu *Naive Bayes*. Model yang dihasilkan ini boleh digunakan untuk penggunaan masa depan jika teks yang hendak ditentukan sentimen ini daripada domain yang sama. Setiap domain mempunyai penentuan sentimen yang berbeza. Oleh itu, jika ingin menentu sentimen untuk teks yang mengandungi emoji daripada domain lain, model ini harus dilatih dan dihasilkan terlebih dahulu. Dengan penggunaan model setiap sentimen teks yang mengandungi emoji ini dapat ditentukan dengan betul dan tepat dalam masa yang sangat singkat. Walaupun jumlah teks yang hendak ditentukan sentimen ini besar dengan penggunaan model ini, masa yang diperuntukkan untuk penghasilan sentimen setiap teks singkat.

K-fold random testing adalah salah satu cara untuk mendapatkan purata ketepatan model yang dihasilkan. Ketepatan (*accuracy*) adalah 0.70, ketepatan (*precision*) adalah 0.70 dan dapatan balik (*recall*) untuk model *Naive Bayes* model adalah 0.68.


PERBANDINGAN ANALISIS SENTIMEN YANG MENGGUNAKAN EMOJI DAN TANPA MENGGUNAKAN EMOJI

Jadual 7 Perbandingan Analisis sentimen yang menggunakan emoji dan tanpa menggunakan emoji

	Ketepatan (Precision)	Dapatan Semula (Recall)	Ketepatan (Accuracy)
Analisis sentimen menggunakan emoji	0.7041	0.6849	0.7005
Analisis sentimen tanpa menggunakan emoji	0.5689	0.4411	0.5676

Ketepatan analisis sentimen yang menggunakan emoji menunjukkan ketepatan, 0.70, yang lebih tinggi berbanding dengan analisis sentimen tanpa menggunakan emoji, 0.57. Dengan ini analisis sentimen yang menggunakan emoji terbukti lebih tepat berbanding kepada analisis sentimen yang tanpa menggunakan emoji. Hal ini kerana, data daripada media sosial sangat bergantung kepada emoji yang digunakan.

Contohnya :-

 <p>Vishnupriya R @PriyaVishnu16</p> <p>Cheese cake 🍷</p> <p>7:27 AM · Jun 22, 2020 · Twitter Web App</p>	<table border="1"> <thead> <tr> <th></th> <th>Sentimen</th> </tr> </thead> <tbody> <tr> <td>Analisis sentimen menggunakan emoji</td> <td>Positif</td> </tr> <tr> <td>Analisis sentimen tanpa menggunakan emoji</td> <td>Neutral</td> </tr> </tbody> </table>		Sentimen	Analisis sentimen menggunakan emoji	Positif	Analisis sentimen tanpa menggunakan emoji	Neutral
	Sentimen						
Analisis sentimen menggunakan emoji	Positif						
Analisis sentimen tanpa menggunakan emoji	Neutral						

Rajah 9 Contoh *tweet* dan perbandingannya

Analisis sentimen tanpa menggunakan emoji akan menjangka sentimen tweet tersebut neutral hal ini disebabkan analisis ini menggunakan hanya menggunakan ['cheese', 'cake'] untuk menjangka sentimen. Skor tweet tersebut adalah 0 jika menggunakan analisis sentimen tanpa emoji.

Analisis sentimen yang menggunakan lebih tepat kerana analisis ini menggunakan ['cheese', 'cake', '\xF0\x9F\x98\x8D'] untuk menjangka sentimen. Emoji tersebut mempunyai skor sentimen yang positif iaitu 0.678. Skor tweet tersebut adalah 0.678 jika menggunakan analisis sentimen emoji. Oleh itu, jangkakan sentimen tweet tersebut adalah positif.

6 KESIMPULAN

Algoritma analisis sentimen yang berfokuskan emoji ini memberi ketepatan yang lebih tinggi daripada analisis sentimen yang menganalisis teks sahaja. Dengan perbandingan ini, kepergantungan sentimen teks kepada emoji dapat dikenal pasti dan analisis sentimen terhadap teks media sosial ini harus menggunakan algoritma analisis sentimen yang mengambilkira emoji yang digunakan. Algoritma analisis sentimen yang berfokuskan emoji ini memberi ketepatan 70.05% dengan penggunaan model pengelas *Naive Bayes*.

Secara konklusinya, algoritma ini boleh digunapakai oleh pengguna yang ingin menganalisis teks yang mengandungi emoji dengan tepat dan efisien. Setiap domain yang ingin dianalisis sentimen ini mempunyai penggunaan tatabahasa dan penentuan sentimen berbeza. Dengan algoritma ini, model yang berbeza untuk setiap domain dapat dihasilkan untuk penentuan sentimen. Model-model yang dihasilkan ini boleh disimpan untuk penggunaan masa depan. Walaupun dataset yang ingin dianalisis sentimen ini besar, algoritma analisis sentimen ini akan memberi output dengan ketepatan yang tinggi dalam masa yang singkat.

7 RUJUKAN

- A. Pak and P. Paroubek, 2010. Twitter as a Corpus for Sentimen Analysis and Opinion Mining. In 7th Conference on International Language Resources and Evaluation (LREC 2010), pages 1320–1326.
- Bing Liu, 2012. Sentimen analysis and opinion mining. Morgan & Claypool Publishers: 7 -8.
- C. Manning, T. Grow, T. Grenager, J. Finkel, and J. Bauer. Stanford Tokenizer, 2010. Available online, <http://nlp.stanford.edu/software/tokenizer.shtml>
- De Smet, W., & Moens, M. F. (2007). Generating a topic hierarchy from dialect texts. In DEXA Workshops (pp. 249–253). IEEE Computer Society.
- Emoji Sentimen Ranking. http://kt.ijs.si/data/Emoji_sentimen_ranking/
- Finn, A., & Kushmerick, N. ,2003. Learning to classify documents according to genre. Journal of the American Society for Information Science, 57, 1506–1518. Special issue on Computational Analysis of Style
- Freund, Y., Seung, H. S., Shamir, E., & Tishby, N. ,1997. Selective sampling using the query by committee algorithm. Machine Learning, 28, 133–168.
- Godsay, M. 2015. The process of sentimen analysis: A study. International journal of computer application (0975 – 8887) , 126(7) : 26 – 30.
- J. Read. Using Emoticons to Reduce Dependency in Machine Learning Techniques for Sentimen Classification. In Student Research Workshop at the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005), pages 43–48. Association for Computational Linguistics, 2005.
- K. Liu, W. Li, and M. Guo, 2012. Emoticon Smoothed Language Models for Twitter Sentimen Analysis. In 26th AAAI Conference on Artificial Intelligence (AAAI 2012), pages 1678–1684. Association for the Advancement of Artificial Intelligence, 2012
- List of emoticons. https://en.wikipedia.org/wiki/List_of_emoticons , 04 2016.
- Michele Di Capua, Alfredo Pertosino, Emanuel Di Nardo, September 2015. An architecture for sentimen analysis in Twitter. International Conference on E-Learning: 3–7.

- Nasukawa, Tetsuya and Jeonghee Yi. Sentimen analysis: Capturing favorability using natural language processing. in Proceedings of the KCAP-03, 2nd Intl. Conf. on Knowledge Capture. 2003.
- Ozan Irsoy and Claire Cardie. 2014. Opinion mining with deep recurrent neural networks. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-14), pages 720-728.
- R. Shepard. Recognition Memory for Words, Sentences, and Pictures. Journal of Verbal Learning and Verbal Behavior, 6(1):156–163, 1967.
- Shiho Hashimoto (1 June 2017). 7 Benefits of sentimen analysis.
<https://blog.insightsatlas.com/7-benefits-of-sentimen-analysis-you-cant-overlook>
- Vaalmeekam Karthik, Dheeraj Nair, Anuradha J, 2018. Opinion mining on emojis using deep learning techniques. Procedia Computer Science 132: 167-173.
- W Parott, 2001. Emotions in social psychology. Essential readings. Psychology Press.
- Wieslaw Wolny, 2016. Emotion analysis of twitter data that use emoticons and emoji ideograms. 25TH International Conference On Information Systems Development.