Vol. 3 No. 1, April 2022, pp. 140-147 **Jurnal Pepadun**

Analisis Sentimen *Tweet* Kasus Kebocoran Data Penggunaan Facebook oleh Cambrigde Analytica

[1]Ridho Sholehurrohman dan [2]Igit Sabda Ilman

[1,2]Jurusan Ilmu Komputer FMIPA, Universitas Lampung, Jl. Prof. Dr. Ir. Sumantri Brojonegoro No.1 Bandar Lampung, Lampung, Indonesia
e-mail:[1]ridho.sholehurrohman@fmipa.unila.ac.id, [2]igit.sabda@fmipa.unila.ac.id

Abstract *— The case of the Facebook user data leak by Cambridge Analytica has been spotlight in the public lately. Many of the citizens has participated discussing this case, especially in social media Twitter. Sentiment analysis is a computational research of opinions and emotions sentiment that are expressed textually. This study aims to classify positive and negative sentiment from Twitter data and to determine the accuracy of the classification model using Naïve Bayes Classifier method. Based on experiment conducted by tweet data with the "Zuckerberg" and "Cambridge Analytics" keywords, it has been produced Naïve Bayes Classifier with an accuracy of 83.06%.*

Keywords**:** *Facebook; Cambridge Analytics; Sentiment Analysis; Naïve Bayes Classifier.*

1. PENDAHULUAN

Perkembangan percepatan kemajuan teknologi yang semakin pesatsangat dirasakan dalam kehidupan sehari-hari. Kemajuan teknologi ini beriringan dengan penggunaan internet yang memudahkan masyarakat diseluruh dunia [1]. Berkembangnya internet tersebut dibuktikan dari munculnya aplikasi-aplikasi yang dapat membantu memudahkan kita semua. Contohnya kemudahan berkomunikasi jarak jauh seperti aplikasi *whatsapp*, *line*, *twitter*, *facebook*, *Instagram*, dan aplikasi sejenisnya. Hanya dengan *Handphone* yang canggih, kita dapat bersilaturrahmi dengan sanak saudara meskipun dengan jarak yang jauh.

Banyak penelitian pemanfaatan internet dan aplikasi yang telah dilakukan, seperti halnya [1] membahas tentang aplikasi facebook yang dapat menunjang kegiatan perkuliahan. [2] Yasya Wichitra dkk membahas tentang pengaruh penggunaan media sosial facebook terhadap perilaku pemberian asi. Namun, teknologi informasi saat ini juga memiliki dampak negatif karena menjadi sarana efektif perbuatan melawan hukum tindak pidana (kejahatan) yang biasa disebut "*cybercrime*".

Pada negara Amerika Serikat, Facebook merupakan salah satu situs jejaring sosial media dengan 164,58 juta pengguna [3] dan merupakan yang terbesar pertama di negara adidaya tersebut. Sayangnya pada tahun 2018, Facebook ditimpa kasus kebocoran data pengguna yang dilakukan oleh pihak Cambridge Analytica, dimana dampak dari kebocoran data ini berpengaruh terhadap hasil pemilihan umum Presiden AS ke 45 yang berhasil memenangkan Donald Trump sebagai calon terpilih.

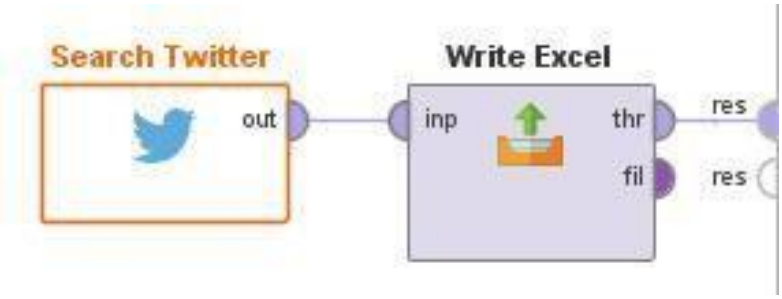
Dalam penelitian ini, penulis akan berfokus pada analisis sentimen serta pembobotan nilai akurasi hasil klasifikasi dengan metode *Naïve Bayes Classifier* dari data Twitter yang diperoleh dengan kata kunci "*Zuckerberg*" dan "*Cambridge Analytics*".

2. METODOLOGI PENELITIAN

Pada bagian ini penulis berfokus pada metode yang akan digunakan untuk melakukan penelitian terkait dengan analisis sentimen.

2.1. Akuisisi Data

Untuk mendapatkan data *tweet* penulis menggunakan aplikasi RapidMiner yang terhubung dengan API Twitter. Proses yang dilakukan ialah mengambil sebanyak 1000 data *tweet* yang mengandung kata "*Zuckerberg*" atau kata "*Cambridge Analytics*", yang selanjutnya data tersebut disimpan ke dalam format spreadsheet.



Gambar 1. Proses akuisisi data dari twitter

Data yang diperoleh dari twitter ini berisikan atribut ID pengguna, isi *tweet* yang dibuat, waktu posting *tweet* serta atribut lainnya. Dari data ini, kemudian penulis mengambil atribut *Text* atau isi *tweet* saja untuk dilakukan *Text pre- processing*.

R... ↑	Id	Text	Created-At	From-User	From-User-Id	To-User	To-User-Id
1	1004508114...	Cambridge Analytics CEO "allegedly withdrew more t...	Jun 7, 2018 6...	Adam Rawns...	198939070	?	-1
2	1001630864...	AIQ executives' prior testimony to MPs 'completely fal...	May 30, 2018 ...	David Carroll ...	15384720	?	-1
3	1001526907...	One former Cambridge Analytica employee is startin...	May 30, 2018 ...	WIRED	1344951	?	-1
4	1004512788...	RT @arawnsley: Cambridge Analytics CEO "allegedl...	Jun 7, 2018 6...	NastyElizabet...	37080933	?	-1
5	1004512048...	RT @arawnsley: Cambridge Analytics CEO "allegedl...	Jun 7, 2018 6...	jenjdnnwkkn...	467488893	?	-1
6	1004511786...	gerald_bader : Cambridge Analytica's Nix said it lice...	Jun 7, 2018 6...	SelfDriving car	8953316225...	?	-1
7	1004511663...	RT @arawnsley: Cambridge Analytics CEO "allegedl...	Jun 7, 2018 6...	M.R.M	2491425985	?	-1
8	1004509419...	RT @arawnsley: Cambridge Analytics CEO "allegedl...	Jun 7, 2018 6...	News Reader	2945911918	?	-1
9	1004508489...	RT @StandUpAmerica: #BREAKING: A Cambridge A...	Jun 7, 2018 6...	Yôhan	9120672877...	?	-1
10	1004508114...	Cambridge Analytics CEO "allegedly withdrew more t...	Jun 7, 2018 6...	Adam Rawns...	198939070	?	-1
11	1004507025...	@SavageLucia @hmkyale Apple will allow user to giv...	Jun 7, 2018 6...	ZibdyHealth	460598968	SavageLucia	2332653321
12	1004506844...	RT @StandUpAmerica: #BREAKING: A Cambridge A...	Jun 7, 2018 6...	Mike Conry	7882656404...	?	-1
13	1004501724...	@Rodrigo_Merinos Cambridge Analytics	Jun 7, 2018 6...	Ale	1275565278	Rodrigo_Meri...	301669969
14	1004493445...	@kylegriffin1 Hmm, this comes out as the Wikileaks ...	Jun 7, 2018 5...	Fall of Roses...	30865805	kylegriffin1	32871086
15	1004491546...	RT @bronze_bombSHEL: #Media Christopher Wylie ...	Jun 7, 2018 5...	Carter	121987601	?	-1

Gambar 2. Hasil akuisisi data dari Twitter

2.2. *Text Pre-processing*

Text pre-processing adalah tahapan yang harus dikerjakan dalam melakukan analisis teks dimana terbagi ke dalam beberapa tahapan adalah sebagaib berikut.

1) *Case folding*

Case Folding adalah proses yang paling sering digunakan sebagai tahapan pertama dalam *preprocessing*. *Case Folding* bertujuan untuk menghapus semua huruf kapital atau huruf besar yang ada pada data dokumen, atau secara lebih gampang *Case Folding* mengembalikan semua huruf besar

ke huruf kecilnya. *Case folding* yang dilakukan dalam penelitian ini ialah mengubah semua kata menjadi *lowercase* dalam isi *tweet* dari data twitter yang sudah didapat.

2) *Tokenizing*

Tokenizing merupakan suatu proses normalisasi data tekstual yang sebelumnya kalimat yang kemudian di pecah menjadi kata perkata. Sebagai contoh dari proses *tokenizing* yang diterapkan pada suatu kalimat dapat dilihat pada tabel dibawah ini. *Tokenizing* ialah memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata, tujuannya untuk membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan.

3) *Filtering*

Filtering merupakan tahapan pengambilan kata yang penting dari hasil proses sebelumnya. Tahapan *Filtering* ini dapat diterapkan dengan bantuan algoritma *stop list* atau *word list*. *Filtering* juga biasanya diartikan sebagai tahapan penghapusan *stopwords*, dimana *stopwords* sendiri adalah kosa kata yang sering digunakan sebagai kata penghubung atau bukan kata unik dari suatu dokumen

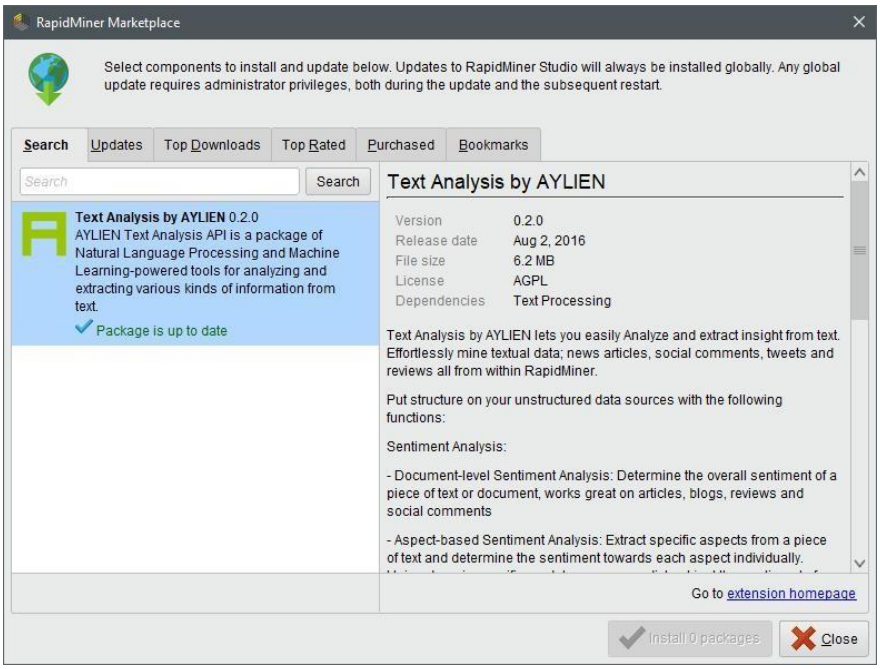
Procces Filtering dilakukan untuk melakukan pembersihan data terhadap kata-kata yang dianggap tidak penting atau tidak memiliki arti. Dalam proses ini dapat menggunakan metode *stopword* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting).

4) *Stemming*

Tahapan terakhir ialah melakukan *Stemming* guna mengelompokan kata-kata lain yang memiliki kata dasar dan arti yang serupa namun memiliki bentuk atau *form* yang berbeda karena mendapatkan imbuhan yang berbeda. *Stemming* adalah proses penguraian berbagai bentuk atau variasi kata dari hasil tahapan sebelumnya untuk mengembalikan kata tersebut menjadi kata dasarnya (*stem*).

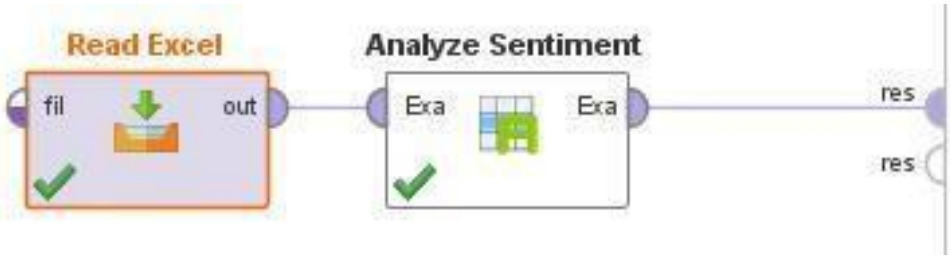
2.3. *Natural Language Processing* (NLP)

NLP adalah cabang ilmu komputer dan linguistik yang mengkaji interaksi antara komputer dengan bahasa (alami) manusia [3]. Dalam penelitian ini algoritma NLP yang digunakan terdapat dalam AYLIEN Text Analysis API yang merupakan suatu ekstensi pada aplikasi RapidMiner. Tujuannya untuk mendeteksi sentimen pada data *tweet*, baik dalam hal polaritas (positif atau negatif) atau dalam hal subjektivitas (subjektif atau objektif) [4].



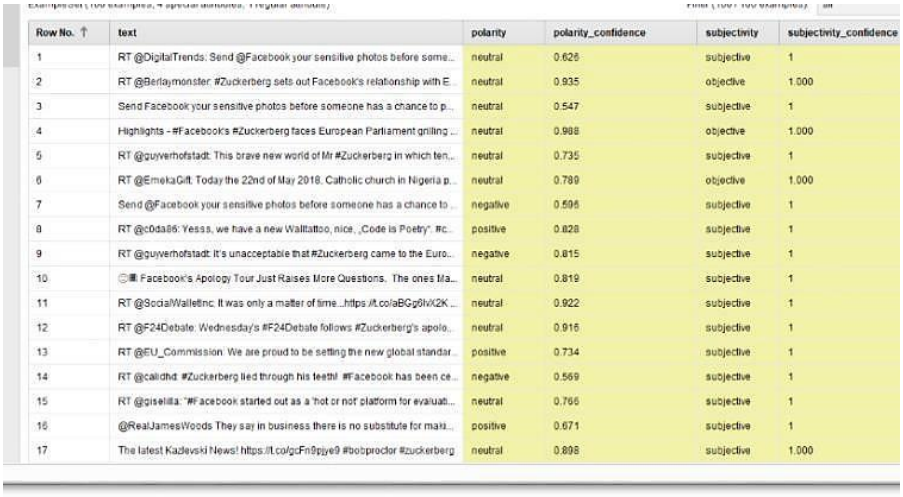
Gambar 3. Ekstensi *text analysis* by aylien pada aplikasi rapidminer

Berikut adalah proses dan hasil dari penggunaan operaton *Analyze Sentiment* yang disediakan dalam ekstensi *Text Analysis* by AYLIEN pada aplikasi RapidMiner.



Gambar 4. Proses klasifikasi sentimen pada aplikasi rapidminer

Hasillnya dapat dilihat pada Gambar 5 yang menunjukan hasil pelabelan sentimen untuk kategori polaritas (positif atau negatif) serta subjektivitas (subjektif atau objektif).



Row No.	text	polarity	polarity_confidence	subjectivity	subjectivity_confidence
1	RT @DigitalTrends: Send @Facebook your sensitive photos before some...	neutral	0.626	subjective	1
2	RT @Berlaymonster: #Zuckerberg sets out Facebook's relationship with E...	neutral	0.935	objective	1.000
3	Send Facebook your sensitive photos before someone has a chance to p...	neutral	0.547	subjective	1
4	Highlights - #Facebook's #Zuckerberg faces European Parliament grilling ...	neutral	0.988	objective	1.000
5	RT @guyverhofstadt: This brave new world of Mr #Zuckerberg in which ten...	neutral	0.735	subjective	1
6	RT @EmekaGift: Today the 22nd of May 2018, Catholic church in Nigeria p...	neutral	0.789	objective	1.000
7	Send @Facebook your sensitive photos before someone has a chance to ...	negative	0.596	subjective	1
8	RT @c0da86: Yesss, we have a new Walltattoo, nice, „Code is Poetry". #c...	positive	0.828	subjective	1
9	RT @guyverhofstadt: It's unacceptable that #Zuckerberg came to the Euro...	negative	0.815	subjective	1
10	🇪🇺 Facebook's Apology Tour Just Raises More Questions. The ones Ma...	neutral	0.819	subjective	1
11	RT @SocialWalletInc: It was only a matter of time...https://t.co/aBGg0lvK2K ...	neutral	0.922	subjective	1
12	RT @F24Debate: Wednesday's #F24Debate follows #Zuckerberg's apolo...	neutral	0.916	subjective	1
13	RT @EU_Commission: We are proud to be setting the new global standar...	positive	0.734	subjective	1
14	RT @calicthd: #Zuckerberg lied through his teeth! #Facebook has been ce...	negative	0.569	subjective	1
15	RT @giselitta: "#Facebook started out as a 'hot or not' platform for evaluati...	neutral	0.766	subjective	1
16	@RealJamesWoods They say in business there is no substitute for mark...	positive	0.671	subjective	1
17	The latest Kadevski News! https://t.co/gcFn9pjye9 #bobproctor #zuckerberg	neutral	0.898	subjective	1.000

Gambar 5. Hasil pelabelan polaritas dan subjektifitas pada data *tweet*

Dalam proses ini dilakukan penghapusan data *tweet* yang mengandung polaritas netral, sehingga untuk proses selanjutnya hanya digunakan data *tweet* dengan sentimen positif dan negatif saja.

2.4. *Naïve Bayes Classifier* (NBC)

Naïve Bayes Classifier merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing kondisi/kejadian [5]. Pada teorema Bayes yang menjadi dasar dari metode tersebut, bila terdapat dua kejadian yang terpisah (misalkan A dan B), maka teorema Bayes dirumuskan sebagai berikut [6]:

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A) \quad (1)$$

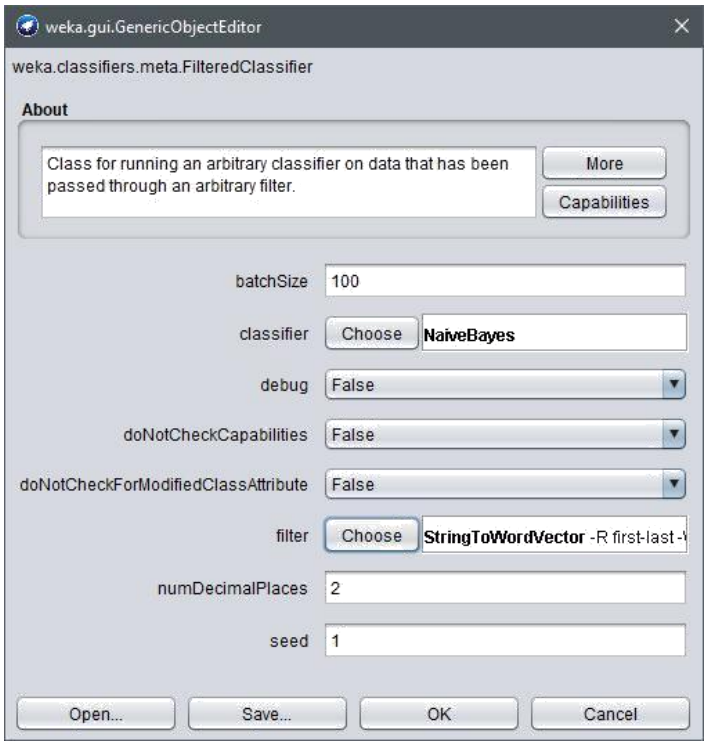
Teorema Bayes tersebut dapat dikembangkan mengingat berlakunya hukum probabilitas total, menjadi:

$$P(A|B) = \frac{P(A)P(B|A)}{\sum_{i=1}^{n} P(Ai|B)} \quad (2)$$

dimana $A_1 \cup A_2 \cup \ldots \cup A_n = S$

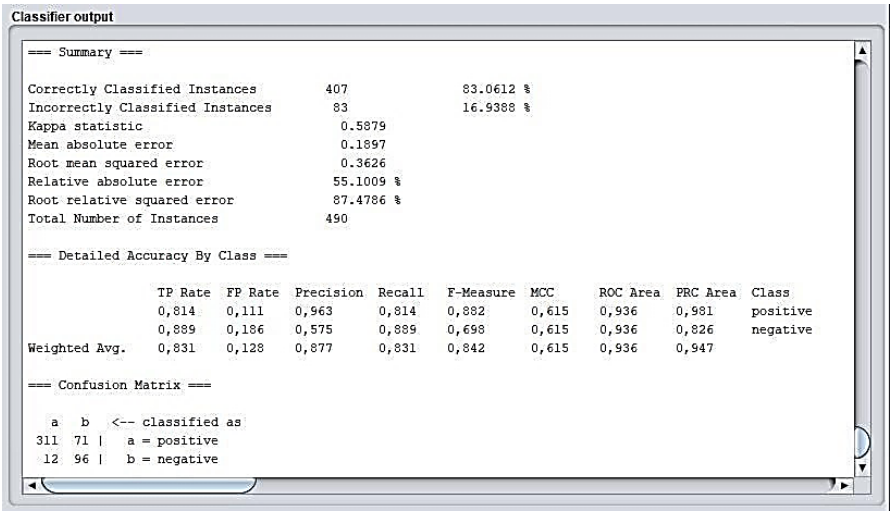
3. HASIL DAN PEMBAHASAN

Dalam penelitian ini data *tweet* yang sudah diberi label sentimen positif dan negatif akan dicari bobot tingkat akurasinya dengan metode *Naïve Bayes Classifier* yang terdapat dalam aplikasi Weka. Dimana dilakukan proses klasifikasi dengan metode *classifier NaiveBayes* dan filter *unsupervised StringToWordVector* adalah sebagai berikut.



Gambar 6. Filter klasifikasi *naïve bayes* pada apilkasi weka

Selanjutnya, diberikan hasil dari proses klasifikasi dengan filter *Naïve Bayes classifier* pada aplikasi Weka adalah sebagai berikut.



Gambar 7. Hasil pengklasifikasian *naïve bayes* pada aplikasi weka

Hasil pengklasifikasian *NaiveBayes* pada aplikasi Weka menunjukan bahwa dari total 490 *instance* yang dimuat, terdapat sebanyak 407 atau setara dengan 83,0612% *instance* yang diklasifikasikan dengan tepat. Untuk hasil lebih detailnya akan disajikan dalam beberapa tabel, sebagai berikut.

Tabel 1. *Output summary*

Output	***Result***
Correctly Classified Instances	407 83.0612 %
Incorrectly Classified Instances	83 16.9388 %
Kappa statistic	0.5879
Mean absolute error	0.1897
Root mean squared error	0.3626
Relative absolute error	55.1009 %
Root relative squared error	87.4786 %
Total Number of Instances	490

Berdasarkan Tabel 1 *output summary* diatas terdapat sebanyak 407 data yang terklasifikasi dengan tepat dari total data 490 data dengan persentase 83,0612%. Hal ini sebanding dengan 83 data *error* atau tidak terklasifikasinya data atau 16,9382%. Selanjutnya diberikan akurasi hasil uji coba klasifikasi tersebut sebagai berikut.

Tabel 2. Hasil uji coba *naïve bayes classifier*

Class	***TP Rate***	***FP Rate***
Positive	0,814	0,111
Negative	0,889	0,186
Weighted Avg	0,831	0,128

Tabel 2. Hasil uji coba *naïve bayes classifier* (2)

Class	**F-Measure**	**MCC**
Positive	0,882	0,615
Negative	0,698	0,615
Weighted Avg	0,842	0,615

Berdasarkan Tabel 2 Hasil Uji Coba *Naïve Bayes Classifier*, akan didapatkan nilai akurasi, presisi dan *recall* dengan rumus performa matrik sebagai berikut.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

$$Persisi = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

Maka, didapat sebagai berikut.

Tabel 3. *Detailed accuracy by class*

Class	*Accuracy*	*Precision*	*Recall*
Positive	1,963	0,963	0,814
Negative	0,889	0,889	0,889
Weighted Avg	1,693	0,831	0,831

Tabel 3. *Detailed accuracy by class* (2)

Class	*All Area*	*ROC Area*	*PRC Area*
Positive	1,936	0,936	0,981
Negative	0,936	0,936	0,826
Weighted Avg	1,693	0,936	0,947

Berdasarkan hasil akurasi, presisi dan *recall* pada Tabel 2 *Detailed Accuracy By Class* dan Tabel 3 *Confusion Matrix* diatas menyatakan bahwa proses klasifikasi dari 490 data sangat baik dan efektif. Dengan rata-rata *error* presisi 0,873 dan *error recall* 0,851. Sedangkan *error* rata-rata akurasi 1,693.

4. KESIMPULAN

Pada penelitian ini, pengimpelentasian metode *Naïve Bayes Classifier* untuk 490 data sentimen *tweet* kasus kebocoran data penggunaan facebook oleh cambrigde analytica berhasil dan tepat. Hasil analisis performansi *Naïve Bayes Classifier* menunjukkan bahwa nilai rata-rata akurasi klasifikasi yang cukup tinggi untuk pembobotan sentimen data *tweet* dengan kata kunci "*Zuckerberg*" dan "*Cambridge Analytics*" yakni sebesar 83,06%.

DAFTAR PUSTAKA

[1] M. Ziveria, "Pemanfaatan Media Sosial Facebook Sebagai Sarana Efektif Pendukung Kegiatan Perkuliahan di Program Studi Sistem Informasi Institut Teknologi dan Bisnis Kalbe," *Jurnal Science dan Teknologi Kalbis Scientia*, Vol 4, No. 2, 2017.

[2] W. Yasya, P. Muljono, K. B. Seminar & H. Hardinsyah, "Pengaruh Penggunaan Media Sosial Facebook Dan Dukungan Sosial Online Terhadap Perilaku Pemberian Asi," *Jurnal Studi Komunikasi dan Media*, 2019

[3] Statista: The Statistics Portal. (2018). Most popular social media apps in the U.S. 2018, by audience. [Online]. Available: https://www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/

[4] Wikipedia. (2013). "Pemrosesan Bahasa Alami". [Online]. Available: https://id.wikipedia.org/wiki/Pemrosesan_bahasa_alami

[5] Aylien. (2018). Text Analysis API: Natural Language Processing for Effective Understanding of Human-Generated Text. [Online]. Available: https://aylien.com/text-api/

[6] S. Natalius, Metoda Naïve Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen, Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung, 2010.