

Klasifikasi Opini Masyarakat Terhadap Pertanian di Indonesia dengan Naive Bayes pada Twitter

Hafiz Irsyad¹, M Rizky Pribadi²

¹Teknik Informatika, STMIK Global Informatika MDP Jalan Rajawali No.14, Palembang

²Teknik Informatika, STMIK Global Informatika MDP Jalan Rajawali No.14, Palembang
e-mail: ¹hafizirsyad@mdp.ac.id,² rizky@mdp.ac.id

Abstrak

Pertanian di Indonesia dalam beberapa tahun terakhir menjadi sebuah perbincangan yang sangat penting. Sektor pertanian Indonesia mengalami pertumbuhan yang sangat signifikan sampai mencapai di angka 2.56% pada tahun 2017. Dengan angka yang disebutkan tadi membuat pemerintah Indonesia mengeluarkan target yang sangat tinggi. Target tersebut adalah menjadikan Indonesia sebagai lumbung pangan dunia pada tahun 2045. Agar tercapainya target tersebut perlu adanya perluasan bisnis, salah satunya adalah melalui media sosial. Dalam perluasan bisnis dengan media sosial membuat masyarakat banyak mengeluarkan pendapat, pernyataan, maupun tanggapan tentang target yang telah ditetapkan oleh pemerintah.

Prediksi dalam mencapai hasil dari target pemerintah dapat dilakukan dengan menganalisa opini masyarakat atau biasanya disebut dengan sentimen analisa yang dapat dicurahkan dalam teks yang dituliskan oleh penggiat sosial media. Metode penelitian yang digunakan oleh peneliti terdiri dari beberapa tahap, yaitu: pengambilan data pada twitter dengan memanfaatkan fasilitas API Twitter, membuat metode analisis sentimen (pre-processing, case folding, tokenizing dan stop removal) menggunakan tools orange dan metode analisis menggunakan metode Naïve Bayes. Data yang digunakan adalah sebanyak 391 dengan masing-masing kaegori. Metode Naive Bayes pada penelitian ini menghasilkan akurasi dengan rata-rata 0.523% pada seluruh kategori, untuk Precision menghasilkan rata-rata 0,865% diseluruh kategori dan untuk Recall sendiri mendapatkan persentase 0,701 % dari keseluruhan kategori opini masyarakat di twitter. Dengan tingkat akurasi tersebut dapat disimpulkan metode Naïve Bayes berjalan sesuai dengan harapan.

Kata Kunci : Naïve Bayes, Pertanian, Klasifikasi Opini

Abstract

Agriculture in Indonesia in recent years has become a very important conversation. Indonesia's agricultural sector experienced very significant growth until it reached 2.56% in 2017. With the figures mentioned above, the Indonesian government issued a very high target. The target is to make Indonesia a world food barn in 2045. In order to achieve this target, business expansion is needed, one of which is through social media. In expanding the business with social media, it makes the public issue many opinions, statements, and responses about the targets set by the government.

Prediction in achieving the results of government targets can be done by analyzing public opinion or usually referred to as sentiment analysis that can be poured out in a text written by social media activists. The research method used by the researcher consists of several stages, namely: data collection on twitter by utilizing the Twitter API facility, making sentiment analysis methods (pre-processing, folding cases, tokenizing and stop removal) using orange

tools and analysis methods using the Naïve Bayes method . The data used is 391 with each category. The Naive Bayes method in this study resulted in an accuracy of 0.523% in all categories, for Precision produced an average of 0.865% in all categories and for Recall alone earned a percentage of 0.701% of the overall category of public opinion on twitter. With this level of accuracy, it can be concluded that the Naïve Bayes method goes according to expectations.

Keywords : Naïve Bayes, Agriculture, Opinion Classification.

1. PENDAHULUAN

Indonesia adalah negara yang memiliki letak geografis yang strategis. Dengan letak geografis yang strategis Indonesia memiliki Sumber Daya Alam yang berlimpah salah satunya terdapat di sektor pertanian, Berdasarkan data dari Badan Pusat Statistika (BPS), tiga sektor utama yang ikut andil dalam pergerakan ekonomi Indonesia adalah sektor pertanian, sektor perdagangan dan sektor perindustrian. Pada 2017 produksi padi nasional mengalami pertumbuhan sebesar 2.56% dibandingkan pada tahun-tahun sebelumnya [1]. Pada tahun yang sama pemerintah berhasil menghentikan impor beras. Sehingga pemerintah memiliki target yang sangat tinggi yaitu mewujudkan Indonesia menjadi lumbung pangan di dunia pada tahun 2045 [2]. Untuk menjaga agar tercapainya Indonesia sebagai lumbung pangan dunia, sektor pertanian yang pada umumnya berkembang secara tradisional sekarang diubah menjadi berkembang secara industri. Media sosial salah satu peran utama dalam penggerakan perluasan bisnis serta sebagai pemasaran produk.

Twitter merupakan salah satu dari sekian banyak media sosial yang difungsikan sebagai pemasaran produk, dimana Twitter sebagai media sosial yang sangat populer setelah Facebook, Instagram dan lain-lain. Pada tahun 2016 Indonesia memiliki pengguna aktif sebanyak 24.340 juta orang. Twitter dapat menampung sebanyak 280 karakter, kemudian diolah menjadi sebuah *statement* [3]. Dewasa ini media sosial Twitter dimanfaatkan sebagai perluasan bisnis, dimana pengguna bisa menemukan pelaku bisnis lain sehingga bisa menjadi teman atau pengikut (*followers*) dan tentu bisa saling berinteraksi.

Twitter menyediakan fasilitas-fasilitas agar bisa dimanfaatkan oleh para pengguna akun Twitter, salah satunya adalah fasilitas *retweet*. Banyaknya *retweet* atau komentar masyarakat tentang perkembangan maupun penyebaran informasi sektor pertanian di twitter, *retweet* tersebut dapat diolah kemudian disajikan dalam bentuk informasi sehingga dapat mengetahui pendapat masyarakat tentang perkembangan sektor pertanian. *Retweet* dari *followers* akan dapat mempengaruhi pola pikir dari *followers* sehingga bisa membuat *retweet* bisa memungkinkan terjadinya *overload retweet*. Dengan demikian *retweet* dapat di klasifikasikan kalimat yang berhubungan dengan *retweet* perkembangan pada sektor pertanian, sehingga dapat menemukan pola dalam pengelompokan kalimat yang positif, negatif dan netral. Pengumpulan data *tweet* dari twitter dapat dilakukan dengan menghubungkan Twitter API dan Orange Tools.

Metode yang digunakan dalam klasifikasi adalah metode Naïve Bayes dan K-Means untuk menghitung *precision*, *recall*, *F-measure* . Adapun keuntungan yang didapatkan dalam penerepan metode Naïve Bayes ini digunakan oleh beberapa peneliti diantara lain adalah memprediksi besarnya penggunaan listrik untuk tiap rumah tangga dengan hasil yang diperoleh dalam pengujian penelitian tersebut yang menggunakan 60 data yang diperoleh, 47 data yang mampu diklasifikasikan dengan benar dan nilai yang diperoleh dari akurasi 78,333% untuk keakuratan prediksi [4]. Sementara untuk metode K-Means hasil penelitian terhadap analisis sentimen yang menunjukkan bahwasan akurasi dari K-Means dengan dataset digunakan 300 positif dan 300 negatif dan akurasi 57.83%, 700 dokumen positif dan 700 dokumen negative akurasi 56.71% dari hasil tersebut maka disimpulkan semakin banyak dataset maka semakin rendah akurasi dari metode K-Means [6].

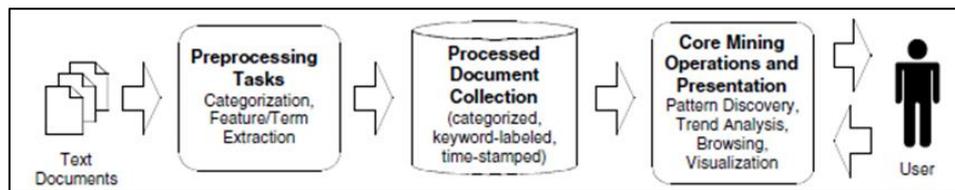
Berdasarkan penelitian yang telah dituliskan diatas, penelitian ini akan melakukan sebuah analisis sentiment dengan mengklasifikasikan data pada twitter. Data yang telah dipereloh dari Twitter akan diproses dengan text mining untuk menghindari data yang tidak diinginkan, setelah itu dikelompokan data tweet atau komentar berdasarkan klasifikasi positif, negatif dan netral. Pengelompokan ini menggunakan perbandingan metode Naïve Bayes.

2. METODOLOGI PENELITIAN

2.1. Text Mining

Text mining adalah proses mengekstrak informasi yang berguna dari sekumpulan dokumen dari waktu ke waktu melalui identifikasi pola yang menarik [3]. *Text mining* adalah proses penemuan informasi baru dengan mengekstrak pola secara otomatis dari berbagai sumber teks [4].

Text mining mempunyai beberapa tingkatan fungsional, *text mining* mengikuti model yang telah disediakan oleh beberapa aplikasi data mining sehingga dapat terbagi menjadi empat tingkatan fungsional utama, yakni *preprocessing task*, *processed document collection*, dan *core mining operations* [3]. Tingkatan arsitektur dari *text mining* dapat dilihat pada gambar. 1.



Gambar 1. Arsitektur *Text Mining* [3]

2.2. Preprocessing

Text preprocessing merupakan awal mula dari *text mining*. Tujuan dari *text preprocessing* adalah membersihkan data dari noise sehingga data menjadi lebih kecil dan lebih terstruktur. Berikut ini adalah tahapan proses dari *text preprocessing* [3].

a. Tokenizing

Tokenizing merupakan proses pemecahan teks menjadi kata tunggal dan menghapus tanda baca serta angka, sesuai dengan kamus yang telah ditentukan.

b. Stopword Removing

Stopword removing merupakan proses menghilangkan kata tidak penting dalam teks. Hal ini dilakukan untuk memperbesar akurasi dari pembobotan. Dalam penelitian ini *Stopword removing* digunakan untuk menghilangkan kata-kata seperti: dan, atau, mungkin, ini, itu dan sebagainya merupakan kata yang dapat dihilangkan.

c. Stemming

Stemming merupakan perubahan berbagai kata imbuhan menjadi sebuah kata dasar. Dalam penelitian ini kata imbuhan yang dihilangkan adalah teks yang berbahasa inggris, dikarenakan teks Bahasa inggris memiliki bentuk lampau.

2.3. Naïve Bayes

Naïve Bayes adalah salah salah satu algoritma klasifikasi yang ditemukan oleh Tomas Bayes. Klasifikasi adalah proses untuk memperoleh aturan atau model yang dapat mengklasifikasikan data baru yang belum pernah dipelajari dengan mempelajari sekumpulan

data yang lama. Naïve Bayes adalah algoritma klasifikasi probabilistik yang sederhana berdasarkan Teorema Bayes [5]. Prinsip umumnya adalah mengasumsikan bahwa nilai suatu atribut tidak bergantung dan mempengaruhi atribut yang lainnya. Model Naïve bayes memungkinkan setiap atribut memiliki kontribusi yang sama terhadap keputusan akhir dan komputasinya lebih efisien bila dibandingkan dengan algoritma pengklasifikasi teks lainnya. Model yang diperoleh dari proses training berisikan kumpulan konstanta untuk setiap data train. Model tersebut akan digunakan pada data testing untuk melihat seberapa akurat model yang telah diperoleh [6]. Naïve bayes menjumlahkan frekuensi dan kombinasi nilai dari dataset untuk menghitung probabilitas [7]. Probabilitas dihitung dengan menggunakan persamaan 1 [5].

$$P(H|X) = (P(X|H).P(H))/P(X) \quad (1)$$

Pada persamaan diatas adalah :

X = Data dengan class yang belum diketahui
 H = Hiptotesis data merupakan suatu class spesifik
 P(H|X)[^] = Probabilitas hipotesis berdasar kondisi
 P(H) = Probabilitas hipotesis
 P(X|H) = Probabilitas berdasarkan kondisi pada hipotesis
 P(X) = Probabilitas H

2.4. Sentimen Analysis

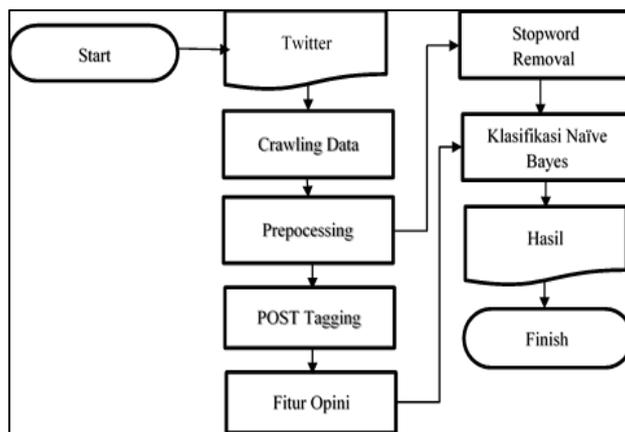
Analisis sentimen, juga disebut penambangan opini, adalah bidang studi itu menganalisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap entitas seperti produk, layanan, organisasi, individu, masalah, peristiwa, topik, dan atributnya. Itu mewakili besar ruang masalah. Ada juga banyak nama dan tugas yang sedikit berbeda, misalnya, analisis sentimen, penambangan opini, ekstraksi pendapat, sentimen penambangan, analisis subjektivitas, analisis pengaruh, analisis emosi, peninjauan penambangan, dan lain-lainnya. Namun, sekarang semuanya berada di bawah payung analisis sentimen atau penambangan opini. Sementara di industri, analisis sentimen jangka lebih umum digunakan, tetapi di akademisi baik analisis sentimen dan penambangan opini sering digunakan.

2.5. Twitter Rest API

Twitter adalah jejaring sosial populer di mana pengguna dapat berbagi pesan singkat seperti SMS (*Short Message Service*) yang disebut *tweet*. Pengguna berbagi pemikiran, tautan, dan gambar di Twitter, jurnalis mengomentari secara langsung, perusahaan mempromosikan produk, dan terlibat dengan pelanggan. Berbagai cara dalam menggunakan Twitter, dengan 500 jutaan *tweet* per hari, ada banyak data untuk dianalisis dan dimainkan. Twitter Rest APIs menyediakan akses terprogram untuk membaca dan menulis data Twitter. Buat *tweet* baru, baca profil penulis dan data pengikut, dan banyak lagi. REST API mengidentifikasi aplikasi Twitter dan pengguna yang menggunakan OAuth; tanggapan tersedia di JSON [8].

2.6. Tahapan Penelitian

Gambaran umum dalam proses skenario penelitian dapat dilihat pada gambar 2. dapat dilihat sebagai berikut:



Gambar. 2 Gambaran Umum Penelitian

Skenario dari penelitian ini adalah sebagai berikut:

- a. *Tools Orange* akan menerima inputan berupa dataset twitter. Data tersebut diambil berdasarkan inputan data “Pertanian”.
- b. Melakukan *preprocessing* terhadap dataset agar dapat memenuhi klasifikasi dan mempermudah dalam memproses data.
- c. *POST Tagging* berfungsi sebagai pencarian faktor opini. *POST Tagging* berfungsi juga sebagai ekstraksi fitur opini.
- d. Setelah fitur opini didapatkan, maka akan dilakukan langkah selanjutnya dengan menggunakan *stopword removal*. *Stopword removal* sendiri berfungsi untuk menghapus kata-kata yang sering muncul akan tetapi tidak memiliki makna dalam penelitian ini.
- e. Proses selanjutnya adalah melakukan pengklasifikasian sentimen data dengan menggunakan Naïve Bayes.

Dataset yang digunakan dalam penelitian ini merupakan opini-opini masyarakat tentang pertanian yang ada di Indonesia pada saat sekarang. Algoritma yang digunakan untuk klasifikasi adalah Naïve Bayes dengan menggunakan *tools Orange*.

- a. *Crawling Data Tweets*
 Dataset yang digunakan dalam penelitian ini adalah data dari berbagai akun yang selalu mentweetkan pertanian indonesia. Dataset yang diambil menggunakan *Application Programming Interface (API)* Twitter menggunakan *Tools Orange*. *Keyword* yang digunakan adalah pertanian, penarikan data mulai dari tanggal 1 Juni 2019 sampai dengan 08 Juni 2019. Dari hasil *crawling* data diperoleh *tweet* sebanyak 501 *tweet*. Adapun hasil dari perolehan *crawling* data adalah:

Tabel 1. Hasil Crawling dari Twitter

| No | Komentar Opini |
|----|--|
| 1. | Banyak juga yg akhirnya hasil panen ngga dijual tapi dikonsumsi sendiri. Akhirnya mereka juga harus kerja di luar sektor pertanian untuk memenuhi kebutuhan. Atau kadang memberlakukan sistem barter karena kadang sifatnya masih kekeluargaan. |
| 2. | @shitlicious Pernah ngobrol sama org kementrian, salah satu problem dunia pertanian di indonesia terutama sawah, teknologi minim sekali peranannya. Entah petani menolak perubahan cara menanam, dan alasan2 lainnya yg intinya memang tidak berubah cara bertaniya. |

| | |
|----|--|
| 3. | @Pangan_Kita Kementan harus berhati-hati.. niatnya ingin membuat RI jadi lumbung pangan dunia malah bisa jadi blunder untuk sektor pertanian. |
| 4. | RT @republikaonline: Tanggal 7 Juni diperingati sebagai hari keselamatan pangan sedunia. https://t.co/fnkg1A2cpH |
| 5. | [Fakta Melody] Melody suka menanam pohon, maka dari itu dia ambil jurusan pertanian, biar bisa menanam benih cinta gue ke dia |
| 6. | "@shitlicious Dulu kepikiran, gimana kalo para petani petani itu dijadikan PNS ato karyawan pada dinas pertanian. Mereka punya gaji tetap tiap bulan, tapi juga memiliki target untuk produksi, dan jika produksinya berlimpah, dapat bonus tambahan.." |
| 7. | "Indonesia menempati peringkat 1 dalam produk pertanian, yaitu : cengkeh (cloves) & pala (nutmeg), serta no.2 dalam karet alam (Natural Rubber) dan minyak sawit mentah (Crude Palm Oil)." |
| 8. | RT @ardi_riau: Dari hal pertanian, Aceh adalah penghasil nilam terbesar di dunia, sama halnya dengan Riau. Cadangan pertambangan minyak bum... |
| 9. | Sebanyak 398 pasar tani di seluruh negara beroperasi lima hari mulai semalam sehingga Selasa ini (4 Jun) bagi memberi peluang para pengguna mendapatkan bekal keperluan bersempena Hari Raya Aidilfitri - Menteri Pertanian dan Industri Asas Tani @MyDSA_official |

b. *Preprocessing*

Pada tabel 1. Diatas telah didapatkan hasil dari *crawling* data dari Twitter, maka selanjutnya kita akan ketahap *preprocessing* agar data-data yang telah *crawling* tersebut dapat menjadi terstruktur. Berikut ini adalah tahapan dari proses *preprocessing*:

- *Tokenizing*

Sebagaimana telah dijelaskan di atas, *Tokenizing* merupakan proses pemecahan teks menjadi kata tunggal dan menghapus tanda baca serta angka, sesuai dengan kamus yang telah ditentukan.

Maka berdasarkan dataset ditwitter, pada umumnya pemilik akun hanya menggunakan kata-kata baku bahkan menggunakan kata-kata gaul untuk membuat sebuah ciutan atau tweet. Tugas pada *Tokenizing* ini adalah mengubah kata-kata baku tersebut sehingga lebih mudah dimengerti.

- *Stopword Removing*

Stopword removing merupakan proses menghilangkan kata tidak penting dalam teks. Hal ini dilakukan untuk memperbesar akurasi dari pembobotan. Dalam penelitian ini *Stopword removing* digunakan untuk menghilangkan kata-kata seperti: dan, atau, mungkin, ini, itu dan sebagainya merupakan kata yang dapat dihilangkan.

- *Post Tagging Tweet*

Selanjutnya adalah melakukan pelabelan secara manual dari beberapa proses yang telah diciptakan tadi. Dengan adanya sebuah label maka tweet akan mudah dikelompokan, adapun label yang digunakan sebanyak tiga label yakni Positif, Negatif dan Netral.

Preprocessing telah dibuat berdasarkan point-point diatas ada beberapa data yang tidak terpakai juga bakalan dibuang, seperti tweet yang mengandung unsur *emoticon*, *url*, *rt*, gambar (biasanya menjadi tweet yang kosong pada saat terjadi *crawling* data). Dari data tweet sebanyak

501 terdapat data yang sama, sehingga dihapus dan setelah dihapus maka data yang tersisa untuk digunakan adalah sebanyak 391 data training. Dari data tersebut dapat dilihat pada Tabel 2.

Tabel 2. Data *training* dan Data *Testing*

| Data | Positif | Negatif | Netral |
|----------------------|---------|---------|--------|
| Data <i>Training</i> | 142 | 108 | 141 |

Dari tabel 2. Diatas proses data *training* akan menggunakan algoritma Naïve Bayes yang sudah tersedia pada *tools* Orange.

Evaluasi performansi dilakukan untuk menguji hasil dari klasifikasi dengan mengukur nilai performansi. Parameter pengujian yang digunakan untuk evaluasi yaitu akurasi yang perhitungannya diperoleh dari tabel *confution matrix* [7]. Dari *confution matrix* dapat dihitung nilai akurasi, *precision* dan *recall* [8].

Akurasi tersebut dapat dihitung dengan menggunakan persamaan 2.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Untuk menghitung *precision* menggunakan persamaan 3.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Untuk menghitung nilai *recall* dapat dilihat pada Persamaan 4.

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Dimana:

TP = *True Positif Count*.

TN = *True Negative Count*.

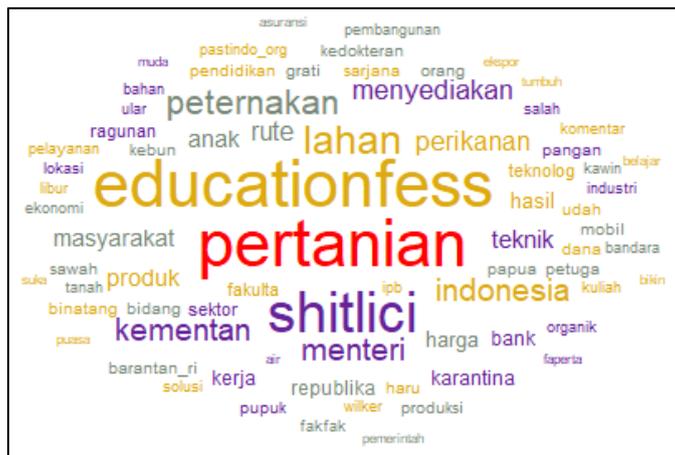
FP = *False Positive Count*.

FN = *False Negative Count*.

3. HASIL DAN PEMBAHASAN

Dari hasil data training maka akan digunakan probabilitas untuk mendapatkan hasil perolehan yang mendominasi atau yang banyak digunakan sehingga mendapatkan opini yang dikeluarkan oleh masyarakat. Adapun nilai yang dihasilkan untuk kategori positif adalah 0.358, untuk kategori negatif adalah 0.276 dan untuk netral adalah 0.360. dari hasil tersebut yang paling banyak digunakan adalah kategri positif.

Setelah mendapatkan kategori positif, negatif serta netral pada *Post Tagging Data* dilanjutkan dalam penggunaan *wordcloud*. *Wordcloud* digunakan pengklasifikasian berdasarkan kategori yang telah diketahui, pada gambar 2 dapat dilihat *wordcloud* yang sering muncul dalam pengkategorian:



Gambar 5. Wordcloud Sentimen Negative, Positif dan Netral

Dari hasil wordcloud tersebut, terdapat 9 kata dominan dari setiap masing-masing opini yang sudah melalui klasifikasi yakni:

Tabel 3. Wordcloud yang Dominan

| Weight | Word |
|--------|---------------|
| 348 | Pertanian |
| 70 | Educationfess |
| 52 | Shitlici |
| 34 | Lahan |
| 29 | Peternakan |
| 27 | Menteri |
| 26 | Indonesia |
| 26 | Kementan |
| 23 | Menyediakan |

Tahapan selanjutnya adalah melakukan pengujian data *tweets*. Pengujian tersebut berdasarkan kategori positif adalah 142, untuk kategori negatif adalah 108 dan untuk kategori netral 141 sehingga hasil tersebut dapat dilihat pada tabel 4.

Tabel 4. Wordcloud yang Dominan

| Kategori | Positif | Negatif | Netral |
|----------|---------|---------|--------|
| Positif | 110 | 32 | 0 |
| Netral | 1 | 40 | 100 |
| Negatif | 20 | 87 | 1 |

Berdasarkan data pada tabel 4 diatas, maka dapat diubah menjadi *confusion matrix* yang akan digunakan untuk menghitung akurasi, *precision* dan *recall*, sehingga menghasilkan setiap kategori yang dapat dilihat pada Tabel 5.

Tabel 5. Wordcloud yang Dominan

| Kategori | Akurasi | Precision | Recall |
|----------|---------|-----------|--------|
| Positif | 0.570 | 0.846 | 0.558 |
| Netral | 0.566 | 0.921 | 0.715 |
| Negatif | 0.433 | 0.830 | 0.830 |
| Average | 0.523 | 0.865 | 0.701 |

Dari hasil table 5. dapat dilihat bahwasan kategori yang memiliki tingkat akurasi yang lebih tinggi adalah Positif dengan nilai 0,570%, untuk kategori netral dengan nilai 0,566% dan negatif berada pada posisi paling rendah yakni 0,433% sehingga didapatkan rata-rata dari keseluruhan tingkat akurasi adalah 0,523%. Dari hasil tersebut terdapat selisih pada Positif dan Netral sebesar 0.004 % dengan terdapat selisih yang tidak signifikan membuktikan terdapat kesalahan. Kesalahan ini diakibatkan ada beberapa factor berupa tweet dari penggiat yang bisa mengirim berulang kali, bahkan ada juga kemiripan dari beberapa tweet dari pengguna akun, sehingga ada kendala dalam melakukan pengkategorian.

4. KESIMPULAN

Hasil dari opini masyarakat terhadap pertanian Indonesia dengan menggunakan dataset pada twitter dan penerapan metode Naïve Bayes menghasilkan akurasi dengan rata-rata 0.523% pada seluruh kategori, untuk Precision menghasilkan rata-rata 0,865% diseluruh kategori dan untuk *Recall* sendiri mendapatkan persentase 0,701 % dari keseluruhan kategori opini masyarakat di twitter. Hanya terdapat sedikit permasalahan pada akurasi antara Positif dan Netral sehingga *tools* dari Orange kurang berjalan baik.

Opini yang mendominasi dari hasil pembahasan diatas terdapat 9 pokok kata domain yakni, Pertanian, *Educationfess*, *shitlici*, lahan, peternakan, Menteri, Indonesia, Kementan dan menyediakan. Saran yang dapat disampaikan untuk pembahasan adalah dapat melanjutkan penelitian yang lebih mendalam, sehingga mendapatkan hasil yang lebih bagus lagi pada saat sekarang.

DAFTAR PUSTAKA

- [1] B. P. Statistik., www.bps.go.id., 5-2-2018.[Online]. Available: <https://www.bps.go.id/pressrelease/2018/02/05/1519/ekonomi-indonesia-triwulan-iv-2017--tumbuh-5-19-persen.html>. [Accessed 10 8 2018].
- [2] A. W. F. Aisyah., www.goodnewsfromindonesia.id., 03-10-2017. [Online]. Available: <https://www.goodnewsfromindonesia.id/2017/10/03/tahun-2045-wujudkan-indonesia-menjadi-lambung-pangan-dunia>. [Accessed 10 08 2018].
- [3] M. Rani and A. J, Twitter Data Predicting Geolocation Using Data Mining Techniques, *International Journal of Innovative Research in Computer*, vol. 4, no. 6, p. 10446, 2016.
- [4] A. Saleh, Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga, *Creative Information Technology Journal* , vol. 2, no. 3, pp. 207-2017, 2015.
- [5] S. Ting, A. H. Tsang and W. H. Ip, Is Naïve Bayes a Good Classifier for Document Classification?, *International Journal of Software Engineering and its Applications*, vol. 5, no. 3, p. 398, 2011.
- [6] S. Budi, Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Means, *Techno.COM*, vol. 16, no. 1, pp. 1-8, 2017.
- [7] R. Feldam and J. Sanger, *The Text Mining Handbook*, New York: Cambridge University Press, 2007.
- [8] M. S. Kini, Devi, D. P.G and N. Chiplunkar, "Text mining Approach to Classify Technical Research Document using Naïve Bayes," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 7, pp. 386-391, 2015.
- [9] A. Valdivia, M. V. Luzón and F. Herrera, "Sentiment analysis in tripadvisor," *IEEE Intelligent Systems*, vol. 4, no. 72-77, p. 32, 2017.
- [10] R. F. Clayton, J. J. Kopecky, D. S. Chou and A. J. Llorens, "Coarse-and Fine-Grained Sentiment Analysis of Social Media Text," *Johns Hopkins Apl Technical Digest*,

- vol. 30, no. 1, p. 50, 2011.
- [11] G. A, H. L and B. R, "Twitter sentiment analysis," *CS224N - Final Project Report*, 2009.
 - [12] Orange, <https://orange.biolab.si/>, Orange, [Online]. Available: <https://orange.biolab.si/#Orange-Features>. [Accessed 20 Agustus 2018].
 - [13] H. J, M. Kamber and J. Pei, "Data Mining Concept and Techniques3rd," *Elsivier Inc, USA*.
 - [14] I. Ahmed, D. Guan and T. Chung, SMS Classification Based on Naïve Bayes Classifier and Apriori Algorithm Frequent Itemset, *International Journal of Machine Learning and Computing* , vol. 4, no. 2, pp. 183-187, 2014.
 - [15] E. Yulian, Text Mining dengan K-Means Clustering pada Tema LGBT dalam Arsip Tweet Masyarakat Kota Bandung, *JURNAL MATEMATIKA "MANTIK"* , vol. 4, no. 1, pp. 53-58, 2018.
 - [16] Srihari. [Online]. Available: <https://cedar.buffalo.edu/~srihari/CSE626/Lecture-Slides/>. [Accessed 11 08 2018].
 - [17] F. Febrianti, M. Hafiyusholeh and A. Asyhar, "Perbandingan Pengklusteran Data Iris Menggunakan Metode K-Means dan Fuzzy C-Means," *Jurnal Matematika MANTIK*, vol. 2, no. 1, pp. 7-13, 2016.
 - [18] A. P. Putra, "www.medium.com," 18 2 2018. [Online]. Available: <https://medium.com/@mandes95/belajar-data-science-langkah-awal-mengenal-r-dan-rstudio-198ec2246f78>. [Accessed 10 08 2018].
 - [19] P. A. Putri, . A. Ridok and I. , "Implementasi Metode Improved K-Nearest Neighbor pada Analisis Sentimen Twitter Berbahasa Indonesia," *Repositori Jurnal Mahasiswa PTIIK UB*, vol. 2, no. 2, 2013.