

ANALISIS SENTIMEN TERHADAP BANTUAN KUOTA INTERNET DARI KEMENDIKBUD DIMASA COVID-19 MENGGUNAKAN ALGORITMA C5.0

ANALYSIS OF SENTIMENT ON INTERNET QUOTA ASSISTANCE FROM THE MINISTRY OF EDUCATION AND CULTURE DURING COVID-19 USING C5.0 ALGORITHM

Nur Maelani Asih¹, Jajam Haerul Jaman², Yuyun Umidah³

¹²³Universitas Singaperbangsa Karawang
nur.maelani17166@student.unsika.ac.id

ABSTRACT

The Ministry of Education and Culture quota is internet quota assistance provided by the Ministry of Education and Culture for students and teaching staff to facilitate online learning caused by the Covid-19 pandemic. The quota assistance was carried out because students complained about the increasing number of internet quotas being issued during the online learning period. The purpose of this study is to analyze the sentiment of Twitter social media users towards the Ministry of Education and Culture quota using the Knowledge Discovery in Databases (KDD) methodology and the C5.0 algorithm and test the algorithm. As well as for the performance results of the C5.0 algorithm using a confusion matrix. The results of the accuracy of C5.0 are 74%. This has increased from sentiment analysis research with previous twitter data, with an accuracy of 66%. However, in this study, the presentation for the neutral class on precision and recall and the negative class on recall is low because the training data used can be said to be unbalanced.

Keywords: C5.0, Kemendikbud Quota, Knowledge Discovery in Database Process.

ABSTRAK

Kuota kemendikbud merupakan bantuan kuota internet yang diberikan kemendikbud untuk para pelajar dan tenaga pengajar guna memfasilitasi pembelajaran daring yang diakibatkan pandemi Covid-19. Bantuan kuota dilakukan karena para pelajar yang mengeluhkan semakin banyak kuota internet yang dikeluarkan pada masa pembelajaran daring. Tujuan dari penelitian ini adalah melakukan analisis sentimen pengguna media sosial twitter terhadap kuota kemendikbud dengan menggunakan metodologi *Knowledge Discovery in Databases* (KDD) dan algoritma C5.0 dan menguji algoritmanya. Serta untuk hasil performa dari algoritma C5.0 menggunakan confusion matrix. Hasil akurasi C5.0 yaitu 74%. Ini mengalami kenaikan dari penelitian sentimen analisis dengan data twitter sebelumnya yaitu dengan akurasi sebesar 66%. Tetapi untuk pada penelitian ini presentasi untuk kelas netral pada *precision* dan *recall* serta kelas negatif pada *recall* rendah karena data training yang digunakan bisa dikatakan tidak *balance*.

Kata Kunci: C5.0, Kuota Kemendikbud, Knowledge Discovery in Database Process.

PENDAHULUAN

Di tahun 2020 ditemukan virus baru yang dinamakan virus *Covid-19*. Virus ini awalnya berasal dari Wuhan, China yang telah menyebar dengan sangat cepat ke negara-negara di seluruh dunia termasuk Indonesia. Untuk mengurangi angka penyebaran virus *Covid-19* pemerintah memberlakukan pembelajaran daring. Pembelajaran daring menimbulkan pro dan kontra karena keterbatasan media, jaringannya, dan biaya internet. Para pelajar

mengeluhkan tentang kuota internet yang mereka gunakan untuk pembelajaran daring, karena jadi lebih banyak mengeluarkan kuota internet.

Karena hal itu Kementerian Pendidikan dan Kebudayaan (Kemendikbud) jadi *leading sector* pemerintah yang menyalurkan bantuan subsidi kuota internet yang setiap bulannya diberikan kepada para pelajar dan tenaga pengajar. Di tahun 2020 tepatnya September sampai Desember 2020, ada dua jenis bantuan kuota

internet yaitu kuota belajar dan kuota umum. Karena banyak para pelajar yang mengeluhkan kuota belajar lebih banyak dari pada kuota umum maka pada Maret 2021 kemendikbud mengeluarkan bantuan kuota internet kembali dengan presentasi kuota umum semua tetapi dengan jumlah kuota yang lebih sedikit. Walaupun jumlah kuota internet diubah masih tetap ada pendapat yang pro dan kontra tentang kebijakan ini. Pendapat pro dan kontra dari para pelajar tentang pembagian kuota kemendikbud tersebut. Apalagi saat ini pembelajaran daring diperpanjang sampai bulan agustus karena adanya PPKM (Permbelakuan Pembatasan Kegiatan Masyarakat). Tetapi sampai saat ini kuota kemendikbud tidak diperpanjang kembali.

Media sosial *twitter* merupakan media social yang digunakan masyarakat dengan bebas dan mengungkapkan pendapat mereka secara bebas. Beberapa tahun terakhir, *twitter* memberikan pengaruh yang sangat besar dalam memberikan informasi. *Twitter* dalam data mining bisa dicari sesuatu yang menarik seperti bagaimana opini atau pendapat masyarakat terhadap sesuatu hal seperti kebijakan pemerintah contohnya kemendikbud dalam membagikan bantuan subsidi kuota internet. Terdapat banyak opini dengan sentimen negatif, positif, dan netral yang dituangkan para pelajar di *twitter*. Namun untuk dalam menghasilkan sentimen oleh para user *twitter* dibutuhkan waktu dan usaha yang memakan waktu lama karena banyaknya jumlah tweet yang digunakan. Dibutuhkan machine learning yang efektif dalam pengklasifikasian tweet dengan sentimen negatif, positif dan netral. Dengan analisis sentimen begitu diperlukan karena dapat membantu menjadi bahan pertimbangan untuk pemerintah dalam menanggapi sikap para pelajar.

Pada penelitian (Romadloni,

Santoso, & Budilaksono, 2019) yang melakukan analisis sentimen terhadap transportasi umum KRL dengan metode Decision Tree menghasilkan akurasi yang tinggi yaitu 100% dibanding dengan metode *Naïve Bayes Classifier* dan KNN dengan akurasi masing-masing metode yaitu sebesar 80%. Dalam penelitian (Balamurugan & Kannan, 2016) melakukan 2 kali percobaan dengan data set yang berbeda dan teknik sampling yang berbeda. Dari 2 kali percobaan dengan dataset yang berbeda, algoritma C5.0 menunjukkan akurasi yang tinggi semua yaitu sebesar 78,79% - 93,82%. Dan dalam penelitian (Albances. et al, 2018) melakukan penelitian dengan penerapan algoritma C5.0 untuk prediksi flu menggunakan data *twitter*, hasil ringkasan akurasi, notasi Big O, dan presisi dari kedua algoritma. Dari segi presisi dan efisiensi, algoritma Naive Bayes lebih baik dari pada algoritma C5.0. Namun, C5.0 lebih baik dalam hal akurasi yang menghasilkan 66%. Solusi yang diusulkan, Algoritma C5.0, memiliki akurasi yang lebih baik dibandingkan dengan pengklasifikasi *Naive Bayes*.

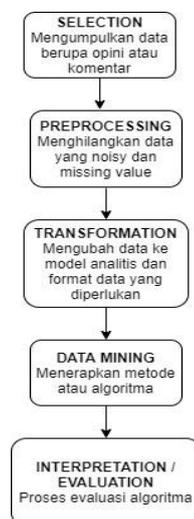
Algoritma C5.0 mampu mengatasi data berskala besar dengan baik dan mendapatkan nilai akurasi lebih tinggi dibandingkan dengan algoritma C4.5. (Revathy & Lawrence, 2019). Dibanding algoritma C4.5, algoritma C5.0 memiliki performa yang lebih cepat, juga penggunaan memori yang lebih efisien dan pohon keputusannya lebih kecil. Pada aturan C5.0 memiliki tingkat kesalahan yang lebih rendah untuk kasus yang tidak terlihat. C5.0 juga memiliki nilai akurasi tinggi dibanding dengan C4.5. algoritma C5.0 juga dapat memungkinkan secara otomatis dalam penghapusan atribut yang tidak membantu.

Tujuan penelitian ini adalah untuk meningkatkan performa algoritma C5.0 dalam hal analisis sentimen pada data

twitter dan mengklasifikasi sentiment berupa positif, negatif dan netral pada twitter untuk kata kunci “Kuota Kemendikbud”. Dengan menggunakan kata kunci “Kuota Kemendikbud” menghasilkan tweet yang berfokus pada “Kuota Kemendikbud”. Pada penelitian ini metode *Knowledge Discovery in Databases* (KDD) dan menggunakan algoritma C5.0, dan Algoritma TFIDF (*Term Frequency Invert Document Frequency*) digunakan untuk pembobotan fitur.

METODE

Dalam pengembangan penelitian ini, metodologi yang akan digunakan metode *Knowledge Discovery in Databases* (KDD). Dalam pengumpulan informasi dengan metodologi ini sangat baik dan benar dan juga dalam format yang benar dan dalam tujuan pengambilan keputusan waktu yang digunakan sangat tepat. Berikut metode *Knowledge Discovery in Databases* (KDD) terdapat pada Gambar 1.



Gambar 1. Metodologi Penelitian

Menurut (Fayyad, Piatetsky-Shapiro & Smith, 1996) tahap penelitian menggunakan metodologi *Knowledge Discovery in Databases* (KDD) terdiri dari:

1. Selection

Pada tahap *selection*, proses yang dilakukan adalah melakukan crawling data secara online dari sosial media twitter. Kemudian proses pelabelan untuk menentukan kalimat bernilai positif, negatif, dan netral. Pelabelan dilakukan secara manual oleh dosen bahasa Indonesia IT Telkom Purwokerto.

2. Preprocessing

Pada tahap *preprocessing* yaitu menghilangkan noisy dan *missing value*. Di *preprocessing* ada 5 tahapan yaitu:

1) Cleaning

Cleaning untuk menghilangkan tanda baca seperti titik, koma, dll serta angka dalam *tweet* dan komponen lainnya seperti HTML, URL dan hastag.

2) Case Folding

Case folding yaitu tahap mengubah bentuk huruf agar menjadi bentuk huruf yang sama. *Case folding* yang dilakukan pada penelitian ini yaitu mengubah bentuk huruf menjadi huruf kecil semua atau *lower case*.

3) Tokenizing

Tokenizing yaitu proses pemisahan kalimat menjadi kata per kata yang saling terpisah. Tujuannya agar didapat potongan kata yang akan dijadikan sebagai entitas yang memiliki nilai matriks dokumen yang akan dianalisis.

4) Filtering

Tahap filtering dilakukan karena untuk pemilihan kata pada suatu dokumen teks atau dimensi kata dalam corpus dikurangi, yang sering disebut stopword. Stopwords yaitu tahap dimana kata-kata yang tidak informatif tetapi sering muncul untuk dihilangkan.

5) Stemming

Proses stemming yaitu proses untuk mencari kata dasar pada setiap kata dengan menghilangkan imbuhan pada kata tersebut. Proses ini dilakukan

dengan menghilangkan imbuhan pada awal dan akhir pada suatu kata.

3. Transformation

Tahap transformation dalam penelitian ini menggunakan pembobotan TF-IDF. Pada pembobotan term akan dilakukan pemberian bobot atau nilai terhadap setiap *term* yang terkandung dalam setiap *tweet* yang sudah dilakukan tahap *preprocessing*. Pembobotan ini bertujuan untuk memberikan nilai pada suatu term yang dimana nilai term tersebut akan dijadikan sebagai input pada proses klasifikasi.

4. Data Mining

Pada proses data mining yaitu mengimplementasikan algoritma C5.0 tersebut dilakukan dengan menggunakan *library* pada R studio.

5. Evaluation

Evaluation yaitu melakukan pengukuran performa terhadap model yang telah diimplementasikan. Yang jadikan parameter untuk *confussion matrik* yaitu *accuracy*, *precission*, *recall* dan *f-measure*.

HASIL DAN PEMBAHASAN

Hasil dari penelitian yang dilakukan adalah sentimen analisis sentimen terhadap kuota kemendikbud dengan metode C5.0. berikut ini pembahasan mengenai tahapan penelitian yang telah dilakukan:

1. Selection

Mengumpulkan data tweet yang diperoleh dengan menggunakan *stream twitter* API dengan R studio. Kemudian lanjut proses pelabelan. Hasil *crawling* data dan pelabelan seperti pada Gambar 2.

| user_id | text | label |
|----------|--|---------|
| x1278260 | Kuota Kemendikbud lekaslah cair <U+0001F62 | positif |
| x1337190 | Kebiasaan boros kuota edukasi kemendikbud | negatif |
| x5846678 | Ini kuota kemendikbud emg ga turun lagi apa | negatif |
| x1234773 | Kuota kemendikbud cepet turun deh. PLN sini | positif |
| x1252852 | senin dah daring tapi kuota kemendikbud dah | netral |
| x1247319 | Kuota kemendikbud plsssss cepetann turun, b | positif |
| x1263045 | pantesan kuota gue cepet abiss gataunya uda | netral |
| x1223806 | bulan ini paling boros keknya deh pusinggg m | negatif |
| x3268958 | Anjir sampe sekarang kaga dikasih2 itu kuota | negatif |

Gambar 2. Data Tweet dan Label

2. Preprocessing

Dalam preprocessing ada 5 tahap yaitu:

1) Cleaning

Berikut hasil dari data *tweet* sebelum *cleaning* dan sesudah *cleaning* pada Gambar 3 dan Gambar 4.

| |
|--|
| Kuota Kemendikbud lekaslah cair <U+0001F623> |
| Kebiasaan boros kuota edukasi kemendikbud jadi keterusan ke kuota berbayar deh... https |
| Ini kuota kemendikbud emg ga turun lagi apa gmn... |
| Kuota kemendikbud cepet turun deh. PLN sini cupu banget gerimis dikit langsung mati listri |
| senin dah daring tapi kuota kemendikbud dah expired |
| pantesan kuota gue cepet abiss gataunya udah ga dapat subsidi dari kemendikbud |
| bulan ini paling boros keknya deh pusinggg mana kuota kemendikbud kaga diperpanjang |

Gambar 3. Teks tweet Sebelum Cleaning

| |
|--|
| kuota kemendikbud lekaslah cair |
| kebiasaan boros kuota edukasi kemendikbud jadi keterusan ke kuota berbayar deh |
| ini kuota kemendikbud emg ga turun lagi apa gmn |
| kuota kemendikbud cepet turun deh pln sini cupu banget gerimis dikit langsung mati listrik |
| senin dah daring tapi kuota kemendikbud dah expired |
| pantesan kuota gue cepet abiss gataunya udah ga dapat subsidi dari kemendikbud |
| bulan ini paling boros keknya deh pusinggg mana kuota kemendikbud kaga diperpanjang |

Gambar 4. Hasil Cleaning

2) Case Folding

Mengubah teks tweet menjadi *lower case* semua. Berikut hasil sebelum dilakukan *case folding* dan sesudah dilakukan *case folding* terlihat pada Gambar 5 dan Gambar 6.

| |
|--|
| Kuota Kemendikbud lekaslah cair |
| Kebiasaan boros kuota edukasi kemendikbud jadi keterusan ke kuota berbayar deh. |
| Ini kuota kemendikbud emg ga turun lagi apa gmn... |
| Kuota kemendikbud cepet turun deh. PLN sini cupu banget gerimis dikit langsung n |
| senin dah daring tapi kuota kemendikbud dah expired |
| pantesan kuota gue cepet abiss gataunya udah ga dapat subsidi dari kemendikbud |
| bulan ini paling boros keknya deh pusinggg mana kuota kemendikbud kaga diperp. |

Gambar 5. Teks Tweet Sebelum

Case Folding

| |
|--|
| kuota kemendikbud lekaslah cair |
| kebiasaan boros kuota edukasi kemendikbud jadi keterusan ke kuota berbayar deh |
| ini kuota kemendikbud emg ga turun lagi apa gmn |
| kuota kemendikbud cepet turun deh pln sini cupu banget gerimis dikit langsung mati listrik |
| senin dah daring tapi kuota kemendikbud dah expired |
| pantesan kuota gue cepet abiss gataunya udah ga dapet subsidi dari kemendikbud |
| bulan ini paling boros keknya deh pusinggg mana kuota kemendikbud kaga diperpanjang |

Gambar 6. Hasil Case Folding

3) Tokenizing

Memisahkan kalimat menjadi kata per kata yang saling terpisah. Berikut hasil dari tokenizing dapat dilihat pada Gambar 7.

| | | |
|-----------------------------------|---------|---|
| kuota kemendikbud lekaslah cair | positif | list(("kuota", "kemendikbud", "lekaslah", "cair")) |
| kebiasaan boros kuota edukasi... | negatif | list(("kebiasaan", "boros", "kuota", "edukasi", "kemendikbud", "jadi", "keterusan", "ke", "kuota", "berbayar", "deh")) |
| ini kuota kemendikbud emg g... | negatif | list(("ini", "kuota", "kemendikbud", "emg", "ga", "turun", "lagi", "apa", "gmn")) |
| kuota kemendikbud cepet turu... | positif | list(("kuota", "kemendikbud", "cepat", "turun", "deh", "pln", "sini", "cupu", "banget", "gerimis", "dikit", "langsung", "mati", "listrik")) |
| senin dah daring tapi kuota ke... | netral | list(("senin", "dah", "daring", "tapi", "kuota", "kemendikbud", "dah", "expired")) |
| pantesan kuota gue cepet abis... | netral | list(("pantesan", "kuota", "gue", "cepat", "abis", "gataunya", "udah", "ga", "dapet", "subsidi", "dari", "kemendikbud")) |
| bulan ini paling boros keknya ... | negatif | list(("bulan", "ini", "paling", "boros", "keknya", "deh", "pusinggg", "mana", "kuota", "kemendikbud", "kaga", "diperpanjang")) |

Gambar 7. Hasil Tokenizing

4) Filtering

Tahap *filtering* dilakukan untuk menghasilkan *stopwords*. Sistem akan mengambil data dari dokumen *stopword.txt* atau nama file dokumen dalam penelitian ini adalah *stoplist.txt*, kemudian sistem akan secara otomatis mengecek apakah data yang ada di file *stoplist.txt* ada di dalam data tweet, jika mengandung kata-kata tersebut dalam *tweet* maka kata-kata tersebut akan otomatis dihapus. Hasil *stopwords* dapat dilihat pada Gambar 8.

ada
adalah
adanya
adapun
agak
agakny
agar
akan
akankah
akhir

Gambar 8. Hasil Stopwords

5) Stemming

Menghilangkan imbuhan pada awalan dan akhiran kata. Berikut hasil sebelum proses *stemming* dan sesudah proses *stemming* dapat dilihat pada Gambar 9 dan Gambar 10.

| | | |
|-----------------------------------|---------|---|
| kuota kemendikbud lekaslah cair | positif | list(("kuota", "kemendikbud", "lekaslah", "cair")) |
| kebiasaan boros kuota edukasi... | negatif | list(("kebiasaan", "boros", "kuota", "edukasi", "kemendikbud", "jadi", "keterusan", "ke", "kuota", "berbayar", "deh")) |
| ini kuota kemendikbud emg g... | negatif | list(("ini", "kuota", "kemendikbud", "emg", "ga", "turun", "lagi", "apa", "gmn")) |
| kuota kemendikbud cepet turu... | positif | list(("kuota", "kemendikbud", "cepat", "turun", "deh", "pln", "sini", "cupu", "banget", "gerimis", "dikit", "langsung", "mati", "listrik")) |
| senin dah daring tapi kuota ke... | netral | list(("senin", "dah", "daring", "tapi", "kuota", "kemendikbud", "dah", "expired")) |
| pantesan kuota gue cepet abis... | netral | list(("pantesan", "kuota", "gue", "cepat", "abis", "gataunya", "udah", "ga", "dapet", "subsidi", "dari", "kemendikbud")) |
| bulan ini paling boros keknya ... | negatif | list(("bulan", "ini", "paling", "boros", "keknya", "deh", "pusinggg", "mana", "kuota", "kemendikbud", "kaga", "diperpanjang")) |

Gambar 9. Sebelum Proses Stemming

| text | label | token |
|------------------------------|---------|---|
| kuota kemendikbud leka... | positif | list(("kuota", "kemendikbud", "lekas", "cair")) |
| biasa boros kuota eduka... | negatif | list(("boros", "kuota", "edukasi", "kemendikbud", "kuota", "bayar", "deh")) |
| ini kuota kemendikbud e... | negatif | list(("kuota", "kemendikbud", "emg", "ga", "turun", "gmn")) |
| kuota kemendikbud cep... | positif | list(("kuota", "kemendikbud", "cepat", "turun", "deh", "pln", "cupu", "banget", "gerimis", "dikit", "langsung", "mati", "listrik")) |
| senin dah daring tapi ku... | netral | list(("senin", "dah", "daring", "tapi", "kuota", "kemendikbud", "dah", "expired")) |
| pantesan kuota gue cep... | netral | list(("pantesan", "kuota", "gue", "cepat", "abis", "gataunya", "udah", "ga", "dapet", "subsidi", "dari", "kemendikbud")) |
| bulan ini paling boros ke... | negatif | list(("boros", "kek", "deh", "pusinggg", "kuota", "kemendikbud", "kaga")) |

Gambar 10. Hasil Stemming

3. Transformation

Tahap transformation menggunakan pembobotan TF-IDF dengan tujuan untuk memberikan nilai pada suatu term yang dimana nilai term tersebut akan dijadikan sebagai input pada proses klasifikasi. Berikut hasil hitungan dari TF-IDF dapat dilihat pada Gambar 11.

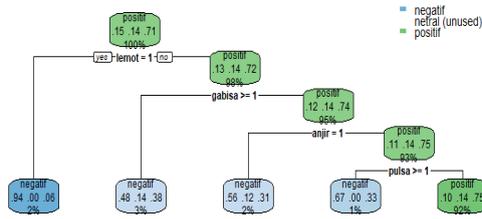
| label | word | n | total | tf | idf | tf_idf |
|---------|--------|---|-------|--------------|-----------|--------------|
| positif | bentar | 1 | 5134 | 0.0001947799 | 1.0986123 | 2.139876e-04 |
| positif | berapa | 1 | 5134 | 0.0001947799 | 1.0986123 | 2.139876e-04 |
| positif | berat | 1 | 5134 | 0.0001947799 | 1.0986123 | 2.139876e-04 |
| positif | beri | 1 | 5134 | 0.0001947799 | 1.0986123 | 2.139876e-04 |
| negatif | kudu | 1 | 1234 | 0.0008103728 | 1.0986123 | 0.0008902855 |
| negatif | kurang | 1 | 1234 | 0.0008103728 | 0.4054651 | 0.0003285779 |

Gambar 11. Perhitungan Nilai TF-IDF

4. Data Mining

Implementasi algoritma C5.0. tahap pertama yaitu mengambil *dataset* yang akan dijadikan untuk data training, kemudian lakukan 4 proses yaitu *randomize data*, *vectorize data*, *split data*, dan *create model*. Pada proses klasifikasi pada penelitian ini menggunakan data *test* yang diambil secara acak yaitu data train 80% dan data

test 20%. Berikut hasil pohon keputusan model dari algoritma C5.0 dapat dilihat pada Gambar 12.



Gambar 12. Pohon Keputusan Model Algoritma C5.0

Gambar diatas menunjukkan pohon keputusan yang didapatkan pada model, terdapat beberapa rules informasi yang didapatkan dari model pohon keputusan tersebut diantaranya sebagai berikut:

1. Rules 1, jika terdapat data kata lemt !=1, gabisa <= 1, anjir !=1, pulsa <=1 maka kalimat tersebut dari hasil informasi pohon keputusan yaitu menunjukkan kalimat positif bernilai 92%.
2. Rules 2, jika terdapat data kata lemt =1, gabisa >= 1, anjir = 1, pulsa >= 1 maka kalimat tersebut dari hasil informasi pohon keputusan yaitu menunjukkan kalimat negative bernilai 1 %.

Untuk kelas netral *unused* atau tidak terpakai karena tidak ada *datatest* yang digunakan untuk kelas netral, hal ini disebabkan data yang tidak *balance*. Model pohon keputusan tersebut untuk memberikan informasi yang dihasilkan jika dimasukkan data baru ke dalam model sehingga dapat menghasilkan informasi kalimat tersebut bernilai positif, negatif atau netral. Setelah itu dapat jelaskan prediksi dari data yang dimasukkan dan menghitung performa dari model algoritma C5.0 yang digunakan.

5. Evaluation

Menghitung performa dari model algoritma C5.0 dengan *confusion matrix* dengan parameter *accuracy*, *precission*, *recall* dan *f-measure*.

Berikut hasil dari *confusion matrix* model C5.0 dapat dilihat pada Gambar 13.

| | p | | |
|---------|---------|--------|---------|
| testy | negatif | netral | positif |
| negatif | 9 | 0 | 12 |
| netral | 0 | 1 | 23 |
| positif | 3 | 5 | 115 |

Gambar 13. Hasil Confusion Matrix Model C5.0

Dari hasil *confusion matrix* yang ditunjukkan pada Gambar 4.18 didapat hasil nilai akurasi dari model, nilai *precision*, *recall*, dan *f-measure* berdasarkan nilai-nilai berikut:

- a. *True Positive* (TP) yaitu data positif yang diprediksi benar atau dimana kondisi jumlah dari kelas *True* yang dapat diprediksi dengan benar pada kelas *True*
- b. *True Negative* (TN) yaitu data negatif yang diprediksi benar atau dimana jumlah kelas *False* yang dapat diprediksi dengan benar pada kelas *False*.
- c. *False Positive* (FP) yaitu data negatif namun diprediksi sebagai positif atau dimana kelas *True* yang diprediksi salah pada kelas *False*.
- d. *False Negative* (FN) yaitu data positif namun diprediksi sebagai data negatif atau dimana kelas *False* yang diprediksi salah pada kelas *True*.

Berikut perhitungan nilai *accuracy*, *precision*, *recall*, dan *f-measure* berdasarkan *confusion matrix* diatas yaitu sebagai berikut:

$$\begin{aligned}
 TP &= 115 + 1 + 9 = 125 \\
 Accuracy &= \frac{TP}{Jumlah\ Data} = \frac{125}{168} = 0,74 \\
 &= 74\% \\
 Precision &= \frac{TP}{TP+FP} = \frac{115}{115+(23+12)} = \frac{115}{150} = 0,77 \\
 &= \frac{TP}{TP+FP} = \frac{1}{1+(0+5)} = \frac{1}{6} = 0,17 \\
 &= \frac{TP}{TP+FP} = \frac{9}{9+(3+0)} = \frac{9}{12} = 0,75 \\
 Precision &= \frac{0,77+0,17+0,75}{3} = \frac{1,69}{3} = 0,56 \text{ atau } 56\% \\
 Recall &= \frac{TP}{TP+FN} = \frac{115}{115+(0+3)} = \frac{115}{118} = 0,97 \\
 &= \frac{TP}{TP+FN} = \frac{1}{1+(23+0)} = \frac{1}{24} = 0,04
 \end{aligned}$$

$$Recall = \frac{TP}{TP+FN} = \frac{9}{9+(12+5)} = \frac{9}{26} = 0,34$$

$$Recall = \frac{0,94+0,04+0,34}{3} = \frac{1,35}{3} = 0,45 \text{ atau } 45\%$$

F-Measure

$$= \frac{2 \times recall \times precision}{recall + precision} = \frac{2 \times 0,97 \times 0,77}{0,97+0,77} = 0,85$$

$$= \frac{2 \times recall \times precision}{recall + precision} = \frac{2 \times 0,04 \times 0,17}{0,04+0,17} = 0,064$$

$$= \frac{2 \times recall \times precision}{recall + precision} = \frac{2 \times 0,34 \times 0,75}{0,34+0,75} = 0,46$$

F-Measure

$$= \frac{0,85+0,064+0,46}{3} = 0,45 \text{ atau } 45\%$$

Untuk akurasi algoritma C5.0 juga bisa dihitung dengan dengan program seperti pada Gambar 14.

```
> # Accuracy
> sum(p==testY) / length(p)
[1] 0.7440476
```

Gambar 14. Hasil Akurasi Algoritma C5.0

Hasil *precision*, *recall* dan *f-measure* disetiap kelasnya dapat dilihat pada Tabel 1.

Tabel 1. Nilai Precision, Recall, dan F-Measure

| Jenis Klasifikasi | Precision | Recall | F-Measure |
|--------------------------|-----------|--------|-----------|
| Positif | 0,77 | 0,97 | 0,85 |
| Netral | 0,17 | 0,04 | 0,064 |
| Negatif | 0,75 | 0,34 | 0,46 |
| Rata-rata Akurasi | 0,56 | 0,45 | 0,45 |

Hasil nilai akurasi algoritma C5.0 adalah **74%**. Sedangkan hasil dari evaluasi model dapat dilihat nilai *precision* dan *recall* di setiap kelasnya. Dapat dikatakan tingkat kemampuan sistem dalam mencari ketepatan antara informasi yang diminta oleh pengguna (*precision*) untuk kelas positif sebesar **77%**, kelas negatif sebesar **75%** dan untuk kelas netral sebesar **17%**, **17%** termasuk rendah itu berarti tingkat kemampuan sistem dalam mencari ketepatan antara informasi yang diminta oleh pengguna kelas netral rendah. Sedangkan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi (*recall*) untuk kelas positif

sebesar **97%**, untuk kelas netral sebesar **0,04%**, kelas negatif sebesar **34%**. Untuk prosentase tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi untuk kelas netral sebesar **0,04%**, kelas negatif sebesar **34%** termasuk rendah itu artinya kinerja sistem keberhasilan sistem dalam menemukan kembali sebuah informasi yang bernilai netral dan negatif dalam dokumen rendah.

KESIMPULAN

Berdasarkan hasil pengujian algoritman C5.0 yang dilakukan berikut beberapa hasil hal yang dihasilkan adalah sebagai berikut:

1. Analisis sentimen menggunakan algoritma C5.0 yang dikerjakan dengan R Studio sesuai dengan metodologi *Knowledge Discovery in Databases* (KDD), dari selection data sampai evaluasi performa model algoritma C.50. C.50 digunakan untuk pengklasifikasian kelas menjadi kelas negatif, positif, serta netral.
2. Pada penelitian ini, algoritma C5.0 terbukti algoritma yang cukup akurat karena menghasilkan nilai akurasi sebesar 0,74 atau 74%. Mengalami peningkatan dari penelitian sebelumnya yaitu 66%. Untuk nilai rata-rata dari *precision* 56% serta *recall* dan *f-measure* sama 45%. Pada nilai akurasi *precision*, *recall*, dan *f-Measure* pada kelas positif memiliki akurasi yang tinggi, itu berarti data *tweet* atau pendapat pengguna *twitter* mengenai bantuan kemendikbud menanggapi dengan positif atau berguna bagi mereka. Sehingga diharapkan kuota kemendikbud untuk diperpanjang kembali, apalagi belajar daring diperpanjang sampai waktu yang tidak ditentukan.

Dari hasil penelitian ini masih banyak kekurangan, dengan demikian peneliti berharap penelitian ini untuk

dikembangkan, beberapa saran dari penulis, di antaranya sebagai berikut:

1. Sebaiknya menggunakan feature selection karena memiliki kemampuan mengurangi dimensionalitas suatu data sehingga dapat meningkatkan performansi.
2. *Recall* pada kelas netral dan negatif, yaitu masing-masing hanya sebesar 0.04% dan 34%. Hal itu disebabkan karena jumlah data training pada kelas netral dan negatif lebih sedikit dibandingkan dengan kelas positif atau bisa dikatakan data training yang digunakan pada penelitian ini tidak balance. Begitu juga *precision* untuk kelas netral juga rendah yaitu hanya 17%. Sehingga sebaiknya untuk menambah data training dan mengupayakan untuk menyeimbangkan jumlah data di setiap kelasnya untuk penelitian selanjutnya.
3. Membuat *interface* dari proses pengujian model dan visualisasi dari performa metode yang digunakan.

DAFTAR PUSTAKA

- Albances, L. Z., Bungar, B. A., Patio, J. P., Sevilla, R. J. M., & Acula, D. (2018, January). Application of C5.0 Algorithm to Flu Prediction Using Twitter Data. In *2018 International Conference on Platform Technology and Service (PlatCon)* (1-6). IEEE.
- Balamurugan. M & Kannan. S. (2016). Performance Analysis of Cart and C5.0 using Sampling Techniques. *International Conference on Advances in Computer Applications (ICACA)*. 72-75.
- Debora Yantina. (2020). *Kuota Gratis Kemendikbud: Syarat, Cara Daftar dan Rincian Bantuan*. Retrieved from tirto.id website: <https://tirto.id/kuota-gratis-kemendikbud-syarat-cara-daftar-dan-rincian-bantuan-f2ZU>.
- Dhara. A. F., (2017). Rancang Bangun Ekstraksi Ekspresi Kata Kerja Pada Ulasan Pelanggan Dengan Text Chunking Untuk Memaparkan Pengalaman Penggunaan Produk. Skripsi: Teknologi Informasi, Institut teknologi Sepuluh Nopember.
- Efendi. (2020). *Apa itu Kuota Internet? Mengenal Pengertian Kuota Internet*. Fayyad UM, Piatetsky-Shapiro G & Smith P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27–34.
- Ferdiana, R., Jatmiko, F., Purwanti, D.D., Ayu, A.S.T., & Dicka, W.F. (2019). Dataset Indonesia untuk Analisis Sentimen. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, 8(4), 334-339.
- Fitriyyah. S. N. j., Safriadi. n., & Pratama. E. E. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *Jurnal Edukasi dan Penelitian Informatika*, 5(3), 279-285.
- Imron, A. (2019). Analisis Sentimen Terhadap Tempat Wisata Di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier. Skripsi: Teknologi Industri, Universitas Islam Indonesia.
- Informatikalogi. (2016). *Pembobotan Kata atau Term Weighting TF-IDF*. Retrieved April 12, 2018, from [Informatikalogi.com](https://informatikalogi.com): <https://informatikalogi.com/term-weighting-tf-idf/>
- Kementerian dan Pendidikan dan Kebudayaan RI. (2020). *Bantuan Kuota Data Internet 2020*. Retrieved from <https://kuotabelajar.kemdikbud.go.id/index.html>.
- Larose, D. T. (2005). *Discovering knowledge in Data: An Introduction to*

- Data Mining*. Canada: John Willey & Sons, Inc.
- Mujilawati, S., (2016). Pre-processing text mining pada data twitter. Seminar *Nasional Teknologi Informasi dan Komunikasi*, 49-56.
- Pratiwi. R., Hayati. M. N., Prangga, S., & Budilaksono, S. (2019). Perbandingan Klasifikasi Algoritma C5.0 Dengan Classification And Regression Tree (Studi Kasus: Data Sosial Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan Muara Ancalong Tahun 2019). *Jurnal Ilmu Matematika dan Terapan*, 14(2), 267-278.
- Retrieved from Nesabamedia website: <https://www.nesabamedia.com/apa-itu-kuota-internet/>.
- Revathy. R & Lawrance. R. (2017). Comparative Analysis of C4.5 and C5.0 Algorithms on Crop Pest Data. *International Journal of Innovative Research in Computer and Communication Engineering*, 5(1), 50-58.
- Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes , Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl. *Jurnal IKRA-ITH Informatika*, 3(2), 1-9.
- Siregar. R. P., (2016). *Klasifikasi Jenis Pengaduan Konsumen Dari Layanan Media Sosial Dengan Menggunakan Text Mining*. Skripsi: Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara.