

ANALISIS SENTIMEN CYBERBULLYING KPOP DI MEDIA SOSIAL TWITTER MENGGUNAKAN METODE NAIVE BAYES

Anisa Putri, Ari Muzakir

Jurusan Sistem Informasi Fakultas Ilmu Komputer, Universitas Bina Darma, Palembang, Indonesia

Email: nssaptri22@gmail.com, arimuzakir@binadarma.ac.id

Abstrak

Maraknya penggunaan jejaring sosial twitter dikalangan masyarakat modern saat ini tengah sangat populer. Twitter membawa trend baru dalam masyarakat sebagai ajang untuk melakukan tindakan penindasan secara online atau yang lebih dikenal dengan sebutan cyberbullying. Pengguna twitter dengan mudah dapat melakukan cyberbullying kepada pengguna twitter lainnya, pengguna dapat memposting tulisan kejam atau mengunggah foto yang berhubungan dengan individu lain dengan tujuan mengintimidasi dan merusak nama baik seseorang. *Cyberbullying* mudah ditemukan pada penggemar *Korean Wave/Hallyu* sebab aktivitas terbesar penggemar *K-Pop* ada di internet. *K-Pop* memiliki banyak sekali penggemar tidak hanya di negara Korea Selatan saja, tetapi juga ke seluruh dunia termasuk di Indonesia terutama di kalangan remaja. Fokus permasalahan pada memahami dan mencegah *Cyberbullying* pada remaja penggemar *K-Pop* tujuannya memaparkan dan mengaalisa pentingnya penggemar *K-Pop* yang di dominasi oleh remaja, mendapatkan pemahaman tentang *Cyberbullying* baik sebagai pelaku maupun korban. Dalam penelitian ini bertujuan untuk untuk mengklasifikasikan sentimen.pada Komentar tweet berdasarkan kelas negatif, positif, netral dan juga untuk mengetahui trend yang ada di kalangan remaja dalam tindakan terhadap bullying di sosial media. Hasil dari penelitian ini menunjukkan bahwa dalam jejaring sosial twitter, perilaku cyberbullying terjadi karena motivasi yang ada pada dalam diri informan, seperti motif sosiogenis dan motif afektif. Peneliti menggunakan metode *Naive bayes* untuk melakukan klasifikasi sentiment cyberbullying pada sosial media *Twitter*.

Kata Kunci: *K-Pop, Cyberbullying, Twitter, analisis sentimen, naive bayes*

Abstract

The rampant use of the social network twitter among modern society is currently very popular. Twitter brings a new trend in society as a place to carry out acts of oppression online or better known as cyberbullying. Twitter users can easily cyberbully other twitter users, users can post cruel posts or upload photos related to other individuals with the aim of intimidating and damaging someone's good name. Cyberbullying is easy to find on Korean Wave/Hallyu fans because the biggest activity of K-Pop fans is on the internet. K-Pop has a lot of fans not only in South Korea, but also all over the world, including in Indonesia, especially among teenagers. Focusing on the problem of

understanding and preventing Cyberbullying in adolescent K-Pop fans aims to explain and analyze the importance of K-Pop fans who are dominated by teenagers, gaining an understanding of Cyberbullying both as perpetrators and victims. In this study aims to classify sentiments on Tweet comments based on negative, positive, neutral classes and also to find out the trends that exist among adolescents in actions against bullying on social media. The results of this study show that in the social network twitter, cyberbullying behavior occurs due to the motivations that exist in the informant, such as sociogenic motives and affective motives. Researchers used the Naive bayes method to classify sentiment cyberbullying on twitter social media.

Keywords: K-Pop, Cyberbullying, Twitter, sentiment analysis, naive bayes

Pendahuluan

Media sosial merupakan platform media yang memfokuskan pada eksistensi pengguna serta memfasilitasi mereka dalam beraktivitas maupun berkolaborasi. Karena itu media sosial dapat dilihat sebagai medium (fasilitator) online yang menguatkan hubungan antar pengguna sekaligus sebuah ikatan sosial. Melalui media sosial yang semakin banyak berkembang memungkinkan informasi menyebar dengan mudah di masyarakat. Informasi dalam bentuk apa pun dapat disebarluaskan dengan mudah dan cepat sehingga mempengaruhi cara pandang, gaya hidup, serta budaya suatu bangsa (Manning et al., 2008).

Salah satu media sosial yang sangat populer saat ini yaitu Twitter. Twitter merupakan media sosial yang berbentuk microblogging atau ngeblog secara singkat dalam satu paragraf dengan maksimal 280 huruf, karena jumlah huruf dalam satu kali tweet terbatas/dibatasi. Pengguna Twitter terdiri dari berbagai kalangan, seperti pejabat, selebritis, artis, hingga masyarakat biasa pada umumnya. Akan tetapi, kemudahan yang diberikan untuk berbagi informasi melalui media sosial tak luput dari penyalahgunaan yang dilakukan oleh penggunanya. Salah satu bentuk penyalahgunaan tersebut yaitu cyberbullying.

Cyberbullying merupakan bentuk tindak kekerasan yang dilakukan oleh seseorang terhadap korbannya di dunia maya, dimana korban dihina, diejek, dipermalukan dan diintimidasi oleh pelaku. Dampak dari cyberbullying yaitu mempengaruhi mental korban, bahkan banyak dari korban bullying berakhir dengan bunuh diri karena tidak tahan dengan banyak tekanan. Ada hubungan positif dan signifikan antara perilaku pelaku cyberbullying dan perilaku korban cyberbullying yang mana semakin reaktif perilaku pelaku cyberbullying maka semakin reaktif pula perilaku korban cyberbullying. Hal ini membuktikan betapa besarnya pengaruh cyberbullying terhadap kehidupan social (Khaira et al., 2020).

Oleh karena itu, sampai Saat ini, tidak ada data statistik yang tersedia informasi konkret tentang anak-anak yang terkena dampak kasus cyberbullying di Indonesia. Interaksi di internet tersebut membuat anak-anak dan remaja khususnya remaja penggemar K-Pop, rentan menjadi korban bahkan pelaku cyberbullying di media sosial, mengingat aktivitas yang mereka lakukan banyak dihabiskan di internet dan media

sosial. Kekhawatiran tersebut juga tergambar dalam Survei Penetrasi Internet dan Perilaku Pengguna Internet di Indonesia 2018 yang dirilis oleh Asosiasi Penyelenggara Jasa Internet di Indonesia (APJII) yang menunjukkan, 49% pengguna internet pernah dirisak (di-bully) dalam bentuk diejek atau dilecehkan di media sosial. hal tersebut menjadikan baik orang dewasa, remaja maupun anak-anak rentan untuk mendapatkan perlakuan negatif tersebut bahkan menjadi pelaku.”

K-Pop juga dapat dikenali dengan munculnya fans club K-Pop di media sosial yang menyediakan semua informasi tentang artis K-Pop kepada para penggemar K-Pop. Di Indonesia, khususnya di kalangan remaja. Oleh karena itu, pendekatan pembelajaran mesin dapat digunakan untuk mendeteksi yang mengandung unsur intimidasi. Artinya, menganalisa sentimen. Klasifikasi membutuhkan pendekatan pembelajaran mesin yang dapat membedakan antara kata-kata yang mengandung cyberbullying dan yang tidak (Ortega et al., 2008).

Analisis sentimen merupakan sub bagian dari Natural Language Processing (NLP) yang fokus pada menentukan perasaan yang terkandung pada sebuah teks. Analisis sentimen dikenal dengan opinion mining yang merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat. Ide dasar dari analisis setimen adalah untuk mendeteksi polaritas teks pada dokumen, kalimat, dan tweet. Polaritas sentimen terbagi tiga yaitu positif, negatif, dan netral.

Berbagai penelitian terkait analisis sentimen telah banyak dilakukan. Terdapat dua pendekatan untuk melakukan analisis sentimen, pendekatan yang pertama adalah berbasis machine learning yaitu dengan melatih classifier pada dataset yang telah dilabelkan secara manual. Pendekatan yang kedua adalah berbasis leksikal yang tidak memerlukan pelatihan dataset, mengukur polaritas suatu kalimat atau dokumen berdasarkan pada sentimen kata-kata dan frasa-frasa sambil menerapkan aturan-aturan tertentu yang diambil dari linguistik. Penelitian ini akan menganalisis sentimen terkait cyberbullying dari komentar masyarakat pada media sosial Twitter, untuk itu diperlukan metode yang dapat mengklasifikasikan komentar ke dalam kelas positif, negatif, dan netral. Kelas negatif berarti komentar yang mengandung elemen cyberbullying, kelas positif berarti komentar mengandung unsur motivasi atau dukungan, dan kelas netral adalah komentar yang tidak mengandung elemen cyberbullying.

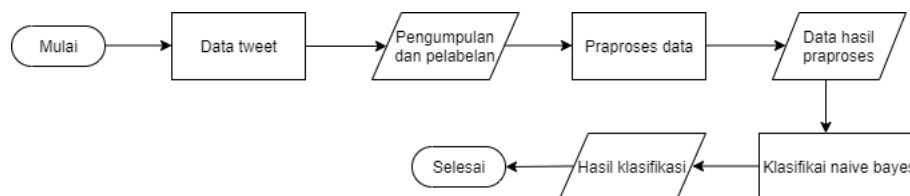
Dalam penelitian sebelumnya, SentiStrength dievaluasi pada beberapa dataset yang terdiri dari teks bahasa Inggris. Dalam penelitian ini, kami mereplikasi studi sebelumnya untuk teks bahasa Indonesia. Penelitian ini menggunakan metode SentiStrength berbasis kamus/leksikon. Kamus/leksikon SentiStrength berisi terms serta bobot kekuatan sentimennya. Metode SentiStrength menggunakan daftar idiom dan emotikon serta beberapa aturan lain seperti peningkatan kekuatan sentimen ketika huruf kapital digunakan (Nurzahputra & Muslim, 2016).

Metode Penelitian

Penelitian tentang analisis sentimen dari tweet twitter tentang keyword Cyberbullying, Kpop, dan Bullying. Secara umum suatu analisis sentimen digunakan untuk memprediksi kecenderungan suatu opini atau pendapat masyarakat terhadap sesuatu hal, apakah cenderung ke opini positif, netral atau negatif. Pada penelitian ini di buat sebuah dashboard berbasis web dimana fitur dari dashboard tersebut berfungsi untuk memprediksi opini atau pendapat user dengan cara live sehingga pengguna lebih mudah mengetahui hasil analisis sentimen dengan keyword yang di inginkan. Selain dengan keyword pengguna juga bisa melakukan input dengan hastag untuk mengetahui hasil sentimen analisis. Data yang dikumpulkan dapat digunakan untuk analisis sentimen. Analisis sentimen memerlukan beberapa langkah untuk mendapatkan hasil pengujian. Di bawah ini adalah langkah-langkah yang saya gunakan.

Pengumpulan Data

Tahap pertama dalam melakukan proses analisis sentimen adalah pengumpulan data. Data dari Twitter dengan 2000 Cari "cyberbullying, kpop, dan bullying" menggunakan aplikasi Rapidminer. Twitter Search digunakan untuk mengambil data dari media sosial Twitter. Data ini disimpan dan disimpan dalam format Excel.csv dengan membuat Excel.csv, untuk membantu mengidentifikasi duplikat atau konten duplikat yang harus dihapus. Pemilihan atribut, penggantian nilai yang hilang, subprocessing, dan analisis sentimen diikuti, dan hasil akhirnya adalah data yang digunakan sebagai data pelatihan dalam algoritma naive Bayes.



Gambar 1. Alur Penelitian

Pengolahan Data

Sebelum diproses, perlu dilakukan pengolahan data yang diperoleh dari twitter. Proses pengolahan data atau text preprocessing berfungsi untuk merubah data teks yang tidak terstruktur menjadi data yang terstruktur. Proses yang dilakukan adalah sebagai berikut: (Afuan, 2013)

1. Case folding

Proses case folding untuk menyeragamkan bentuk huruf menjadi huruf kecil. Hal ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital.

2. Tokenizing

Pada proses tokenization ini, semua kata yang ada di dalam tiap dokumen dipisahkan dan dihilangkan tanda bacanya, serta dihilangkan jika terdapat simbol atau apapun yang bukan huruf.

3. Stopword removal

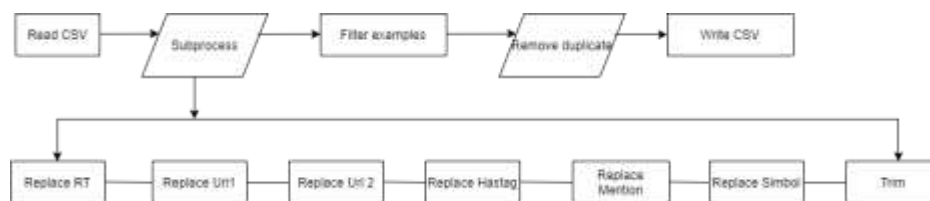
Pada tahap ini, kata-kata yang tidak relevan akan dihapus, kata-kata yang tidak mempunyai makna tersendiri jika dipisahkan dengan kata yang lain dan tidak terkait dengan kata sifat yang berhubungan dengan sentimen.

4. Stemming

Stemming adalah proses pencarian kata dasar dengan menghilangkan imbuhan. Dalam proses ini kata-kata akan dikelompokkan ke dalam beberapa kelompok yang memiliki kata dasar yang sama, seperti lantik, melantik, dan pelantikan di mana kata dasar dari semuanya adalah kata lantik. Pada penelitian ini stemming yang digunakan yaitu dari library sastrawi stemmer yang dibangun berdasarkan algoritma Nazief & Andriani.

Crawling Data

Dalam penelitian ini menggunakan tool rapid miner, Anda perlu mengunduh dan menginstal alat Rapid Miner di komputer Anda dan membuat akun Twitter. Akun ini akan terhubung atau dilampirkan ke Rapid Miner untuk meng-crawl data yang akan diambil nanti, dan dengan memanfaatkan search *API Key Twitter (API)* setelah mendapatkan akses API key maka dapat data yang diinginkan melalui fasilitas kolom pencarian di twitter yang berhubungan dengan penelitian ini, kemudian data yang sudah di ambil akan disimpan dalam database lokal dengan ekstensi excel.csv.



Gambar 2. Proses Crawling Data di Twitter

Pengambilan data menghasilkan 329 data yang di ambil berdasarkan sistem Rapid Miner. Dibawah ini adalah hasilnya adalah:

| Row No. | ID | Created At | From-User | From-User-id | To-User | To-User-id | Language |
|---------|------------|--------------|---------------|--------------|---------|------------|----------|
| 1 | 1553195888 | Jul 30, 2022 | benny blanco | 23052434 | ? | -1 | en |
| 2 | 1553476285 | Jul 31, 2022 | Nickelodeon | 204825315 | ? | -1 | en |
| 3 | 1553248259 | Jul 30, 2022 | GoFeed Info | 1105488886 | ? | -1 | en |
| 4 | 1553729520 | Jul 31, 2022 | 8th army | 1345031222 | ? | -1 | en |
| 5 | 1553728518 | Jul 31, 2022 | Tam | 9148143624 | ? | -1 | en |
| 6 | 1553728504 | Jul 31, 2022 | vanilla sky | 1477252928 | ? | -1 | en |
| 7 | 1553726486 | Jul 31, 2022 | BE Urban | 59160177 | ? | -1 | en |
| 8 | 1553726484 | Jul 31, 2022 | 00000 | 1428384251 | ? | -1 | en |
| 9 | 1553726484 | Jul 31, 2022 | J N J E E E | 2588860335 | ? | -1 | en |
| 10 | 1553726484 | Jul 31, 2022 | 00 | 1501802327 | ? | -1 | en |
| 11 | 1553726483 | Jul 31, 2022 | sa (reg user) | 1031791017 | ? | -1 | en |
| 12 | 1553726483 | Jul 31, 2022 | L_jungkaan19 | 1001248978 | ? | -1 | en |
| 13 | 1553726479 | Jul 31, 2022 | ... | 1354117714 | ? | -1 | en |

Gambar 3. Hasil Data Crawling

Sebagai hasil dari pemeriksaan dan crawling manual, kami menemukan banyak data yang bukan teks tweet. Data dihapus karena merupakan data iklan, jumlah duplikat lebih besar dari 1, yang dapat mempengaruhi skor sentimen yang digunakan sebagai data pelatihan.



| | A | B |
|----|--------------------------------------------------------------------------------|-------------|
| 1 | Text | id |
| 2 | Bock auf #KPOP? Morgen zeigen wir um 22 Uhr erneut das beliebte #MTVUn | 1,54287E+18 |
| 3 | BTS' J-Hope Reveals All the Songs of 'Jack in the Box' Album#Kpop #BTS #JHo | 1,54526E+18 |
| 4 | RT @RohitRa08117884: Movie has family flick go guys watch this movie in w | 1,54526E+18 |
| 5 | Movie has family flick go guys watch this movie in weekends all overseas fan | 1,54526E+18 |
| 6 | @bts_bighit @BTS_twt #BTS #BTS9thAnniversary #BTSARMY #ARMY #KPOP# | 1,54523E+18 |
| 7 | @bts_bighit There is a group of people in America, South Korea and other cc | 1,54522E+18 |
| 8 | @btsukrprojects @BIGHIT_MUSIC @HYBEOFFICIALtwt @HYBE_EDU @BIGHI | 1,54509E+18 |
| 9 | @bts_bighit @BTS_twt There is a group of people in America, South Korea a | 1,54507E+18 |
| 10 | Breaking: @bts_bighit break for National Military Service once again in headl | 1,54504E+18 |
| 11 | Jungkookie?@BTS_twt @BIGHIT_MUSIC @bts_bighit #JungKook #JK #Kookie | 1,545E+18 |
| 12 | K-pop supergroup #BTS has ranked high on the US Billboard's main albums c | 1,54494E+18 |
| 13 | K-pop supergroup #BTS has ranked high on the US Billboard's main albums c | 1,54493E+18 |
| 14 | @bts_bighit ?? ? ? ? ? ? ? ? ? ? ? #kpop ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? ? | 1,54469E+18 |
| 15 | K-pop Dance: Fandoming Yourself on Social Media (Routledge, 2022) https:// | 1,54465E+18 |
| 16 | ??703 NO / BTS??@BTS_twt @bts_bighit #bts #????? #army #kpop #cover # | 1,54465E+18 |
| 17 | Mi mundo@BTS_twt @bts_bighit #KPOP #SUGA #BTS_YetToCome #?????????! | 1,54455E+18 |
| 18 | Now, tell me who amongst the #KPOP artist both PAST & PRESENT have read | 1,54452E+18 |
| 19 | Wow! @bts_bighit Members Are Muscular Than They Seem, Pictures Here Fi | 1,54449E+18 |
| 20 | #unboxing @BTS_twt // #BTS #KPOP #ARMY #TIKTOK #????? @bts_bighithtt | 1,54433E+18 |
| 21 | @BTS_twt @t_gasp_official @bts_bighit @HopingDaily #SomeoneInSeoul A | 1,54432E+18 |
| 22 | Viathemind knew big things would happen for our boys...so glad she was rig | 1,5442E+18 |

Gambar 4. Dataset.csv

Kemudian, setelah menyaring Sebagai hasil dari duplikat data, para peneliti juga menghapus data yang mengandung frase iklan. Setelah semua data disaring, sentimen yang dihasilkan oleh alat rapid miner tidak akurat dan masih ada hasil sentimen yang salah, sehingga pelabelan dilakukan secara manual. Misalnya, hasil sentimen untuk atribut Polaritas adalah Tidak Ada, P (Positif), P +, Baru (netral), N (negatif).

Data Preprocessing

Preprocessing adalah langkah terpenting dalam melakukan analisis sentimen. Preprocessing atau pemrosesan text merupakan mempersiapkan sebuah teks menjadi daya yang siap diolah lebih lanjut dan bertujuan untuk menghilangkan noise, serta mengambil fitur penting yang ada pada dokumen teks. Preprocessing dalam penelitian ini terdiri dari enam tahap yaitu case folding, punctuation removal, tokenizing, normalization, stopwords removal dan stemming. Selanjutnya dilakukannya preprocessing dokumen seperti di atas untuk menghilangkan noise, menyeragamkan bentuk kata dan mengurangi volume kata.

Tahapan yang dilakukan dari dokumen preprocessing yaitu Pembersihan Data Pembersihan data adalah proses mendeteksi dan mengoreksi (atau menghapus) catatan yang rusak atau tidak akurat dari kumpulan catatan, tabel, atau basis data dan mengacu pada pengidentifikasian bagian data yang tidak lengkap, tidak benar, tidak akurat, atau tidak relevan dan kemudian mengganti, memodifikasi, atau menghapus data kotor (Winahyu & Suharjo, 2021). Pembersihan data dilakukan untuk menghilangkan beberapa data tidak memiliki value. Data yang harus dihilangkan ialah seperti username

yang di-mention, menghapus hashtag, RT, dan menghapus @. kemudian yang kedua melakukan case folding untuk menyamaratakan penggunaan huruf kapital, Misalnya data teks yang kita dapat berupa tulisan "DaTA SCIENCE" maka dengan case folding artinya kita mengubah semua huruf menjadi huruf kecil (lowercase) semua. Sementara itu, karakter lain yang bukan termasuk huruf dan angka, seperti tanda baca dan spasi dianggap sebagai delimiter. Ketiga melakukan *Tokenizing* proses pemisahan teks menjadi potongan-potongan yang disebut sebagai token untuk kemudian di analisa. Kata, angka, simbol, tanda baca dan entitas penting lainnya dapat dianggap sebagai token. Keempat melakukan *Stopword* adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh *stopword* dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll. Makna di balik penggunaan *stopword* yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya. Yang terakhir adalah melakukan *Stemming* yaitu proses menghilangkan infleksi kata ke bentuk dasarnya, namun bentuk dasar tersebut tidak berarti sama dengan akar kata (*root word*). Misalnya kata “mendengarkan”, “dengarkan”, “didengarkan” akan ditransformasi menjadi kata “dengar”.

Tabel 1. Proses Data Preprocessing

| No | Nama proses | Contoh |
|----|---------------------|-------------------------------------------------------------------------------|
| 1. | <i>Case folding</i> | Dua tahun lamaya ga ketemu miss you kalian semua |
| 2. | <i>Tokenizing</i> | ['dua', 'tahun', 'lamanya', 'ga', 'ketemu', 'miss', 'you', 'kalian', 'semua'] |
| 3. | <i>Stopwords</i> | “dua tahun lama ga ketemu miss you kalian semua |
| 4. | <i>Stemming</i> | dua tahun lama ga ketemu miss you kalian semua |

Data training

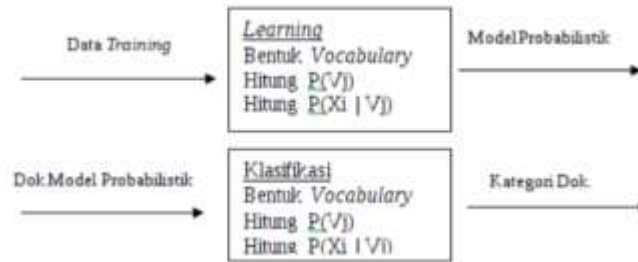
Data pelatihan yang ditandai sebelumnya memasuki sistem untuk diproses. Data yang diterima dari Twitter diberikan kategori 1, 0, -1 (positif, netral, negatif). Setelah pelabelan, data diproses sebelumnya, dibobot dan dilatih dengan Naive Bayes.

Data Testing

Data uji adalah data yang berasal dari kumpulan data tetapi tidak diberi label seperti artikel pelatihan. Item tes dimasukkan ke dalam sistem dan tujuan atau label diprediksi. Item tes juga diproses sebelum klasifikasi.

Algoritma Naive bayes

Algoritma Naive Bayes Classifier merupakan pengklasifikasian dengan metode probabilitas, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes (Handayani & Pribadi, 2015).



Gambar 5. Algoritma Naïve Bayes Classifier

Pendekatan naïve bayes membuat asumsi sederhana bahwa semua atribut bersifat independen. Hal ini menyebabkan penggolongan yang jauh lebih sederhana, ini membuat efektif dalam praktiknya.

Confusion Matrix adalah tool yang berguna untuk menganalisis seberapa baik classifier mengenali kelas yang berbeda. TP dan TN menjelaskan ketika pengklasifikasi mendapatkan sesuatu dengan benar, sementara FP dan FN menjelaskan ketika pengklasifikasi mendapatkan hal yang salah. Dibawah ini adalah rumus confusion matrix untuk menghitung nilai tingkat akurasi.

$$Accuracy = \frac{TP + FN}{TP + FP + TN + FN}$$

- TP = True positif
- TN = True negatif
- FP = False negatif
- FN = False positif

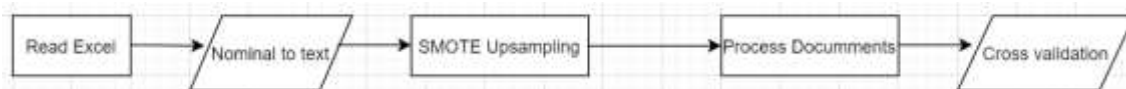
Hasil Dan Pembahasan

Hasil

Hasil penelitian rapid miner dilakukan berdasarkan eksperimen menghasilkan dataset yang sudah terlabel sebanyak 329 data, dari data tersebut dilakukan training dan testing. Dengan pembagian data 80% training dan 20% data testing

Pembahasan

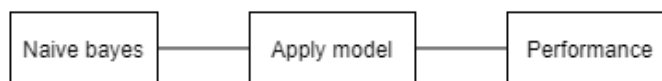
Uji akurasi ini dilakukan dengan menggunakan alat bantuan rapidminer. Pengujian dilakukan pada 329 data tweet. Di bawah ini adalah rancangan prosedur pengujian naïve Bayes.



Gambar 6. Design Pengujian Metode Naïve Bayes



Gambar 7. Tahapan Process Document From Data



Gambar 8. Validasi Pengujian Naïve Bayes

Hasil akurasi model naïve bayes menunjukkan tingkat akurasinya 67.27% artinya model klasifikasi kelulusan menggunakan naïve bayes terbukti baik hal ini dilihat dari tingkat akurasinya yang mencapai 67.27% akan tetapi hal ini perlu di tinjau ulang dari sudut pandang kompleksitas dan jumlah datasetnya. Percobaan pada penelitian ini menggunakan Rapidminer 9.10.011. Algoritma yang digunakan adalah naïve bayes. Validasinya menggunakan x-validation dan untuk testing menggunakan Apply Model untuk menjalankan algoritma atau model naïve bayes serta Performance untuk mengukur performa dari model naïve bayes tersebut.

Accuracy: 67.27% +/- 3.34% (micro average: 67.29%)

**Tabel 2
Confusion Matrix Pengujian Naïve Bayes**

| | true Positif | true Netral | true Negatif | class precision |
|---------------|--------------|-------------|--------------|-----------------|
| pred. Positif | 87 | 38 | 0 | 69.60% |
| pred. Netral | 52 | 59 | 0 | 53.15% |
| pred. Negatif | 38 | 29 | 177 | 72.54% |
| class recal | 49.15% | 46.83% | 100.00% | |

Anda dapat melihat dari gambar di atas bahwa hasil akurasi dari pengujian metode Naive Bayes. 67.27%.

Memahami Makna Kalimat

Kendala tersebut terjadi dikarenakan di dalam sebuah kalimat mengandung beberapa makna. Sebagai contoh: keluhan, sindiran, serta apresiasi.

Kendala Pelabelan Secara Manual

Beberapa catatan yang berisi kata-kata yang tergolong kata positif di awal, tengah, dan akhir kalimat, dan kata-kata yang tergolong kata negatif dalam kalimat. Hal ini menyebabkan pernyataan ambigu dan kesalahan dalam proses pelabelan.

Data Tidak Seimbang

Ketidakseimbangan data mendasari hambatan dalam memahami pendapat dan hambatan dalam proses pelabelan. Meskipun jumlah data bervariasi, perbedaan antara data berlabel positif dan data berlabel negatif mempengaruhi hasil.

Aplikasi Sentimen Analisis



Gambar 8. Proses Sentimen Analisis



Gambar 9. Grafik Sentimen Analisis

Gambar 8 merupakan halaman untuk mengecek kata yang terkandung kata negatif, positif, atau netral. Sedangkan gambar 9 itu merupakan halaman grafik, pada halaman ini pengguna dapat melihat grafik hasil data yang di cek.

Kesimpulan

Sebagai kesimpulan Analisis sentimen tweet dengan keyword aplikasi web berbagai opini yang ada di bagi 3 klasifikasi yaitu opini yang bermakna positif, opini bermakna netral dan opini bermakna negatif. Dari semua item data menghasilkan output dengan metode algoritma multinomial naive bayes karna klasifikasi lebih dari dua kelas dan proses di lakukan dengan 80% data untuk training dan 20 % untuk testing. Pada sistem melakukan uji evaluasi dengan confusion matrix. Berdasarkan data yang di uji dan pembagian data train dan tes yang berbeda akan mendapatkan performa akurasi yang berbeda. Kemudian hasil dari preproses dan uji algoritma di import ke rapid miner untuk dilakukan input sehingga pengguna dengan sentimen analisis bisa mengoperasikan dan dengan mudah mengetahui sentimen analisis dengan input keyword atau hastag yang di cari atau di inginkan. Berdasarkan penelitian yang sudah dilakukan dapat dihasilkan output data memberikan pengaruh terhadap precision maka hal ini dikarenakan Naïve Bayes Classifier dengan tipe Multinomial Naïve Bayes menghitung jumlah istilah dan kata yang ada pada sebuah kelas. Dari hasil penelitian diatas lebih banyak mengandung sentimen netral yang berisi makna informasi dan diskusi dan negatif yang berisi makna ketakutan, bullying, sedih, kecewa, unsur protes lebih sedikit. Pengujian analisis sentimen pada data crawling dengan keyword @bts_bighit pada sosial media twitter dengan metode Multinomial Naive Bayes.

Saran yang dapat penulis berikan adalah Melakukan pengembangan untuk penelitian analisis sentimen menggunakan algoritma klasifikasi yang lain atau menggunakan dua algoritma atau lebih yang berbeda supaya di hasilkan output klasifikasi analisis sentimen yang lebih baik dan bermanfaat bagi pengguna. Menambah konversi gambar ke dalam teks pada sistem ini, karena selain teks banyak sekali user yang membuat capture untuk memberikan opini atau pendapat mereka pada hasil crawling dataset. Selanjutnya perbaikan menggunakan data latih yang lebih banyak agar mendapat hasil performa yang lebih akurat.

BIBLIOGRAFI

- Afuan, L. (2013). Stemming Dokumen Teks Bahasa Indonesia. *Jurnal Telematika*, 6(2), 34–40. [Google Scholar](#)
- Handayani, F., & Pribadi, S. (2015). Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110. *Jurnal Teknik Elektro*, 7(1), 19–24. [Google Scholar](#)
- Hofmann, M., & Klinkenberg, R. (2014). RapidMiner: Data Mining Use Cases and Business Analytics Applications (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series) 1st Edition. In *Chapter 17-Medical Data Mining*. [Google Scholar](#)
- Khaira, U., Johanda, R., Utomo, P. E. P., & Suratno, T. (2020). Sentiment Analysis Of Cyberbullying On Twitter Using SentiStrength. *Indonesian Journal of Artificial Intelligence and Data Mining*, 3(1), 21. <https://doi.org/10.24014/ijaidm.v3i1.9145> [Google Scholar](#)
- Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of Natural Language Processing, Second Edition*, 627–666. [Google Scholar](#)
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. *Introduction to Information Retrieval, July 2008*, 10–12. <https://doi.org/10.1017/cbo9780511809071> [Google Scholar](#)
- Nurzahputra, A., & Muslim, M. A. (2016). Analisis Sentimen pada Opini Mahasiswa Menggunakan Natural Language Processing. *Seminar Nasional Ilmu Komputer, Snik*, 114–118. [Google Scholar](#)
- Ortega, R., Calmaestra, J., & Merchán, J. M. (2008). *Cyberbullying / Cyberbullying*. 43, 22–31. <http://www.scopus.com/inward/record.url?eid=2-s2.0-54349105131&partnerID=MN8TOARS> [Google Scholar](#)
- Winahyu, J., & Suharjo, I. (2021). Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes. *Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)*, 10(2), 206. <https://doi.org/10.23887/karmapati.v10i2.36609> [Google Scholar](#)

Copyright holder:

Anisa putri, Ari Muzakir (2022)

First publication right:

Syntax Literate: Jurnal Ilmiah Indonesia

This article is licensed under:

