

ANALISIS SENTIMEN MASYARAKAT TERHADAP PELAKSANAAN P3K GURU DENGAN ALGORITMA NAIVE BAYES DAN DECISION TREE

(THE ANALYSIS OF COMMUNITY SENTIMENT ON THE IMPLEMENTATION OF GOVERNMENT EMPLOYEES WITH WORK AGREEMENT (P3K) TEACHERS WITH NAIVE BAYES ALGORITHM AND DECISION TREE)

Fitriani¹, Ema Utami², Anggit Dwi Hartanto³

PJJ Megister Teknik Informatika, Universitas Amikom Yogyakarta
Yogyakarta, Indonesia

e-mail: ¹fitriani31121984@students.amikom.ac.id, ²ema.u@amikom.ac.id, ³anggit@amikom.ac.id

ABSTRAK

Penyelenggaraan pendidikan di Indonesia masih diwarnai dengan persoalan manajemen guru, khususnya di kalangan pengajar honorer, serta reformasi birokrasi. Aparatur Sipil Negara berupaya meningkatkan kualitas pelayanan. Kementerian Pendidikan dan Kebudayaan bekerjasama dengan Kementerian Pendayagunaan Aparatur Negara, Reformasi Birokrasi, dan Keuangan menyepakati perubahan sistem rekrutmen pegawai pemerintah dari penerimaan Calon Pegawai Negeri Sipil (CPNS) menjadi Pegawai Pemerintah dengan Perjanjian Kerja (PPPK), yang masih terjadi pro dan kontra, oleh karena itu, peneliti akan menggunakan data mining untuk melakukan analisis sentimen terhadap pelaksanaan Pegawai Pemerintah dengan Perjanjian Kontrak (PPPK), di salah satu media sosial yaitu Twitter, menggunakan jumlah data sebanyak 871 data, yg selanjutnya difiltering dan cleaning sehingga tersisa data yg sudah bersih sebanyak 519 data, sesudah menghapus data yang ganda atau terduplikat, data yg kosong serta membersihkan data. Penulis memakai metode Naive Bayes dan Decision Tree untuk melihat tingkat akurasi dari ketiga algoritma tersebut. Tool yang digunakan yaitu tools RapidMiner versi 9.10.1, hasil akurasi yang berasal dari kedua algoritma tersebut yaitu, metode naive Bayes mencapai tingkat akurasi sebanyak 100.00%. Dimana presisi kelas-pred diperlukan. Positif untuk pred adalah 100,00% dan negatif untuk pred adalah 100,00%. Persentase akurasi pendekatan Decision Tree adalah 53,95 persen. Pred-negatif memiliki kelas presisi 0,00 persen, sedangkan pred-positif memiliki presisi 53,95 persen. Seperti yang dapat diamati, pendekatan Naive Bayes memiliki tingkat akurasi tertinggi jika dibandingkan dengan decision tree, dengan tingkat akurasi 100,00 persen.

Kata kunci: Analisis Sentimen; PPPK; Twitter; Data Mining

ABSTRACT

The implementation of education in Indonesia is still inseparable from the problems of teacher management, honorary teachers, and bureaucratic reforms that affect the quality of education and the work climate in it. In an effort to improve the quality of public services by the State Civil Apparatus (ASN), the Ministry of Education and Culture agreed with the Ministry of Empowerment of State Apparatus and Bureaucratic Reform and the Ministry of Finance to change the government employee teacher recruitment system from accepting Civil Servant Candidates (CPNS) to Government Employees by Work Agreement (PPPK) which in its implementation still leaves several problems and pros and cons. Therefore, the researchers conducted a sentiment analysis in the field of data mining on the Implementation of Teacher Teacher Training on social media Twitter as many as 871 data which were then filtered and cleaned into 519 data due to duplicate data, empty data and data cleaning. The author uses the Naive Bayes and Decision Tree methods to compare the accuracy of the two methods. The researcher uses RapidMiner version 9.10.1 tools. The results showed that the sentiment analysis of Twitter data on teacher PPPK using the Naive Bayes method achieved an accuracy rate of 100.00%. Where is the class precision for pred. negative is 100.00% and pred positive is 100.00%, in the Decision Tree method the accuracy rate reaches 53.95%. Where is the class precision for pred. negative was 0.00% and pred positive was 53.95%. In this study, it can be seen that the Naive Bayes method is a method that has a higher accuracy rate than other methods with an accuracy rate of 100.00%.

Keywords: Sentiment Analysis; PPPK; Indonesia; Data Mining

I. PENDAHULUAN

Penyelenggaraan pendidikan di Indonesia masih diwarnai dengan persoalan manajemen guru, khususnya di kalangan pengajar honorer, serta reformasi birokrasi yang berdampak pada kualitas pendidikan dan lingkungan kerja kita [1]. Aparatur Sipil Negara berupaya meningkatkan kualitas pelayanan. Kementerian Pendidikan dan Kebudayaan bekerjasama dengan Kementerian Pendayagunaan Aparatur Negara, Reformasi Birokrasi, dan Keuangan menyepakati perubahan sistem rekrutmen pegawai pemerintah dari penerimaan Calon Pegawai Negeri Sipil (CPNS) menjadi Pegawai Pemerintah dengan Perjanjian Kerja (PPPK)[1]. Di Indonesia, pemerintah memiliki rencana besar untuk guru honorer, dengan tujuan mengangkat satu juta instruktur honorer pada acara rekrutmen pegawai pemerintah dengan perjanjian kerja, atau PPPK, pada tahun 2021 [2]. PPPK adalah warga negara Indonesia yang memenuhi syarat tertentu dan diangkat berdasarkan perjanjian kerja waktu tertentu untuk menyelenggarakan penyelenggaraan pemerintahan, menurut Pasal 1 ayat (4) Undang-Undang Nomor 5 Tahun 2014. PPPK dikontrak minimal satu tahun dan dapat diperpanjang maksimal 30 tahun, sesuai Undang-Undang (UU) Nomor 5 Tahun 2014 yang mengatur tentang Aparatur Sipil Negara (ASN). Ini semua tergantung situasi dan kondisi yang ada[2]. Yang dalam pelaksanaannya masih menyisakan beberapa masalah dan pro kontra, ada yang setuju dengan adanya PPPK ini ada juga yang tidak setuju.

Oleh karena itu, penting untuk mengetahui apa yang dipikirkan masyarakat Indonesia tentang PPPK. Rata-rata orang menggunakan media sosial untuk menyuarakan pemikiran mereka. Twitter merupakan salah satu aplikasi yang paling populer di kalangan masyarakat Indonesia. Twitter adalah platform bagi orang-orang untuk mengekspresikan pikiran dan pendapat mereka. Basis pengguna Twitter di Indonesia berkembang pesat, Indonesia saat ini menduduki peringkat kelima dunia[3]. Twitter banyak digunakan orang guna menyatakan keluhan kesahnya mulai dari keluhan kesah yang menimpa kehidupan tiap hari maupun keluhan kesah terhadap kebijakan-kebijakan pemerintah [4]. Oleh sebab itu sangat efektif bila memakai twitter selaku media buat mengambil informasi mengenai opini publik di Indonesia terkait pelaksanaan PPPK. Menurut temuan tinjauan pustaka peneliti, penelitian tentang analisis sentimen yang dilakukan oleh Nova Tri Romadloni, Imam Santoso, Sularso Budilaksono dari STMIK Nusa Mandiri Jakarta, Indonesia tahun 2019 dengan hasil penelitian antara lain perbedaan hasil akurasi pada pendekatan Naive Bayes akurasi 80 persen, presisi 66,67 persen, sensitivitas 100 persen, dan spesifisitas 66,67 persen. Pada teknik KNN akurasi 80%, presisi 100%,

sensitivitas 50%, dan spesifisitas 100%, tetapi pada pendekatan Decision Tree akurasi 100%, presisi 100%, sensitivitas 50%, dan spesifisitas. adalah 100%[4]. Penelitian kedua dilakukan oleh Tobias Daudert tahun 2021 dari *National University of Ireland Galway* dengan judul *Exploiting textual and relationship information for fine-grained financial sentiment analysis*, hasil penelitiannya yaitu Solusi yang di usulkan dapat meningkatkan kinerja sebanyak 15% dan 234% relatif terhadap beberapa baseline. Pekerjaan kami menunjukkan dampak sentimen implisit serta pentingnya hubungan yang berbeda untuk prediksi sentimen pada laporan perusahaan dan analisis, artikel berita, dan mikroblog. Misalnya, kami mengidentifikasi informasi stempel waktu sebagai tidak penting untuk analisis sentimen terperinci dari laporan perusahaan dan analisis. Meskipun kami dapat menunjukkan peningkatan dalam analisis sentimen keuangan, penalaran sentimen dan konteks terbatas adalah dua masalah umum yang terus terjadi. Oleh karena itu, dengan mendefinisikan ulang analisis sentimen sebagai masalah multi-teks, solusi yang kami usulkan dapat diterapkan di beberapa domain dan jenis teks, seperti ulasan produk[5]. Dan penelitian ketiga dilakukan oleh Murat Demircan, Adem Seller, Fatih Abut, Mehmet Fatih Akay tahun 2021 dari *Department of Geomatics Engineering, Istanbul Technical University, Istanbul, Turkey, Universal Yazılım A.Ş., İstanbul, Turkey dan Department of Computer Engineering, Çukurova University, Adana, Turkey*. Dengan judul penelitian *Developing Turkish Sentiment Analysis Models Using Machine Learning and E-Commerce Data*, hasil penelitiannya yaitu Hasil validasi silang pada data uji independen yang diambil dari situs e-commerce yang sama menunjukkan bahwa model analisis sentimen berbasis SVM dan berbasis RF mengungguli model lainnya. Secara lebih rinci, tidak ada urutan yang ketat antara model prediksi berbasis SVM dan berbasis RF, tetapi hasil model berbasis SVM dan berbasis RF, secara umum, adalah yang tertinggi atau, dalam kasus terburuk, serupa jika kita bandingkan. mereka dengan skor yang diperoleh dengan menggunakan model berbasis DT, berbasis LR, dan berbasis KNN. Dapat disimpulkan bahwa SVM dan RF adalah metode yang layak yang dapat digunakan untuk mengklasifikasikan ulasan produk menjadi tiga kelompok sebagai positif, negatif, dan netral dalam tingkat kesalahan yang dapat diterima[6].

Penting untuk diingat bahwa analisis sentimen adalah bagian dari penambangan data. Data mining adalah teknik untuk mengubah data menjadi informasi yang berguna [4]. Dengan kata lain, data mining dapat didefinisikan sebagai tindakan mengumpulkan pengetahuan mengenai prosedur tertentu. Teknik dan pendekatan penambangan data beragam. Akibatnya, teknik atau prosedur solusi terbaik akan ditentukan oleh hasil yang diinginkan. Dalam skenario ini, peneliti membandingkan akurasi dari dua

pendekatan analisis sentimen yang berbeda, antara lain metode Nave Bayes dan Decision Tree. Untuk sosialisasi pola statistik sering digunakan algoritma Decision Tree [4]. Pohon keputusan terdiri dari tiga simpul: daun, simpul akar yang berfungsi sebagai titik awal pohon keputusan, dan simpul mediator yang dihubungkan menggunakan tes [4]. Pendekatan Nave, Bayes juga diterapkan oleh peneliti. Bayes, atau Nave, adalah algoritma pembelajaran mesin probabilitas. Nave Bayes adalah istilah lain untuk metode membangun pembagian terstruktur teks dengan kecepatan pemrosesan tinggi. Jika Anda berurusan dengan banyak data [6]. Ada juga kepercayaan bahwa Nave, Bayes, mengacu pada pendekatan prediksi karena mengandung probabilistik sederhana yang digunakan dalam teorema Bayes yang menggunakan ketergantungan, yang merupakan alat yang ampuh[7]. Berdasarkan konteks di atas, penelitian ini dilakukan untuk mengetahui tingkat akurasi dua algoritma dengan mengintegrasikan Twitter ke dalam platform bagi peneliti untuk menjelajah data untuk mendapatkan analisis sentimen pada pelaksanaan PPPK Guru menggunakan data mining. Masalah dalam penelitian ini adalah bagaimana peneliti memperoleh data dari Twitter yang kemudian dilabeli manual dan tingkat akurasi yang diberikan oleh algoritma Decision Tree, dan Nave Bayes. Selain itu, tujuannya adalah untuk mengetahui cara mengumpulkan data Twitter menggunakan RapidMiner dengan pelabelan manual dan untuk membandingkan keakuratan kedua algoritma tersebut.

II. STUDI PUSTAKA

Berdasarkan studi literatur yang peneliti amati, penelitian mengenai sentimen analisis yang dilakukan oleh Nova Tri Romadloni, Imam Santoso, Sularso Budilaksono dari Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta, Indonesia tahun 2019, judul penelitian yaitu Perbandingan Metode Naive Bayes , KNN Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl, dengan hasil penelitian yaitu Perbedaan hasil akurasi pada metode Naive Bayes akurasi sebesar 80%, precision 66,67%, sensitivity 100%, specificity 66,67%. Pada metode KNN akurasi sebesar 80%, precision 100%, sensitivity 50%, specificity 100% dan pada metode Decision Tree akurasi sebesar 100%, precision 100%, sensitivity 100%, specificity 100%[4]. Penelitian yang dilakukan oleh Tobias Daudert tahun 2021 dari *National University of Ireland Galway* dengan judul *Exploiting textual and relationship information for fine-grained financial sentiment analysis*, hasil penelitiannya yaitu Solusi yang di usulkan dapat meningkatkan kinerja sebanyak 15% dan 234% relatif terhadap beberapa baseline. Pekerjaan kami menunjukkan dampak sentimen implisit serta pentingnya hubungan yang berbeda untuk prediksi sentimen pada laporan perusahaan dan analisis, artikel berita, dan

mikroblog. Misalnya, kami mengidentifikasi informasi stempel waktu sebagai tidak penting untuk analisis sentimen terperinci dari laporan perusahaan dan analisis. Meskipun kami dapat menunjukkan peningkatan dalam analisis sentimen keuangan, penalaran sentimen dan konteks terbatas adalah dua masalah umum yang terus terjadi. Oleh karena itu, dengan mendefinisikan ulang analisis sentimen sebagai masalah multi-teks, solusi yang kami usulkan dapat diterapkan di beberapa domain dan jenis teks, seperti ulasan produk[5]. Dan penelitian yang dilakukan oleh Murat Demircan , Adem Seller , Fatih Abut , Mehmet Fatih Akay tahun 2021 dari *Department of Geomatics Engineering, Istanbul Technical University, Istanbul, Turkey, Universal Yazılım A.Ş., İstanbul, Turkey dan Department of Computer Engineering, Çukurova University, Adana, Turkey*. Dengan judul penelitian *Developing Turkish Sentiment Analysis Models Using Machine Learning and E-Commerce Data*, hasil penelitiannya yaitu Hasil validasi silang pada data uji independen yang diambil dari situs e-commerce yang sama, menunjukkan bahwa model analisis sentimen berbasis SVM dan berbasis RF mengungguli model lainnya. Secara lebih rinci, tidak ada urutan yang ketat antara model prediksi berbasis SVM dan berbasis RF, tetapi hasil model berbasis SVM dan berbasis RF, secara umum, adalah yang tertinggi atau, dalam kasus terburuk, serupa jika kita bandingkan. mereka dengan skor yang diperoleh dengan menggunakan model berbasis DT, berbasis LR, dan berbasis KNN. Dapat disimpulkan bahwa SVM dan RF adalah metode yang layak yang dapat digunakan untuk mengklasifikasikan ulasan produk menjadi tiga kelompok sebagai positif, negatif, dan netral dalam tingkat kesalahan yang dapat diterima[6].

Penelitian yang dilakukan oleh Eka Wahyu Sholeha, Selviana Yunita, Rifqi Hammad, Veny Cahya Hardita dan Kaharuddin tahun 2022, dengan judul penelitian Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naive Bayes dan K-Nearest Neighbor, hasil dari penelitian ini yaitu: Penelitian ini bertujuan ganda, yaitu untuk membandingkan algoritma, serta menemukan efek dari huruf kapital serta tanda baca. Perbandingan akurasi antara Naive Bayes dan K-Nearest Neighbor diberikan terhadap kumpulan data. Penelitian ini mengumpulkan data dari komentar para pengguna Facebook terhadap tiga agen perjalanan online terbesar di Indonesia. Komentar di kelompokkan menjadi tiga kategori, yaitu positif, negatif, dan netral Hasil dari penelitian ini menemukan jika K-Nearest Neighbor memiliki akurasi yang sedikit lebih tinggi dibandingkan Naive Bayes. Selain itu, penggunaan huruf kecil tanpa tanda baca memiliki akurasi yang lebih baik untuk kedua algoritma[11].

Penelitian yang dilakukan oleh M. Khairul Anam, Bunga Nanti Pikir, Muhammad Bambang Firdaus, Susi Erlinda dan Agustin tahun 2021 dengan judul penelitian Penerapan *Naive Bayes Classifier, K-Nearest Neighbor (KNN)* dan *Decision Tree* untuk Menganalisis

Sentimen pada Interaksi Netizen dan Pemerintah, hasil penelitiannya yaitu Twitter menjadi tempat untuk mendapatkan data yang diungkapkan masyarakat melalui tweet yang diposting ke timeline. Analisa sentimen dilakukan untuk melihat pendapat atau kecenderungan opini netizen terhadap pemerintah Pekanbaru yang mengandung sentimen positif, negatif, dan netral. Data yang digunakan adalah tweet dengan jumlah dataset sebanyak 150 tweets. Data tersebut kemudian di analisa agar menjadi informasi. Analisa dilakukan menggunakan metode data mining yaitu *Naïve Bayes Classifier*, *K-Nearest Neighbor (KNN)*, dan *Decision tree*. Penggunaan ketiga pendekatan ini berupaya untuk mengkategorikan hasil komentar netizen terkait penggunaan teknologi yang telah melalui proses analisis sentimen dan membandingkan keakuratan ketiga cara tersebut. Hasil akurasi yang didapatkan cukup beragam yaitu dari metode *Naïve Bayes* akurasi 100%, metode KNN akurasi 98,25%, dan metode decision tree akurasi 62,28% [12]

Penelitian yang dilakukan oleh Fajar Sodik Pamungkasa dan Iqbal Kharisudin tahun 2021 dengan judul penelitian Analisis Sentimen dengan SVM, Naive Bayes dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter. Hasil penelitiannya yaitu: Pada penelitian ini dilakukan analisis sentimen tanggapan masyarakat Indonesia terhadap pandemi Covid-19 pada media sosial Twitter menggunakan algoritma *Support Vector Machine (SVM)*, *Naive Bayes*, dan *K-Nearest Neighbor*, yang kemudian ketiga algoritma tersebut dibandingkan mana yang paling baik untuk mengklasifikasikan data tanggapan. Berdasarkan tingkat rata-rata akurasi dengan menggunakan evaluasi model *10-Fold Cross Validation*, diperoleh kesimpulan bahwa algoritma SVM memiliki akurasi yang lebih tinggi daripada *Naive Bayes* dan KNN dengan rata-rata akurasinya sebesar 90,01% pada SVM dengan kernel *linear*, 79,20% pada *Naive Bayes* dengan jumlah *laplace* adalah 1, dan 62,10% pada KNN dengan jumlah K adalah 20 dan menggunakan kernel *optimal*[13].

III. METODE PENELITIAN

A. Metode Pengumpulan Data

Metode pengumpulan data yang dilakukan adalah sebagai berikut:

1) Studi Literatur

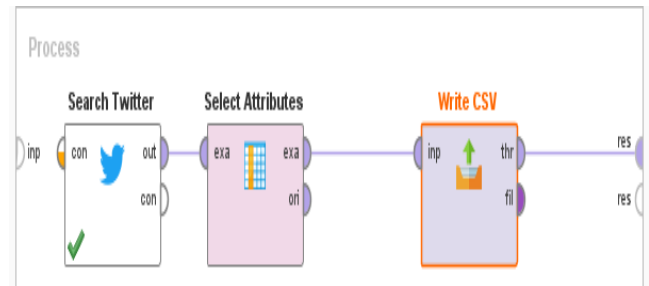
Studi literatur yg dilakukan yaitu menggunakan materi dan referensi yg terkait dengan penelitian yang bersumber dari sumber yg terpercaya, baik itu berasal jurnal nasional maupun internasional, selanjutnya mencari informasi yang berkaitan dengan penelitian yang nantinya akan digunakan pada penyusunan penelitian ini. semua pustaka yg digunakan atau dijadikan acuan mampu dicermati pada *page* daftar pustaka.

2) Wawancara

Wawancara dilakukan dengan serjana bahasa yang terkait dengan text positif dan negatif yang digunakan dalam pelabelan data.

3) Pengumpulan Data

Data dikumpulkan dari media sosial yaitu twitter menggunakan *rapidminer 9.10.1* yang menjadi alat buat mengambil data dari twitter terkait opini masyarakat terhadap aplikasi PPPK guru tahun 2021. Berikut ilustrasi cara crawling data menggunakan *rapidminer 9.10.1*.



Gambar 1. Proses *Crawling* data

Hal pertama yang dilakukan yaitu crawling data atau pengumpulan data. *Pengumpulan* pada sosial media twitter dengan jumlah data yang kita dapatkan sebesar 871 data, kemudian melakukan proses *filtering* data dengan operator *select attributs* yang digunakan untuk memilih text pada data karna attributs lain tidak dibutuhkan, yang selanjutnya proses terakhir di pengumpulan data ini yaitu data di simpan dalam bentuk CSV yg kemudian akan dilakukan preprocessing dan pelabelan data untuk data yg bersifat positif serta negatif yang dilakukan secara manual dengan melibatkan pakar bahasa. Berikut Gambar hasil dari filtering menggunakan *select attributs*, sehingga hanya tersisa baris number, ID, dan text.

Open in Turbo Prep Auto Model Filter (871 / 871 examples):

Row No.	"Text"	"id"
1	Ditematis Kato ASN lain nilai tinggi sekarang tim tentu lolos se...	*1472495726...
2	Bisatapi guru juga bisa PNSkatau ga salah pppk tuh buat yng b...	*1472487833...
3	Forum Honorir PGRI Jatim Minta Formasi Pengangkatan ASN ...	?
4	?	*1472486801...
5	Kabar Menyejukan hari ini bagi peserta pppk guru 2021 tahap ...	*1472485445...
6	PPPK Guru Tahap 2 Picu Migrasi Swasta ke Sekolah Negeri Ad...	?
7	Nilai 717 tetapi Tidak Lolos Formasi PPPK Guru Tahap 2 Pento...	?
8	?	*1472477011...
9	Ketua Paguyuban GTT/PTT Kabupaten Ketumen Musibhin risa...	?

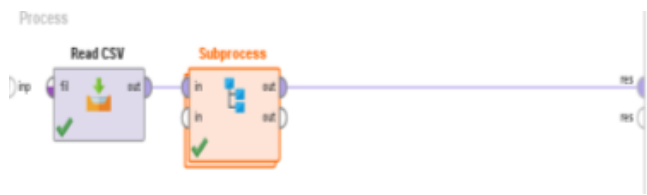
ExampleSet (871 examples, 0 special attributes, 2 regular attributes)

Gambar 2. Data hasil *crawling*

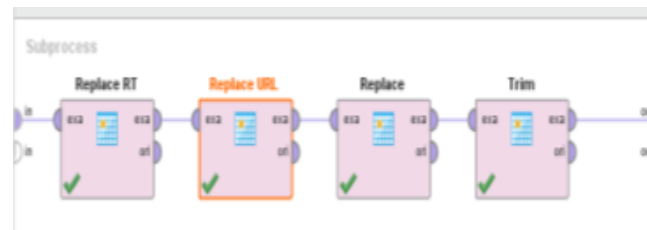
B. Preprocessing

Dalam tahapan ini kita akan membersihkan data yang kita dapatkan dengan operator *subprocess* yang didalamnya berisi *replace* yang digunakan untuk menghapus RT, URL baik yang berada diawal maupun yang di akhir, menghapus hastag, mention, simbo-

Isymbol yang tidak perlu seperti [.,;”!\$%^&*@&()-] dan spasi yang tidak perlu.

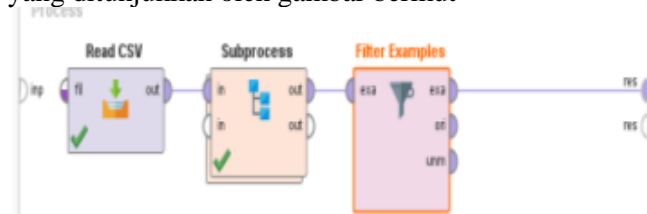


Gambar 3. Subprocess



Gambar 4. Replace

Selanjutnya kita akan membuang data-data yang kosong dengan menggunakan *filter Examples* yang ditunjukkan oleh gambar berikut



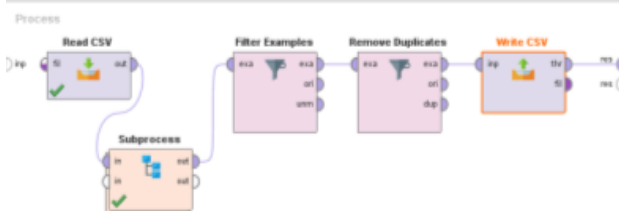
Gambar 5. Filter Examples

Setelah menghapus data yang kosong dari jumlah data sebanyak 871, hasilnya menjadi 731.

Row No.	Text
8	Cara Cek Pengumuman dan Ajuan Sanggah Hasil Seleksi PPPK Guru Tahun 2 Buka
9	via tribunjatim
10	Peserta PPPK Guru Tahun 2 yang Tenang Ada Kabar Menyejukan dari Kemendikbudristek
11	Kemendikbudristek akhirnya memberikan klarifikasi soal hasil pengumuman kelulusan PPPK guru tahun 2 semua peserta bisa tenang
12	Peserta Bersedih Berjaya di PPPK Guru Tahun 2 Pimpinan Honorer: Di Mana Hati Nurani Anda Mas Nadiem
13	Pimpinan honorer melontarkan kritikan pedas untuk Mas Nadiem karena lebih mengutamakan peserta bersedih dalam seleksi PPPK g.
14	Pengumuman Hasil Seleksi PPPK Guru 2021 Tahap 2 Diundur 21 Desember 2021 Inih Jadwal Terbaru
15	Kalau sudah lulus ppgk guru mungkin gagal dipemberikan gaji Kalau kita gi ambil formasi apa ada konsekuensinya
16	Kalau sudah lulus tes ppgk guru masih mungkin gugur gi eh dipemberikan Kalau tidak mengambil formasi apa ada konsekuensinya

Gambar 6. Data setelah di filter examples

Setelah menghapus data yang kosong selanjutnya kita hapus data yang *duplicate* atau data yang ganda yang kemudian filenya kita simpan dalam bentuk csv.



Gambar 7. Remove Duplicate

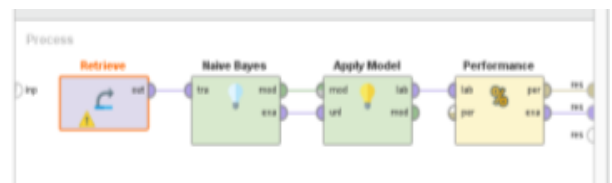
Setelah melakukan *remove duplicate*, dari data 731 setelah dihapus data yang duplicate atau ganda maka hasilnya menjadi 519 data, berikut hasilnya.

Row No.	Text
1	Dilematika Kalau ASN lain nilai tinggi sekarang tim tentu lolos seleksi guru yg gak lolos minta tambahan afirmasi Begitu sama guru sek...
2	Bisatapi guru juga bisa PNSkalau ga salah ppgk tuh buat yng belum PNS tapi udah lebih dari 35 thn umumnya
3	Forum Honorer PGRI Jatim Minta Formasi Pengangkatan ASN PPPK Sesuai Kondisi Real Daerah
4	Kabar Menyejukan hari ini bagi peserta ppgk guru 2021 tahap 2dari Keme lewat YouTube
5	PPPK Guru Tahun 2 Pica Migrasi Swasta ke Sekolah Negeri Ada Bukunya
6	Nilai 717 tetapi Lolos Formasi PPPK Guru Tahun 2 Pantolan Honorer K2 In Kerasal -
7	Kalau Paguyuban GTTPTT Kabupaten Kebumen Mustahil nraau hatinya karena mulai terjadi migrasi guru swasta ke sekolah negeri d..
8	Cara Cek Pengumuman dan Ajuan Sanggah Hasil Seleksi PPPK Guru Tahun 2 Buka
9	via tribunjatim

Gambar 8. Hasil setelah remove duplicate

C. Cross Validation

Validasi silang adalah teknik untuk menentukan keakuratan suatu prosedur dengan membagi data ke dalam fase pelatihan dan pengujian. dalam bentuk kinerja dan model penerapan Validasi silang untuk naive bayes ditunjukkan di bawah ini.



Gambar 9. Cross Validation naive bayes

Proses cross validation algoritma *Naive Bayes* ditunjukkan pada Gambar 9. Pada kolom data latih terdapat pendekatan aksi *Naive Bayes*. Pada kolom pengujian data terdapat kegiatan seperti penerapan model dan performansi. Tujuannya adalah untuk melihat seberapa akurat data untuk algoritma *Naive Bayes*. Setelah itu, kita akan melihat validasi silang untuk pohon keputusan (*decision tree*)

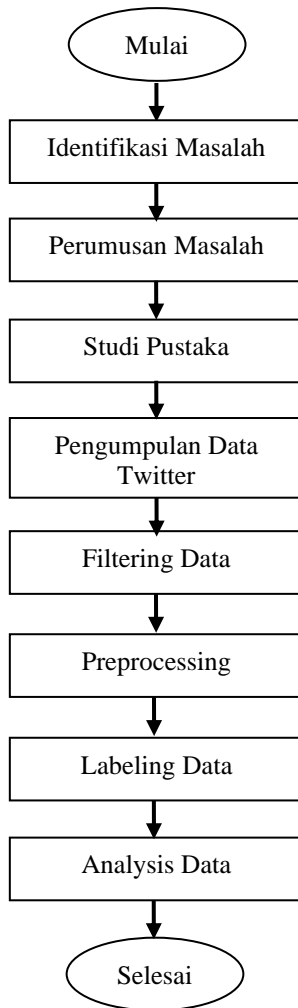


Gambar 10. Cross Validation Decision Tree

Gambar 10 menggambarkan tahap validasi silang. Terdapat teknik Action Decision Tree pada kolom data latih. Tindakan seperti menerapkan contoh dan kinerja dapat ditemukan di kolom pengujian data. Tujuan dari penelitian ini adalah untuk melihat seberapa akurat pendekatan Decision Tree menggunakan data.

D. Alur Progamam

Berikut adalah alur penelitian yang dibuat:



Gambar 11. Alur program

Berikut merupakan penjelasan dari alur program yang dirancang.

- a. Identifikasi masalah adalah mengidentifikasi dari persoalan yg ingin diuji.
- b. Rumusan masalah, yaitu peneliti merumuskan masalah pada titik ini. Masalah yang akan diuji dalam hal ini adalah bagaimana pengumpulan data melalui Twitter menggunakan *tools RapidMiner* dan seberapa akurat pendekatan *Decision Tree*, dan *Nave Bayes*.
- c. Studi literatur, yang meliputi pembahasan pemahaman teoritis atau penjelasan tentang apa itu data mining, apa itu metode *Decision Tree*, dan metode *Nave Bayes*. Akibatnya, penyelidikan awal peneliti hanya terdiri dari tinjauan literatur. Studi literatur yang diterbitkan dalam majalah nasional dan internasional.
- d. Pada titik ini, data dikumpulkan dari Twitter atau crawling data. Jumlah total data yang dikumpulkan adalah 871. Prosedur ini cepat dan tidak menyakitkan. Pada titik ini, memiliki akun Twitter diperlukan untuk mendapatkan

token akses, yang akan dimasukkan ke dalam alat RapidMiner.

- e. Peneliti melakukan metode penyaringan data pada saat ini karena ada data duplikat. Alhasil, data yang tadinya 871 kini tinggal 519.
- f. Pada titik ini, pra-pemrosesan melibatkan penghapusan RT, URL, tagar, sebutan, simbol, dan spasi yang tidak relevan.
- g. Pelabelan Data Pada titik ini, peneliti menggunakan prosedur Pelabelan Data ahli bahasa untuk menentukan sifat teks yang telah dikumpulkan. Ada label positif dan negatif pada label ini.
- h. Menganalisis Data Para peneliti memeriksa keakuratan kedua pendekatan yang digunakan pada saat ini.

IV. HASIL DAN PEMBAHASAN

Analisis sentimen pelaksanaan guru PPPK tahun 2021 dengan crawling data dari twitter dan pelabelan dilakukan secara manual yang selanjutnya mencari akurasi dari algoritma naive bayes dan decision tree, berikut pembahasannya.

A. Labeling Data

Berikut adalah contoh data yang dilabeli secara manual.

1	Text	Sentimen
2	Dillematis Kalo ASN lain nilai tinggi sekarang blm tentu lolos seleksi guru yg gak lolos minta tambahan afirmasi Setuju sama guru sekarang sistem PP	Negatif
3	Bisatapi guru juga bisa PNSkalau ga salah ppkp tuh buat yg belum PNS tapi udah lebih dari 35 thn umurnya	Positif
4	Forum Honorer PGRI Jatim Minta Formasi Pengangkatan ASN PPPK Sesuai Kondisi Real Daerah	Positif
5	Kabar Menyejukkan hari ini bagi peserta ppkp guru 2021 tahap 2 dari Kamu lewat YouTube	Positif
6	PPPK Guru Tahap 2 Picu Migrasi Swasta Ke Sekolah Negeri Ada Bukti	Negatif
7	Nilai 71,7 tetapi Tidak Lolos Formasi PPPK Guru Tahap 2 Pentolan Honorer K2 Ini Kesal -	Negatif
8	Ketua Paguyuban GTT/PTT Kabupaten Kebumen Musibihin risau hatinya karena mulai terjadi migrasi guru swasta ke sekolah negeri dalam PPPK guru	Negatif
9	Cara Cek Pengumuman dan Ajuan Sanggah Hasil Seleksi PPPK Guru Tahap 2 Buka	Positif
10	via tribunjatim	Positif
11	Peserta PPPK Guru Tahap 2 yang Tenang Ada Kabar Menyejukkan dari Kemendikbudristek	Positif
12	Kemendikbudristek akhirnya memberikan klarifikasi soal hasil pengumuman kelulusan PPPK guru tahap 2 semoga peserta bisa tenang	Positif
13	Peserta Berserdik Berjaya di PPPK Guru Tahap 2 Pimpinan Honorer: Di Mana Hati Nurani Anda Mas Nadiem	Negatif
14	Pimpinan honorer melontarkan kritikan pedas untuk Mas Nadiem karena lebih mengutamakan peserta berserdik dalam seleksi PPPK guru tahap 2	Positif
15	Pengumuman Hasil Seleksi PPPK Guru 2021 Tahap 2 Diundur 21 Desember 2021 Inilah Jadwal Terbaru	Negatif
16	Kalau sudah lulus ppkp guru mungkin akan dipemberikan gaji kalau kita gak ambil formasi apa ada konsekuensi nya	Negatif
17	Kalau sudah lulus tapi enak mau resign mungkin mau gak dipanggil mungkin kalo tidak mau ambil formasi apa ada konsekuensi nya	Negatif

Gambar 12. Labeling Data

Gambar 12 menunjukkan data yang telah diberi label dengan tangan. Sebelumnya, itu hanya satu baris teks. Jadi akan ada satu kolom tambahan bernama *sentimen* pada saat ini. Label ini digunakan untuk menentukan apa isi teks tersebut. Ada dua jenis properti data: positif dan negatif.

B. Hasil Accuracy Naive Bayes

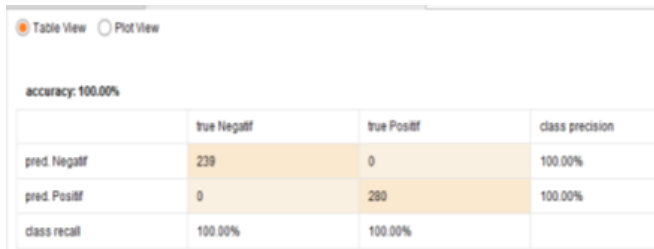


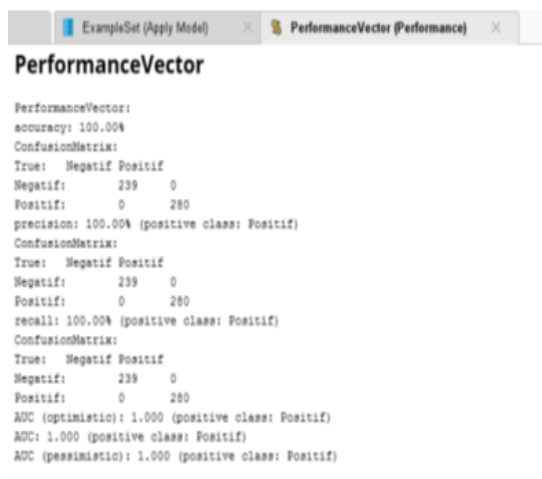
Table View Plot View

accuracy: 100.00%

	true Negatif	true Positif	class precision
pred. Negatif	239	0	100.00%
pred. Positif	0	280	100.00%
class recall	100.00%	100.00%	

Gambar 13. Hasil Accuracy Naive Bayes

Hasil metode *Naive Bayes* menunjukkan bahwa tingkat akurasi metode tersebut adalah 100,00 persen. Pred. negatif memiliki presisi 100,00 persen, sedangkan pred. positif memiliki presisi 100,00 persen. Sebuah vektor kinerja untuk algoritma *Naive Bayes* juga disertakan. Setelah itu, ada tampilan plot dari algoritma ini.



ExampleSet (Apply Model) PerformanceVector (Performance)

PerformanceVector

PerformanceVector:
accuracy: 100.00%

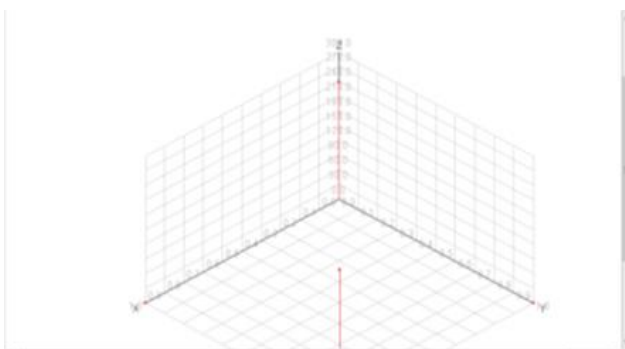
ConfusionMatrix:
True: Negatif Positif
Negatif: 239 0
Positif: 0 280
precision: 100.00% (positive class: Positif)

ConfusionMatrix:
True: Negatif Positif
Negatif: 239 0
Positif: 0 280
recall: 100.00% (positive class: Positif)

ConfusionMatrix:
True: Negatif Positif
Negatif: 239 0
Positif: 0 280
AUC (optimistic): 1.000 (positive class: Positif)
AUC: 1.000 (positive class: Positif)
AUC (pessimistic): 1.000 (positive class: Positif)

Gambar 14. Performance Vector Metode Naive Bayes

Vektor kinerja adalah representasi dari asal tabel yang analisisnya akan dilakukan. Hanya saja, vektor kinerja dari algoritma *Naive Bayes* ini sekarang menyertakan peningkatan seperti kappa.



Gambar 15. Plot View Metode Naive Bayes

C. Hasil Accuracy Algoritma Decision Tree

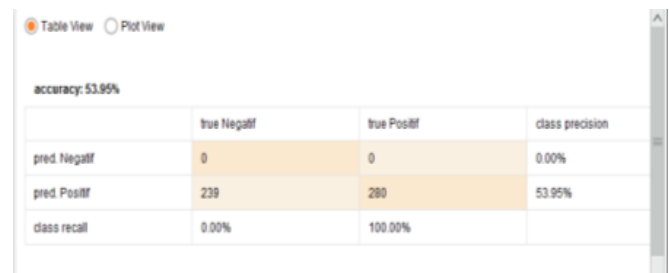


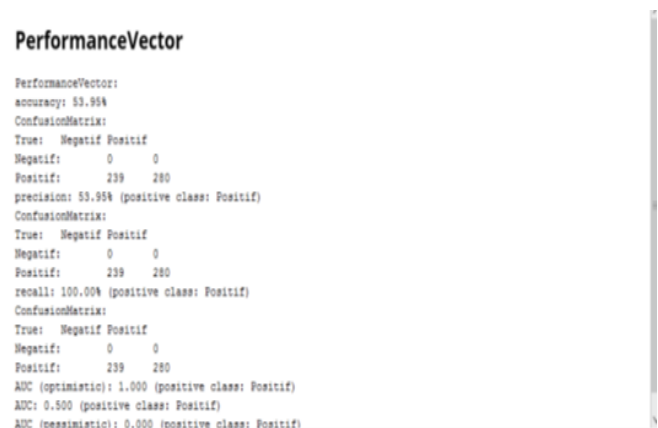
Table View Plot View

accuracy: 53.95%

	true Negatif	true Positif	class precision
pred. Negatif	0	0	0.00%
pred. Positif	239	280	53.95%
class recall	0.00%	100.00%	

Gambar 16. Hasil Accuracy dari Decision Tree

Pendekatan *Decision Tree* memiliki tingkat akurasi 53,95 persen, menurut data. Pred. negatif memiliki kelompok presisi 0,00 persen, sedangkan pred. positif memiliki presisi 53,95 persen. Sebuah vektor kinerja dari teknik *Pohon Keputusan* juga disertakan. Selain itu, ada tampilan plot dari algoritma ini.



PerformanceVector

PerformanceVector:
accuracy: 53.95%

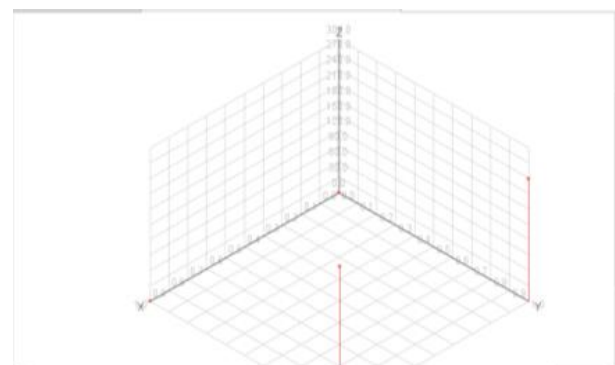
ConfusionMatrix:
True: Negatif Positif
Negatif: 0 0
Positif: 239 280
precision: 53.95% (positive class: Positif)

ConfusionMatrix:
True: Negatif Positif
Negatif: 0 0
Positif: 239 280
recall: 100.00% (positive class: Positif)

ConfusionMatrix:
True: Negatif Positif
Negatif: 0 0
Positif: 239 280
AUC (optimistic): 1.000 (positive class: Positif)
AUC: 0.500 (positive class: Positif)
AUC (pessimistic): 0.000 (positive class: Positif)

Gambar 17. PerformanceVector

Vektor kinerja adalah representasi dari asal tabel yang analisisnya akan dilakukan. Hanya saja, vektor kinerja dari *decision tree* ini sekarang menyertakan peningkatan seperti kappa.



Gambar 18. Plot View Metode Decision Tree

V. KESIMPULAN

Berdasarkan pembahasan diatas dapat disimpulkan bahwa penelitian ini menggunakan RapidMiner untuk menentukan tingkat akurasi dari dua

metode yaitu *Decision Tree*, dan *Nave Bayes*. *Rapidminer versi 9.10.1* adalah alat yang digunakan oleh peneliti. Dalam skenario ini, peneliti menggunakan Twitter untuk mengumpulkan data dan menghubungkan RapidMiner ke api Twitter. Setelah pengumpulan data, prosedur pemrosesan data seperti penyaringan, pelabelan, dan analisis digunakan untuk mengevaluasi tingkat akurasi setiap pendekatan. Algoritma *Decision Tree*, dan *Nave Bayes* semuanya dapat digunakan dalam penambahan data, seperti yang ditunjukkan dalam penelitian ini. Analisis sentimen pada data Twitter tentang guru PPPK menghasilkan tingkat akurasi nave bayes sebesar 100,00 persen. Dimana presisi grup untuk pred. negatif adalah 100,00% dan pred. positif adalah 100,00%, tingkat akurasi untuk pendekatan *decision tree* adalah 53,95 persen. Pred negatif adalah 0,00 persen dan pred positif adalah 53,95 persen. Pada penelitian ini dapat ditunjukkan bahwa pendekatan Naive Bayes memiliki tingkat akurasi yang lebih baik dibandingkan dengan *decision tree* yaitu memiliki tingkat akurasi sebesar 100,00 persen.

DAFTAR PUSTAKA

- [1] M. Goldman, Ian. and Pabari, “PERMASALAHAN GURU HONORER TERKAIT KEBIJAKAN PENGHENTIAN REKRUTMEN GURU PNS MENJADI PPPK,” vol. 2, no. 4, 2021.
- [2] N. Pekanbaru, S. Fahmi, A. Faridhi, and N. Hendayana, “Website : <https://journal.unilak.ac.id/index.php/Respublica> PELAKSANAAN UNDANG-UNDANG NOMOR TENTANG APARATUR SIPIL NEGARA TERHADAP PPPK YANG DILAKUKAN DI SMP Perubahan Undang-Undang Dasar Negara Kesatuan Republik Indonesia Tahun 1945 membawa perubahan dise,” pp. 1–15, 2014.
- [3] T. Mardiana, H. Syahreva, and T. Tuslaela, “Komparasi Metode Klasifikasi Pada Analisis Sentimen Usaha Waralaba Berdasarkan Data Twitter,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 267–274, 2019, doi: 10.33480/pilar.v15i2.752.
- [4] N. T. Romadloni, I. Santoso, and S. Budilaksono, “Perbandingan Metode Naive Bayes , Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl,” *J. IKRA-ITH Inform.*, vol. 3, no. 2, pp. 1–9, 2019.
- [5] T. Daudert, “Exploiting textual and relationship information for fine-grained financial sentiment analysis,” *Knowledge-Based Syst.*, vol. 230, p. 107389, 2021, doi: 10.1016/j.knosys.2021.107389.
- [6] M. Demircan, A. Seller, F. Abut, and M. F. Akay, “Developing Turkish sentiment analysis models using machine learning and e-commerce data,” *Int. J. Cogn. Comput. Eng.*, vol. 2, pp. 202–207, 2021, doi: 10.1016/j.ijcce.2021.11.003.
- [7] C. A. Sugianto, “Analisis Komparasi Algoritma Klasifikasi Untuk Menangani Data Tidak Seimbang Pada Data Kebakaran Hutan,” *Techno.Com*, vol. 14, no. 4, pp. 336–342, 2015, [Online]. Available: <http://publikasi.dinus.ac.id/index.php/technoc/article/view/992>.
- [8] R. Puspita and A. Widodo, “Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS,” *J. Inform. Univ. Pamulang*, vol. 5, no. 4, p. 646, 2021, doi: 10.32493/informatika.v5i4.7622.
- [9] H. Thakkar, V. Shah, H. Yagnik, and M. Shah, “Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis,” *Clin. eHealth*, vol. 4, pp. 12–23, 2021, doi: 10.1016/j.ceh.2020.11.001.
- [10] A. I. Sang, E. Sutoyo, and I. Darmawan, “Analisis Data Mining Untuk Klasifikasi Data Kualitas Udara Dki Jakarta Menggunakan Algoritma Decision Tree Dan Support Vector Machine Data Mining Analysis for Classification of Air Quality Data Dki Jakarta Using Decision Tree Algorithm and Support Vector,” vol. 8, no. 5, pp. 8954–8963, 2021.
- [11] E. W. Sholeha, S. Yunita, R. Hammad, V. C. Hardita, and K. Kaharuddin, “Analisis Sentimen Pada Agen Perjalanan Online Menggunakan Naïve Bayes dan K-Nearest Neighbor,” *JTIM J. Teknol. Inf. dan Multimed.*, vol. 3, no. 4, pp. 203–208, 2022, doi: 10.35746/jtim.v3i4.178.
- [12] M. K. Anam, B. N. Pikir, M. B. Firdaus, and S. Erlinda, “Penerapan Naïve Bayes Classifier , K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemeritah Applications of Naïve Bayes Classifier , K-Nearest Neighbor and Decision Tree to Analyze Sentiment on Ne,” vol. 21, no. 1, 2021, doi: 10.30812/matrik.v21i1.1092.
- [13] F. Sodik and I. Kharisudin, “Analisis Sentimen dengan SVM , NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter,” vol. 4, pp. 628–634, 2021.