

Analisis Sentimen Dalam Pengkategorian Komentar Youtube Terhadap Layanan Akademik dan Non-Akademik Universitas Terbuka Untuk Prediksi Kepuasan

Rhini Fatmasari¹, Virda Mega Ayu^{2*}, Harianto², Bryan Pratama², Windu Gata²

¹ Fakultas Keguruan dan Ilmu Pengetahuan, Universitas Terbuka, Jakarta, Indonesia

² Fakultas Teknologi Informasi, Ilmu Komputer, Universitas Nusa Mandiri, Jakarta, Indonesia

Email: ¹liens@ecampus.ut.ac.id, ^{2*}vir damega8@gmail.com, ³masharry.net@gmail.com, ⁴bryanp2803@gmail.com, ⁵windu@nusamandiri.ac.id

Email Penulis Korespondensi: virdamega8@gmail.com

Submitted: 21/06/2022; Accepted: 15/07/2022; Published: 30/09/2022

Abstrak—Kunci kesuksesan dari suatu organisasi pendidikan dalam mencapai tujuan tentunya tidak terlepas dari kualitas pelayanan baik itu dalam bentuk akademik dan non akademik. Dimana dalam mencapai tujuan tersebut tentunya dengan cara memberikan kepuasan kepada pihak akademis. Studi kasus dilakukan prediksi kepuasan layanan dilakukan pada Universitas Terbuka dengan menggunakan komentar pada sosial media Youtube sebagai pengolahan datanya. Pendekatan *text mining* merupakan salah satu alternatif yang baik dalam tujuan mengartikan makna dalam komentar yang diberikan. Penelitian ini bertujuan untuk menganalisa prediksi kepuasan layanan dari beberapa kategori sebagai tolak ukurnya. Adapun pengkategorianya yaitu: Module, Tutorial, Beasiswa, Pengajar, Ujian, Aplikasi, Non Akademik dan *Others*. Metode penelitian yang digunakan adalah komparatif, dengan menerapkan 4 algoritma yaitu *Decision Tree* (DT), *Support Vector Machine* (SVM), *Naive Bayes* (NB) dan *Random Forest* (RF) untuk *Prediction Accuracy*. Total dataset awal 7776 data dan setelah dilakukan *cleansing* dan *preprocessing* berjumlah 6920 data. Dan kemudian di evaluasi untuk ke 7 kategori setelah di akurasiakan menghasilkan: Kategori Modul akurasi tertinggi sebesar 99,37% dengan menggunakan algoritma DT, Kategori Aplikasi akurasi tertinggi sebesar 100% dengan menggunakan algoritma DT, Kategori Pengajar akurasi tertinggi 99,42% dengan menggunakan algoritma DT. Kategori Tutorial akurasi tertinggi 92,4% dengan menggunakan algoritma SVM, Kategori Ujian akurasi tertinggi 99,7% dengan menggunakan algoritma RF, Kategori Non Akademik akurasi tertinggi 99,90% dengan menggunakan algoritma DT. Dan untuk kategori *Others* akurasi tertinggi 96,58% dengan menggunakan algoritma DT.

Kata Kunci: Akademik; Non Akademik; Kategori; Komentar; Youtube; Decision Tree; SVM; Naive Bayes; Random Forest

Abstract—The key to the success of an educational organization in achieving its goals of course cannot be separated from the quality of service both in academic and non-academic forms. Where in achieving these goals of course by giving satisfaction to the academics. The case study was carried out to predict service satisfaction at the Open University by using comments on social media Youtube as data processing. The text mining approach is a good alternative in terms of interpreting the meaning in the comments given. This study aims to analyze the predictions of service satisfaction from several categories as a benchmark. The categories are: Module, Tutorial, Scholarship, Lecturer, Exam, Application, Non-Academic and Others. The research method used is comparative, by applying 4 algorithms, namely Decision Tree (DT), Support Vector Machine (SVM), Naive Bayes (NB) and Random Forest (RF) for Prediction Accuracy. The total initial dataset is 7776 data and after cleansing and preprocessing is 6920 data. And then evaluated for the 7 categories after being accured to produce: Module category with the highest accuracy of 99.37% using the DT algorithm, Application Category with the highest accuracy of 100% using the DT algorithm, Teaching Category the highest accuracy of 99.42% using the algorithm DT. The tutorial category has the highest accuracy 92.4% using the SVM algorithm, the exam category has the highest accuracy 99.7% using the RF algorithm, the non-academic category has the highest accuracy 99.90% using the DT algorithm. And for the Others category the highest accuracy is 96.58% using the DT algorithm.

Keywords: Academic; Non-Academic; Categories; Comments; Youtube; Decision Tree; SVM; Naive Bayes; Random Forest

1. PENDAHULUAN

Perguruan tinggi merupakan salah satu lembaga yang bergerak dibidang jasa pendidikan, dituntut untuk memberikan pelayanan akademik dan juga non akademik yang mempunyai kualitas yang baik sesuai dengan harapan dan keinginan peserta didiknya tanpa memandang status sosial dari peserta didik. Perguruan tinggi sebagai salah satu bentuk pendidikan yang tinggi yang harus memberikan pelayanan yang memuaskan dan juga mampu memenuhi kebutuhan mahasiswa. Kualitas pelayanan dapat dicapai dengan memenuhi kebutuhan serta keinginan yang diharapkan oleh pelanggan [1]. Kualitas layanan akademik menjadi suatu hal yang sangat penting dalam lembaga pendidikan, dimana pelayanan akademik dapat disebut berkualitas jika sesuai dengan kebutuhan para pelanggannya[2]. Layanan akademik adalah layanan pendidikan terkait langsung kepada pelanggan utama (Mahasiswa) dari Universitas yang termasuk silabus, kurikulum, desain kualitas pendidikan, satuan materi sajian, penyajian materi, evaluasi, magang, bimbingan[3].

Kemajuan sebuah institusi pendidikan tersebut salah satunya dipengaruhi oleh kepuasan dari pengguna jasa pendidikan, dimana mahasiswa yang berperan sebagai pengguna jasa yang paling utama pada suatu perguruan tinggi. Karena mahasiswa yang merasa puas akan memberikan suatu loyalitas tinggi terhadap perguruan tingginya. Sebab, pengguna jasa yang loyal pada umumnya akan menjadi sebuah aset penting yang bernilai tinggi. Dengan tingkat kepuasan akan dapat memberikan umpan balik positif yang menjadi ajang promosi perguruan tinggi tersebut kepada publik. Sehingga dapat memberikan kesan dan pencitraan positif bagi perguruan tinggi tersebut.

Sama seperti halnya dengan salah satu Perguruan tinggi yaitu Universitas Terbuka (UT), dimana UT ini merupakan Perguruan Tinggi Negeri ke-45 di Indonesia yang diresmikan pada tanggal 4 September 1984, berdasarkan keputusan Presiden RI Nomor 41 tahun 1984. UT memiliki 4 fakultas, yaitu Fakultas Ekonomi (FE), Fakultas Hukum, Ilmu Sosial dan Ilmu Politik (FHISIP), Fakultas Dains dan teknologi (FST), Fakultas Keguruan dan Ilmu Pendidikan (FKIP) untuk jenjang Diploma dan Sarjana dan Doktor[4]. UT mempunyai Visi untuk menjadi perguruan tinggi terbuka dan jarak jauh (PTTJJ) berkualitas dunia, untuk mencapai visi tersebut maka UT menempuh Misi yaitu dengan menyediakan akses pendidikan tinggi yang berkualitas dunia bagi semua lapisan masyarakat melalui penyelenggaraan berbagai program PTTJJ untuk menghasilkan lulusan yang berdaya saing tinggi[5]. Dalam rencana strategis bertujuan untuk melakukan perbaikan untuk aspek-aspek yang masih lemah serta peningkatan kualitas aspek-aspek yang sudah berjalan baik, sehingga ditambahkan sasaran dalam Renstra dan Renop UT salah satunya dengan mengukur indeks kepuasan mahasiswa dan pemangku kepentingan[6].

Untuk melihat kesan dan pendapat dalam bentuk komentar dapat lebih cepat dilakukan dengan menggunakan media sosial. Sebab dengan internet menjadikan sosial media sebagai wadah yang lebih berkembang pesat dengan cepat dan efektif. Media sosial kini sarana bagi masyarakat untuk saling berkomunikasi dengan orang lain dengan beragam cara. salah satunya yaitu dengan melalui Youtube. Youtube merupakan media berbagi informasi dalam bentuk video yang terbesar. Youtube sendiri merupakan media sosial yang memungkinkan penggunaannya untuk menonton, menyiarkan, dan berbagi video. Youtube berkembang pesat di Indonesia. Banyak saluran Youtube yang berisi berbagai konten video yang diunggah oleh pengguna Youtube. Youtube memiliki peringkat iklan dengan fitur pilihan *like* dan *dislike*. Pengguna Youtube yang menonton video di saluran Youtube dapat mengungkapkan pendapatnya tentang video tersebut pada kolom komentar. Komentar pengguna dapat digunakan sebagai sumber untuk menilai dan menilai kualitas saluran Youtube[7].

Youtube menjadi Salah satu sosial media yang menarik untuk diteliti. Warganet di Indonesia sebanyak 88% mengakses Youtube[8]. Dimana Youtube menjadi salah satu video sharing dengan tujuan untuk berinteraksi antara pengguna satu dengan yang lain. Selain untuk mendapatkan informasi, Youtube juga dapat memberikan komentar dan berdiskusi tentang suatu hal. Ulasan opini orang-orang tentang suatu topik dapat menjadi sumber data yang potensial untuk mengetahui pendapat dan minat masyarakat terhadap suatu topik. Opini mining atau analisis sentimen merupakan proses memahami, mengekstrak, dan mengolah data terkstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini[9].

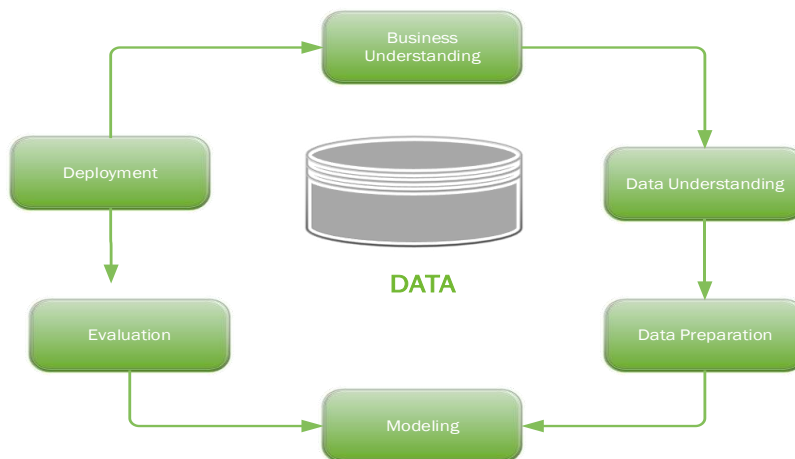
Adapun penelitian serupa mengenai *text mining* dari komentar Youtube yaitu yang dilakukan oleh Pramana Yoga Saputra, dkk. Dengan judul “Implementasi sentimen analisis komentar channel video pelayanan pemerintah di youtub menggunakan algoritma naïve bayes”. Penelitian ini melakukan analisa sentimen menggunakan klasifikasi data Youtube pada komentar video yang dipublish pemerintah tentang kinerjanya kepada masyarakat. Data tersebut diklasifikasikan dengan memperoleh nilai akurasi sebesar untuk KemenPUPR sebesar 69,23% dan 64,10% untuk Kemenkeu[10]. Dan adapula penelitian lainnya dengan judul “Analisis sentimen pengguna Youtube di Indonesia pada *review Smartphone* menggunakan naïve bayes” yang ditulis oleh Balya. Dimana pada penelitian ini melakukan *Text Mining* pada situs Youtube mengenai *review smartpohone* dengan metode *scrapping*, labelisasi, *preprocessing* (*case folding, tokenisasi, filtering*), perhitungan frekuensi kemunculan kata (tf), dan klasifikasi sentimen. Penelitian ini menghasilkan klasifikasi komentar pada video Youtube dalam sentimen positif, negatif. Netral dan membandingkan algoritma Multinomial Naïve Bayes. Hasil akurasi sebesar 83% dengan Algoritma Gaussian Naïve Bayes dan Multinomial Naïve Bayes sebesar 83%[11].

Fungsi dari adanya komentar yang ada maka dapat digali suatu informasi dari komentar-komentar yang ditinggalkan. Oleh sebab itu, untuk mengetahui bagaimana opini dari masyarakat, baik itu *public* atau almamater dari Universitas Terbuka, maka dilakukan penelitian ini. Pendekatan *text mining* menjadi salah satu pilihan yang terbaik dalam mengartikan suatu arti dalam komentar. Untuk memudahkan dari klasifikasi jenis komentar yang ada, dilakukan pengkategorian yang dimana menjadi tujuh kategori (“*Module, Tutorial, Beasiswa, Pengajar, Ujian, Aplikasi, Non Akademik dan Others*”). Dan setelah itu akan dibuatkan klasifikasi dengan menggunakan metode *Decision Tree* (DT), *Support Vector Machine* (SVM), *Naïve Bayes* (NB) dan *Random Forest* (RF) untuk *Prediction Accuracy*. Diharapkan dengan dukungan dari beberapa metode tersebut dapat membantu institusi dalam melakukan penentuan akurasi pada kategori yang ada sehingga dapat menganalisa untuk memberikan pelayanan terbaik.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

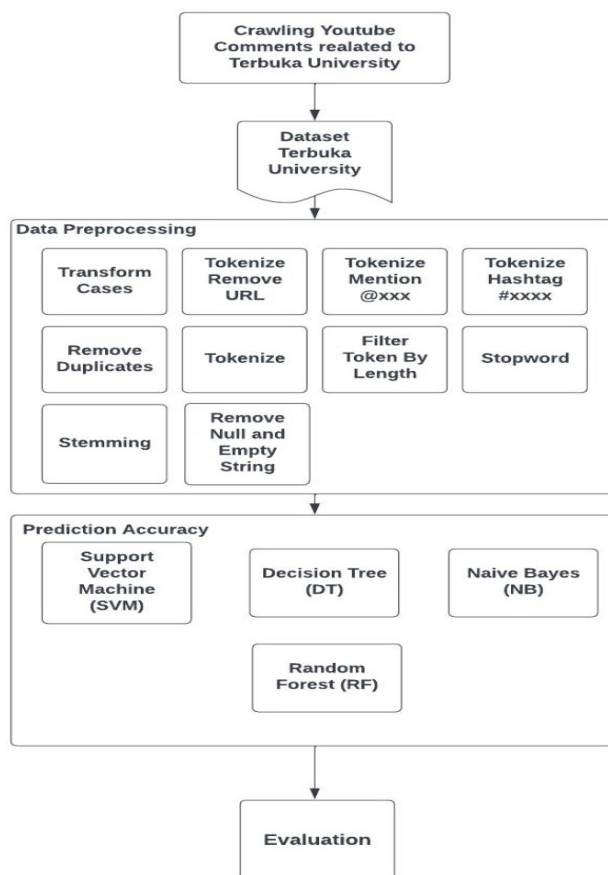
Pada penelitian ini menggunakan pendekatan *metodologi Cross-Industry Standard Process for Data Mining* (CRISP-DM) yang merupakan proses standar terbuka lintas industri untuk penambangan data, metode ini paling banyak digunakan karena penerapannya yang cukup efektif dan memiliki langkah-langkah yang *applicable* (mudah diterapkan)[12]. CRISP-DM juga merupakan standar proses data mining sebagai strategi pemecahan masalah umum dari bisnis ataupun penelitian[13]. Model CRISP-DM terdiri dari enam fase yaitu *business understanding, data understanding, data preparation, modeling, evaluation dan deployment*[14]. Tahapan dalam CRISP-DM terlihat pada Gambar 1. Penelitian ini hanya dilakukan sampai pada fase *evaluation* dan tidak sampai pada fase *deployment* yaitu tahap implementasi dalam sebuah tools.



Gambar 1. Model CRISP-DM[14]

Model CRISP-DM diterapkan dalam penelitian ini, sehingga untuk tahapan penelitiannya disesuaikan menjadi seperti yang ditunjukkan pada Gambar 1. Tahapan penelitian terdiri dari pengumpulan data komentar Youtube, dimana dataset yang digunakan bersumber dari komentar Youtube mengenai komentar Universitas Terbuka yaitu Data *Preprocessing*, *Prediction Accuracy*, dan evaluasi.

Proses diawali dari pengumpulan komentar yang ada dikolom komentar Youtube dengan menggunakan metode *web crawling* dengan kata kunci “Universitas Terbuka”. Sebelum data diproses menggunakan *tools* RapidMiner, dilakukan proses *cleansing* terlebih dahulu dengan menghilangkan kata-kata singkatan dan kata-kata yang tidak memiliki arti. Proses dilanjutkan dengan *pre-processing* menggunakan metode *Transform Cases*, *Tokenize Remove URL*, *Tokenize Mention @xxx*, *Tokenize Hastag #xxx*, *Remove Duplicates*, *Tokenize*, *Filter Tokens (by length)* dan *Filter Stopwords (Dictionary)*. Pembobotan token menggunakan algoritma TF-IDF. Proses selanjutnya adalah *cross validation* menggunakan algoritma *Support Vector Machine (SVM)*, *Decision Tree (DT)*, *Naïve Bayes (NB)*, *Random Forest (RF)*. Bentuk keluaran sistem adalah nilai *accuracy*, *preission* dan *recall* dari *Performance Vector (Performance-SVM)*, *Performance Vector (Performance-DT)*, *Performance Vector (Performance-NB)*, dan *Performance Vector (Performance-RF)*.



Gambar 2. Metodologi Penelitian Pengkategorian Komentar Youtube

2.2 Data Preprocessing

Preprocessing merupakan tahapan penting dalam melakukan text mining. *Preprocessing* digunakan untuk mengubah bentuk dokumen menjadi data yang terstruktur[15]. Tahapan *preprocessing* bertujuan untuk mempersiapkan data sebelum melakukan pembobotan dan klasifikasi menggunakan 4 Algoritma yaitu (SVM, DT, NB, RF).

2.3 Prediction Accuracy

a. Decision Tree (DT)

Decision tree sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat dan banyak di minati. Dalam decision tree ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami dengan bahasa alami. Model pohon keputusan banyak digunakan pada kasus data dengan output yang bernilai diskrit. Walaupun tidak menutup kemungkinan dapat juga digunakan untuk kasus data dengan atribut numeric[16].

b. Support Vector Machine (SVM)

Metode *Support Vector Machine* merupakan metode yang berdasar pada teori pembelajaran statistic dengan tujuan memberikan hasil lebih baik dibandingkan metode lain. *Support Vector Machine* bekerja baik terhadap data yang memiliki dimensi tinggi dengan memanfaatkan teknik kernel. Kernel memiliki fungsi sebagai dasar pembelajaran semua algoritma. Secara umum algoritma ini mengatasi kendala fungsi kernel tertentu, karena mesin linear hanya dapat mengklasifikasi data dalam linear ruang fitur terpisah. Fungsi kernel mendorong sebuah ruang fitur oleh implisit pemetaan[17].

c. Naïve Bayes (NB)

Naïve Bayes merupakan algoritma yang digunakan untuk klasifikasi yang menggunakan teorema bayes dan berasumsi bahwa nilai antar variabel saling bebas (independen) pada suatu nilai output. Dalam hal ini, diasumsikan bahwa kehadiran atau ketiadaan dari suatu variabel tertentu tidak berhubungan dengan kehadiran atau ketiadaan dari variabel lainnya[18].

d. Random Forest (RF)

Random Forest adalah pengembangan dari metode *Decision Tree* yang menggunakan beberapa *Decision Tree*, dimana setiap *Decision Tree* telah dilakukan pelatihan menggunakan sampel individu dan setiap atribut dipecah pada pohon yang dipilih antara atribut subset yang bersifat acak. *Random Forest* memiliki beberapa kelebihan, yaitu dapat meningkatkan hasil akurasi jika terdapat data yang hilang, dan untuk *resisting outliers*, serta efisien untuk penyimpanan sebuah data. Selain itu, *Random Forest* mempunyai proses seleksi fitur dimana mampu mengambil fitur terbaik sehingga dapat meningkatkan performa terhadap model klasifikasi. Dengan adanya seleksi fitur tentu *Random Forest* dapat bekerja pada big data dengan parameter yang kompleks secara efektif[19].

2.4 Evaluation

Evaluation disini akan melihat dan mengevaluasi hasil dari klasifikasi yang ada dengan prediksi yang dilakukan dengan menggunakan beberapa algoritma yang ada.

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Pada tahapan ini, adalah tahap pertama dalam CRISP-DM dan termasuk bagian yang cukup vital. Pada tahap ini membutuhkan pengetahuan dari objek bisnis, bagaimana membangun atau mendapatkan data, dan bagaimana untuk mencocokkan tujuan pemodelan untuk tujuan bisnis sehingga model terbaik dapat dibangun[20]. Pada penelitian ini untuk melakukan kategori pada komentar youtube ini dipersiapkan dari software yang digunakan yaitu rapid miner, data yang diperlukan dari source komentar yang ada pada Youtube.

3.2 Data Understanding

Secara garis besar data understanding bertujuan untuk memeriksa data, sehingga dapat mengidentifikasi masalah dalam data. Tahap ini memberikan fondasi analitik untuk sebuah penelitian dengan membuat ringkasan (summary) dan mengidentifikasi potensi masalah dalam data[20]. Pada penelitian ini dengan tujuan proses pengkategorian adalah untuk mengetahui tingkat akurasi dari beberapa kategori yang ditentukan. Adapun jenis kategori yang dimaksud adalah: Module, Tutorial, Beasiswa, Pengajar, Ujian, Aplikasi, Non Akademik dan Others. Jenis kategori ini adalah klasifikasi yang mempengaruhi kriteria nilai dari kepuasan pelayanan.

3.3 Data Preparation

Secara garis besar untuk memperbaiki masalah dalam data, kemudian membuat *variabel derived*. Tahap ini jelas membutuhkan pemikiran yang cukup matang dan usaha yang cukup tinggi untuk memastikan data tepat untuk algoritma yang digunakan[20]. Pada penelitian ini ditentukan beberapa algoritma yang akan diterapkan yaitu *Decision Tree* (DT), *Support Vector Machine* (SVM), *Naïve Bayes* (NB) dan *Random Forest* (RF).

3.4 Modelling

Secara garis besar untuk membuat model prediktif atau deskriptif. Pada tahap ini dilakukan metode statistika dan *Machine Learning* untuk penentuan terhadap teknik *data mining*, alat bantu *data mining*, dan algoritma *data mining* yang akan diterapkan. Lalu selanjutnya adalah melakukan penerapan teknik dan algoritma data mining tersebut kepada data dengan bantuan alat bantu. Jika diperlukan penyesuaian data terhadap teknik data mining tertentu, dapat kembali ke tahap *data preparation*[20]. Jenis modelling yang diterapkan pada penelitian ini adalah algoritma data mining.

3.5 Deployment

Tahap deployment atau rencana penggunaan model adalah tahap yang paling dihargai dari proses CRISP-DM. Perencanaan untuk Deployment dimulai selama Business Understanding dan harus menggabungkan tidak hanya bagaimana untuk menghasilkan nilai model, tetapi juga bagaimana mengkonversi skor keputusan, dan bagaimana untuk menggabungkan keputusan dalam sistem operasional[20]. Tujuan dari hasil dari tingkat akurasi maka dapat diketahui kategori pendukung untuk kepuasan pelayanan dan kategori mana yang harus dilakukan perbaikan kualitas untuk mencapai tujuan dari pelayanan.

3.6 Tahapan Penelitian

Berikut ini adalah langkah-langkah yang dilakukan yaitu sebagai berikut ini:

a. Pengumpulan Data

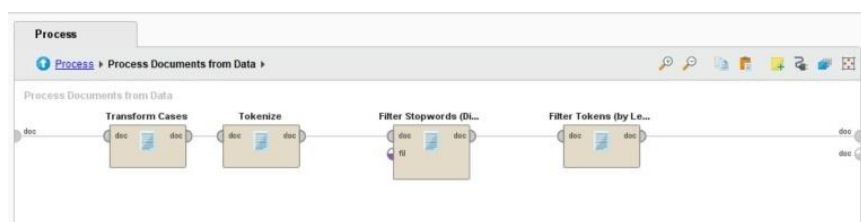
Pada proses ini dilakukan pengambilan data pada situs Youtube. dengan menggunakan API yang disediakan oleh situs Youtube, sehingga video di Youtube dapat diintegrasikan di site yang lain. Dengan adanya API ini user tidak perlu lagi membuka situs Youtube. Youtube API adalah kumpulan fungsi - fungsi yang disediakan oleh pengembang Youtube sehingga memungkinkan konten-konten video Youtube dan fungsionalitasnya dapat diintegrasikan ke situs web, aplikasi perangkat lunak, atau perangkat lainnya. Dengan adanya data API memungkinkan sebuah program dapat memiliki banyak operasi seperti yang ada di website Youtube itu sendiri. Contohnya seperti mencari video, melihat content yang berkaitan. Selain itu, dengan data API sebuah program dapat mengautentikasi user sehingga dapat mengupload video dan menggunakan fitur-fitur lainnya.

b. Data Cleansing

Setelah mendapatkan data komentar yang ada pada Youtube, maka langkah pertama adalah melakukan data cleansing terlebih dahulu dengan tujuan untuk menghilangkan adanya duplikasi data didalamnya. Dari 8113 data termasuk duplikasi dan dilakkan penghapusan sehingga menyisakan 6920 data.

c. Preprocessing

Berikut ini adalah tahapan dalam preprocessing pada penelitian Pengkategorian komentar Youtube terhadap layanan akademik dan non akademik universitas terbuka, yaitu sebagai berikut ini:



Gambar 3. Tahap Preprocessing pada rapid miner

Tools yang digunakan dalam eksperimen ini adalah Rapidminer yang di *update* dengan penambahan *plugin text processing* yang telah memiliki fitur pemrosesan teks di antaranya:

d. Transform Cases

Dengan fitur transform cases kita dapat secara otomatis mengubah semua huruf pada teks menjadi huruf kecil semua atau menjadi huruf kapital semua, pada penelitian ini semua huruf dirubah kedalam huruf kecil karena mayoritas teks berupa tulisan opini yang sebagian besar merupakan huruf kecil semua.

e. Tokenize Remove URL

Langkah berikutnya adalah menghapus URL `http:// www`, remove duplicates, filter token by length, stopword, stemming, Remove null and empty string. Proses ini dilakukan karena sering munculnya URL pada data twitter yang membuat data tidak efektif dan tidak memiliki arti. Penyebab munculnya alamat web atau URL ini adalah karena banyaknya pengguna mempromosikan situs tertentu agar pengguna lain langsung bisa masuk pada halaman web yang dimaksud.

f. Tokenize Mention

Setelah itu dilakukan tokenize mention, yaitu menghapus mention akun yang diawali @teks.

g. Tokenize Hashing

Setelah itu dilakukan kegiatan menghapus hashtag yang diawali # teks.

h. Remove Duplicates

Kemudian dilanjutkan kegiatan menghapus data Duplikat.

i. Tokenize

Lalu tahap berikutnya adalah merubah kalimat menjadi token (per teks). Dimana pada tahapan ini merupakan tahapan memenggal dari setiap kata yang terdapat didalam kalimat, termasuk karakter didalamnya.

j. Filter Length By Token

Setelah itu dilakukan Filter Length by token yaitu hanya mengambil teks dengan length 4 -25.

k. Stopwords

Dan dilanjutkan Stopwords yaitu menghapus teks yang tidak memiliki makna pada proses prediksi dan klasifikasi. Ciri-ciri pada kata ini adalah frekuensi kemunculannya yang jumlahnya cukup banyak dibandingkan dengan kata yang lainnya. Untuk melakukan proses penghapusan kata ini diperlukan sebuah data atau daftar kata yang diinginkan untuk dihapus[21].

l. Stemming

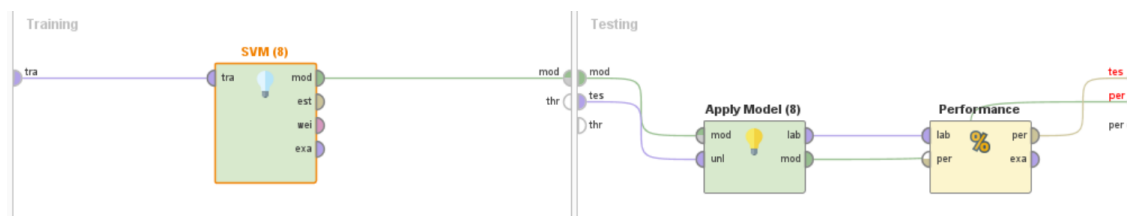
Stemming merupakan suatu proses yang terdapat dalam sistem IR yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (rootword) dengan menggunakan aturanaturan tertentu. Sebagai contoh, kata bersama, kebersamaan, menyamai, akan distem ke root wordnya 'DLWX³VDPD'. Proses stemming pada teks Bahasa Indonesia berbeda dengan stemming pada teks berbahasa Inggris. Pada teks berbahasa Inggris, proses yang diperlukan hanya proses menghilangkan sufiks. Sedangkan pada teks berbahasa Indonesia, selain sufiks, prefiks, dan konfiks juga dihilangkan[22].

m. Remove Null and Empty String

Lalu dilanjutkan dengan Remove null and empty string, yaitu menghapus teks kosong/string kosong

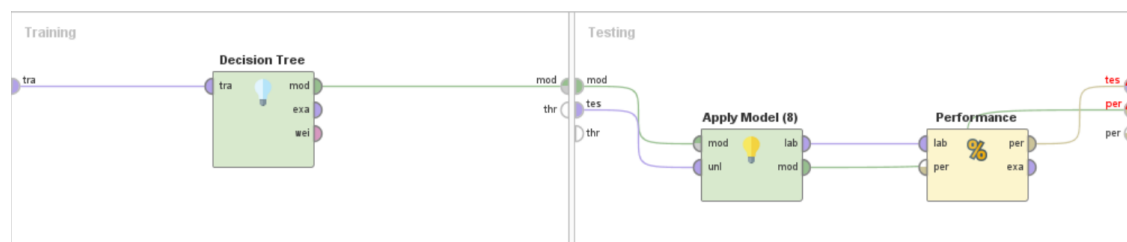
3.7 Cross Validation

Langkah berikutnya dilakukan proses *cross validation* menggunakan algoritma SVM, DT, NB, dan RF yang meliputi proses *training* dan *testing* (*apply model* dan *Performance Vector* (*Performance-SVM*), *Performance Vector* (*Performance-DT*), *Performance Vector* (*Performance-NB*), dan *Performance Vector* (*Performance-RF*). Setelah itu dilakukan proses pemodelan dengan menggunakan Algoritma SVM, DT, NB dan RF untuk mendapatkan nilai *accuracy*, *precision* dan *recall* berdasarkan *performance vector* (*Performance Vector* (*Performance-SVM*), *Performance Vector* (*Performance-DT*), *Performance Vector* (*Performance-NB*), dan *Performance Vector* (*Performance-RF*).



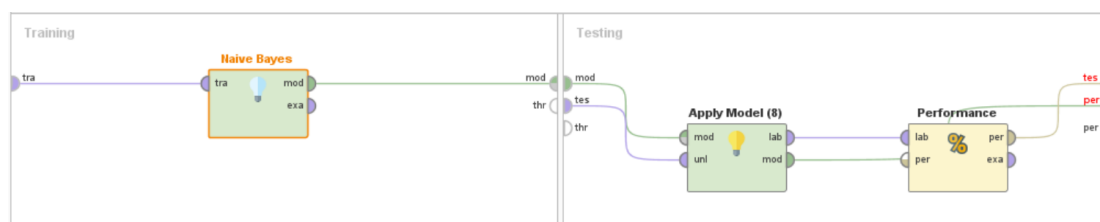
Gambar 4. Cross Validation Process-SVM

Terdiri dari 2 tahap, yaitu tahap *training* dan *testing*. Pada tahap *training*, data dilatih menggunakan algoritma SVM. Kemudian pada tahap *testing*, data diproses melalui *Apply Model* dan *Performance Test*.



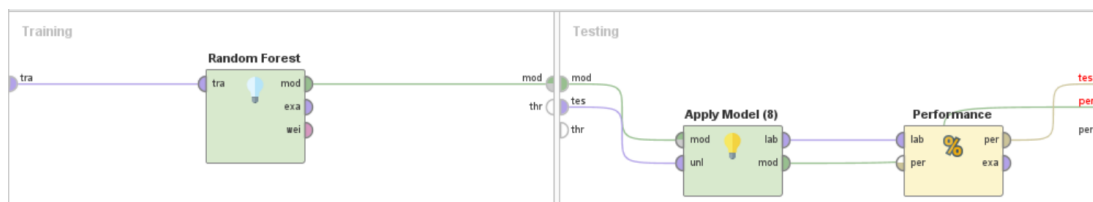
Gambar 5. Cross Validation Process-DT

Terdiri dari 2 tahap, yaitu tahap *training* dan *testing*. Pada tahap *training*, data dilatih menggunakan algoritma DT. Kemudian pada tahap *testing*, data diproses melalui *Apply Model* dan *Performance Test*.



Gambar 6. Cross Validation Process-NB

Terdiri dari 2 tahap, yaitu tahap *training* dan *testing*. Pada tahap *training*, data dilatih menggunakan algoritma NB. Kemudian pada tahap *testing*, data diproses melalui *Apply Model* dan *Performance Test*.



Gambar 7. Cross Validation Process-RF

Terdiri dari 2 tahap, yaitu tahap *training* dan *testing*. Pada tahap *training*, data dilatih menggunakan algoritma RF. Kemudian pada tahap *testing*, data diproses melalui *Apply Model* dan *Performance Test*.

Dari proses *cross validation* yaitu *training* dan *testing*, pada 4 Algoritma yaitu SVM, DT, NB dan RF menghasilkan nilai *accuracy*, *precision*, dan *recall* sebagai berikut ini:

No	Category	Decision Tree				SVM				Naive Bayes				Random Forest			
		Accuracy	Precision	Recall	AUC	Accuracy	Precision	Recall	AUC	Accuracy	Precision	Recall	AUC	Accuracy	Precision	Recall	AUC
1	Modul	99,37	93,48	81,13	90,50%	97,64	62,50	18,87	0,98	73,12	2,09	20,75	50,00%	97,45	0,00	0,00	98,60%
2	Aplikasi	100,00	100,00	100,00	100,00%	97,88	64,71	22,45	0,99	81,93	11,44	100,00	50,00%	97,64	0,00	0,00	98,80%
3	Pengajar	99,42	92,00	85,19	92,50%	97,54	61,54	14,81	0,98	69,25	2,90	33,33	50,00%	97,40	0,00	0,00	96,60%
4	Tutorial	57,69	92,55	58,66	37,10%	92,48	93,02	99,27	0,94	57,69	92,55	58,66	37,10%	91,86	91,86	100,00	81,20%
5	Ujian	99,71	0,00	0,00	50,00%	99,61	0,00	0,00	0,77	89,98	0,49	16,67	50,00%	99,71	0,00	0,00	93,70%
6	Non Akademik	99,90	100,00	95,65	97,80%	97,98	83,33	10,87	0,97	75,39	2,45	26,09	50,00%	97,78	0,00	0,00	96,10%
7	Others	96,58	95,94	100,00	91,10%	85,74	85,39	99,34	0,94	46,05	87,18	38,94	53,00%	80,78	80,78	100,00	93,00%

Gambar 8. Nilai Rekapitulasi Accuracy, Precision, dan Recall Algoritma SVM, DT, NB dan RF

3.8 Klasifikasi pada 7 Kategori

Berikut ini adalah klasifikasi pada 7 kategori (Module, Tutorial, Beasiswa, Pengajar, Ujian, Aplikasi, Non Akademik dan Others) dengan menerapkan 4 algoritma (SVM, DT, NB dan RF).

a. Hasil Klasifikasi Decision Tree

Pada penelitian ini, akan dilakukan pengujian akurasi dengan *Decision Tree* yang ditunjukkan pada penjabaran tabel sebagai berikut ini.

Tabel 1. Klasifikasi Decision Tree

No	Category	Decision Tree			
		Accuracy	Precision	Recall	AUC
1	Modul	99,37	93,48	81,13	90,50%
2	Aplikasi	100,00	100,00	100,00	100,00%
3	Pengajar	99,42	92,00	85,19	92,50%
4	Tutorial	57,69	92,55	58,66	37,10%
5	Ujian	99,71	0,00	0,00	50,00%
6	Non Akademik	99,90	100,00	95,65	97,80%
7	Others	96,58	95,94	100,00	91,10%

Pada pengujian akurasi dengan *Decision Tree* menunjukkan bahwa kategori yang mempunyai akurasi tertinggi adalah Aplikasi. Dilanjutkan dengan Non Akademik, Ujian, pengajar, Modul, Others, dan Tutorial.

b. Hasil Klasifikasi Support Vector Machine (SVM)

Tabel 2. Klasifikasi Support Vector Machine

No	Category	SVM			
		Accuracy	Precision	Recall	AUC
1	Modul	97,64	62,50	18,87	0,98
2	Aplikasi	97,88	64,71	22,45	0,99
3	Pengajar	97,54	61,54	14,81	0,98
4	Tutorial	92,48	93,02	99,27	0,94
5	Ujian	99,61	0,00	0,00	0,77
6	Non Akademik	97,98	83,33	10,87	0,97
7	Others	85,74	85,39	99,34	0,94

Pada pengujian akurasi dengan *Support Vector Machine* menunjukkan bahwa Kategori yang mempunyai akurasi tertinggi adalah Ujian. Dilanjutkan dengan Non Akademik, Aplikasi, Modul, Pengajar, Tutorial dan Others.

c. Hasil Klasifikasi Naïve Bayes

Tabel 3. Klasifikasi *Naïve Bayes*

No	Category	Naïve Bayes			
		Accuracy	Precision	Recall	AUC
1	Modul	73,12	2,09	20,75	50,00%
2	Aplikasi	81,93	11,44	100,00	50,00%
3	Pengajar	69,25	2,90	33,33	50,00%
4	Tutorial	57,69	92,55	58,66	37,10%
5	Ujian	89,98	0,49	16,67	50,00%
6	Non Akademik	75,39	2,45	26,09	50,00%
7	Others	46,05	87,18	38,94	53,00%

Pada pengujian akurasi dengan *Naïve Bayes* menunjukkan bahwa Kategori yang mempunyai akurasi tertinggi adalah Ujian. Dilanjutkan dengan Aplikasi, Non akademik, Pengajar, Modul, Tutorial, dan *Others*.

d. Hasil Klasifikasi Random Forest

Tabel 4. Klasifikasi *Random Forest*

No	Category	Random Forest			
		Accuracy	Precision	Recall	AUC
1	Modul	97,45	0,00	0,00	98,60%
2	Aplikasi	97,64	0,00	0,00	98,80%
3	Pengajar	97,40	0,00	0,00	96,60%
4	Tutorial	91,86	91,86	100,00	81,20%
5	Ujian	99,71	0,00	0,00	93,70%
6	Non Akademik	97,78	0,00	0,00	96,10%
7	Others	80,78	80,78	100,00	93,00%

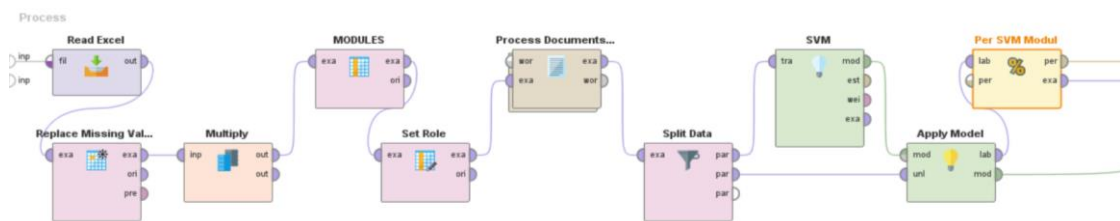
Pada pengujian akurasi dengan *Random Forest* menunjukkan bahwa Kategori yang mempunyai akurasi tertinggi adalah Ujian. Dilanjutkan dengan Non Akademik, Aplikasi, Modul, Pengajar, Tutorial, dan *Others*.

3.5 Modelling Metode

Pada tahap yang ke 3 ini dataset yang telah dibuat pada tahapan sebelumnya digunakan sebagai masukan untuk algoritma. Dimana diterapkan 4 Algoritma yaitu (SVM, DT, NB, RF).

a. Proses Modelling SVM

Berikut adalah desain model SVM dengan menggunakan tools Rapidminer untuk tujuh kategori yang digunakan yaitu (“*Module, Tutorial, Beasiswa, Pengajar, Ujian, Aplikasi, Non Akademik dan Others*”).

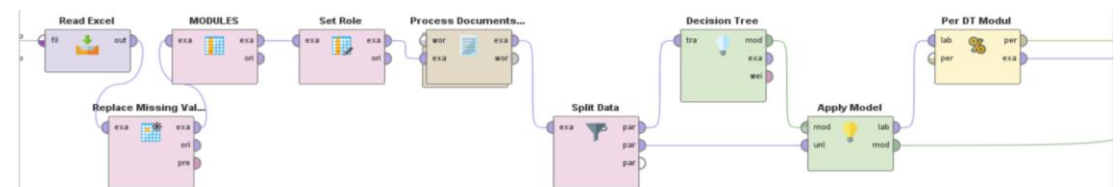


Gambar 9. Desain model SVM

Berdasarkan gambar diatas, dapat dilihat desain model algoritma SVM dengan ketujuh kategori. Pada desain model tersebut diawali dengan operator *Read Excel, Multiply, Select Attribute, Set Role, Process Document, weight by value Average, Select by Weight* dan *Performance* pada masing-masing kategorinya.

b. Proses Modelling DT

Berikut adalah desain model DT dengan menggunakan tools Rapidminer untuk tujuh kategori yang digunakan yaitu (“*Module, Tutorial, Beasiswa, Pengajar, Ujian, Aplikasi, Non Akademik dan Others*”).

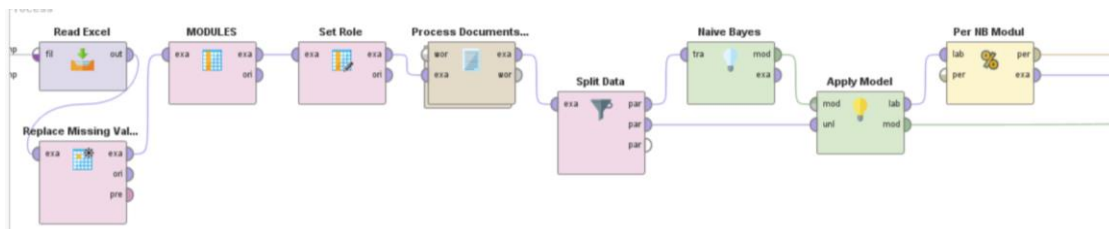


Gambar 10. Desain model DT

Berdasarkan gambar diatas, dapat dilihat desain model algoritma DT dengan ketujuh kategori. Pada desain model tersebut diawali dengan operator *Read Excel*, *Multiply*, *Select Attribute*, *Set Role*, *Process Document*, *weight by value Average*, *Select by Weight* dan *Performance* pada masing-masing kategorinya.

c. Proses Modelling NB

Berikut adalah desain model NB dengan menggunakan tools Rapidminer untuk tujuh kategori yang digunakan yaitu (*“Module, Tutorial, Beasiswa, Pengajar, Ujian, Aplikasi, Non Akademik dan Others”*).

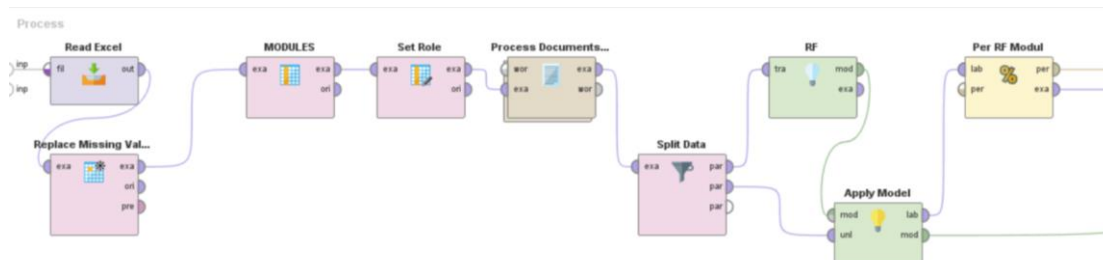


Gambar 11. Desain model NB

Berdasarkan gambar diatas, dapat dilihat desain model algoritma NB dengan ketujuh kategori. Pada desain model tersebut diawali dengan operator *Read Excel*, *Multiply*, *Select Attribute*, *Set Role*, *Process Document*, *weight by value Average*, *Select by Weight* dan *Performance* pada masing-masing kategorinya.

d. Proses Modelling RF

Berikut adalah desain model RF dengan menggunakan tools Rapidminer untuk tujuh kategori yang digunakan yaitu (*“Module, Tutorial, Beasiswa, Pengajar, Ujian, Aplikasi, Non Akademik dan Others”*).



Gambar 12. Desain model RF

Berdasarkan gambar diatas, dapat dilihat desain model algoritma RF dengan ketujuh kategori. Pada desain model tersebut diawali dengan operator *Read Excel*, *Multiply*, *Select Attribute*, *Set Role*, *Process Document*, *weight by value Average*, *Select by Weight* dan *Performance* pada masing-masing kategorinya.

4. KESIMPULAN

Berdasarkan analisis dan pengujian yang dilakukan dalam rangka menganalisa prediksi kepuasan layanan pada Universitas Terbuka dengan data pada komentar Youtube maka dapat diketahui dari beberapa kategori (Module, Tutorial, Beasiswa, Pengajar, Ujian, Aplikasi, Non Akademik dan *Others*) dengan menerapkan 4 Algoritma yaitu *Decision Tree* (DT), *Support Vector Machine* (SVM), *Naïve Bayes* (NB) dan *Random Forest* (RF) untuk menentukan *Prediction Accuracy*. Dan menghasilkan akurasi tertinggi pada kategori modul dengan nilai akurasi 100% dengan menggunakan algoritma DT, kategori Non Akademik dengan nilai akurasi 99,90% dengan menerapkan algoritma DT. kategori Ujian dengan nilai akurasi 99,70% dengan menerapkan algoritma RF, kategori Pengajar dengan nilai akurasi 99,42% dengan menerapkan algoritma DT, kategori modul dengan nilai akurasi 99,37% dengan menerapkan algoritma DT, kategori *others* dengan nilai akurasi 96,58% dengan menerapkan algoritma DT, dan kategori tutorial dengan nilai akurasi 92,40% dengan menerapkan algoritma SVM.

REFERENCES

- [1] Andriana dan Chandra Tjiptono, *Brand Manajemen dan Strategi*. Yogyakarta: Elex Media Komputindo., 2008.
- [2] S. Amin, “Strategi Peningkatan Kualitas Pelayanan Akademik Pada Sekolah Tinggi,” *Wahana Akad.*, vol. 4, no. 2, pp. 194–202, 2017.
- [3] Marthalina, “Analisis Kualitas Pelayanan Akademik Dan Kepuasan Mahasiswa Di Ipdn Kampus Jakarta,” *J. Manaj. Sumber Daya Mns.*, vol. 5, no. 1, pp. 1–18, 2018.
- [4] Universitas Terbuka, “Sejarah UT.” <https://www.ut.ac.id/sejarah-ut>.
- [5] Universitas Terbuka, “Visi & Misi.” <https://www.ut.ac.id/visi-misi>.
- [6] Universitas Terbuka, “Rencana Strategis.” <https://www.ut.ac.id/rencana-strategis>.
- [7] M. P. Munthe, A. S. R. Ansori, and ..., “Analisis Sentimen Komentar Pada Saluran Youtube Food Vlogger Berbahasa Indonesia Menggunakan Algoritma Naïve Bayes,” *eProceedings Eng.*, vol. 8, no. 6, pp. 11909–11916, 2021, [Online].



- Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/16897>.
- [8] D. H. Jayani, "Orang Indonesia Habiskan Hampir 8 Jam untuk Berinternet," Mar. .
- [9] R. F. BUSYRA, R. Primartha, and H. Satria, "Opinion Mining Pada Komentar Youtube Menggunakan Algoritma K-Means," 2018, [Online]. Available: <https://repository.unsri.ac.id/7418/>.
- [10] P. Y. Saputra, D. H. Subhi, and F. Z. A. Winatama, "Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah Di Youtube Menggunakan Algoritma Naïve Bayes," *J. Inform. Polinema*, vol. 5, no. 4, pp. 209–213, 2019, doi: 10.33795/jip.v5i4.259.
- [11] Balya, "Analisis Sentimen Pengguna Youtube Di Indonesia Pada Review Smartphone Menggunakan Naïve Bayes," *Skripsi Univ. Sumatera Utara*, pp. 4–16, 2019.
- [12] Mas Raden Panca Rizqi Wahyu Atmaja Kusuma and Yustanti Wiyli, "Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers)," *Jeisbi*, vol. 02, no. 03, pp. 55–62, 2021.
- [13] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, pp. 1120–1126, 2021, doi: 10.29207/resti.v5i6.3588.
- [14] A. D. Adhi Putra, "Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2021, doi: 10.35957/jatisi.v8i2.962.
- [15] Kurniawan, Taufik, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan Naive Bayes Classifier dan Support Vector Machine," pp. 27–30, 2017.
- [16] Sari Dewi, "KOMPARASI 5 METODE ALGORITMA KLASIFIKASI DATA MINING PADA PREDIKSI KEBERHASILAN PEMASARAN PRODUK LAYANAN PERBANKAN," *J. Techno Nusa Mandiri*, vol. XIII, no. No.1 Maret 2016, 2016.
- [17] A. Bode, "Support Vector Machine Menggunakan Forward Selection untuk Prediksi Penjualan Obat," *Tecnoscienza*, vol. 3, pp. 16–26, 2018.
- [18] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.
- [19] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.
- [20] BINUS UNiversity, "Cross-Industry Standard Process for Data Mining (CRISP-DM)," 2020. <https://mmsi.binus.ac.id/2020/09/18/cross-industry-standard-process-for-data-mining-crisp-dm/#:~:text=dijelaskan sebagai berikut %3A-,Business Understanding,sehingga model terbaik dapat dibangun.>
- [21] L. K. Harsono, Y. Alkhalifi, Nurajijah, and W. Gata, "Analisis Sentimen Stakeholder atas Layanan haiDJPb pada Media Sosial Twitter Dengan Menggunakan Metode Support Vector Machine dan Naïve Bayes," *J. Ilmu-ilmu Inform. dan Manaj.*, vol. 14, no. 1, pp. 36–44, 2020.
- [22] Hafiz Ridha Pramudita, "PENERAPAN ALGORITMA STEMMING NAZIEF & ADRIANI DAN SIMILARITY PADA PENERIMAAN JUDUL THESIS," *J. Ilm. DASI*, vol. 15 no.04, pp. 15–19, 2001.