



OPTIMASI FEATURE SELECTION PADA KOMENTAR MEDIA SOSIAL TERHADAP PERALIHAN TV DIGITAL MENGGUNAKAN NAÏVE BAYES, SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBOR

Nova Tri Romadloni^{1)*}, Nisa Dwi Septiyanti¹⁾

¹⁾Universitas Muhammadiyah Karanganyar, Karanganyar, Indonesia

Email: novatrir@umuka.ac.id

Abstrak

Dalam menghadapi perubahan pada salah satu media informasi yaitu televisi yang semula menggunakan *signal analog* beralih menjadi TV dengan *signal digital*. Hal tersebut disebabkan bahwa siaran dengan transmisi *analog* rentan terhadap gangguan sehingga peralihan ini sebagai upaya dalam menikmati konten siaran televisi yang lebih baik. Namun, pada kenyataannya terdapat beberapa kesulitan yang dialami oleh beberapa pihak atau kalangan dengan berbagai alasan. Hal tersebut dapat diketahui melalui platform media sosial seperti twitter dan instagram. Dengan adanya kejadian tersebut maka dapat diambil beberapa komentar positif dan negatif untuk mengetahui dampak dari peralihan *signal digital* tersebut. Dalam penelitian ini terdapat 1177 data komentar yang didapatkan sehingga membutuhkan teknologi untuk mendeteksi komentar tersebut positif atau negatif. Pada penelitian ini untuk optimasi komentar berbasis *Pearson Correlation* dengan menggunakan metode *Naïve Bayes*, *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (KNN). Hasil akurasi yang didapatkan dari ujicoba tersebut dengan metode *Naïve Bayes* 61,22%, SVM 80,10 %, dan KNN 79,93%. Jika ditambahkan dengan *Feature Selection* mendapatkan hasil *Naïve Bayes* 63,68%, SVM 80,19%, dan KNN 80,02%. Berdasarkan hasil percobaan yang dilakukan digunakan untuk mengetahui berapa banyak perbandingan komentar positif dan negatif serta mengetahui perbandingan dari beberapa macam algoritma dengan seleksi fitur sehingga dapat menjadikan hasil yang optimal.

Kata kunci: data mining; k-nearest neighbor; naïve bayes; pearson correlation; support vector machine.

OPTIMIZATION OF FEATURE SELECTION ON SOCIAL MEDIA COMMENTS ON DIGITAL TV SWITCHING USING NAÏVE BAYES, SUPPORT VECTOR MACHINE AND K-NEAREST NEIGHBOR

Abstract

In the face of changes in one of the information media, namely television which originally used analog signals turned into TVs with digital signals. This is because broadcasts with analog transmissions are prone to interference so this switch is an effort to enjoy better television broadcast content. However, in reality there are some difficulties experienced by some parties or circles for various reasons. This can be known through social media platforms such as Twitter and Instagram. With this work, several positive and negative comments can be taken to find out the impact of the digital signal switch. In this study, there were 1177 comment data obtained so that technology is needed to detect positive or negative comments. In this study for comment optimization based on Pearson Correlation using Naïve Bayes, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) methods. The accuracy results obtained from the trial with the Naïve Bayes method were 61.22%, SVM 80.10%, and KNN 79.93%. When added with Feature Selection get Naïve Bayes results 63.68%, SVM 80.19%, and KNN 80.02%. Based on the results of the experiments conducted, it is used to find out how many positive and negative

comments are compared and find out the comparison of several kinds of algorithms with feature selection so that it can make optimal results.

Keywords: *data mining; k-nearest neighbor; naïve bayes; pearson correlation; support vector machine.*

Submitted: 2 Februari 2023

Reviewed: 22 Februari 2023

Accepted: 12 Maret 2023

Published: 29 April 2023

PENDAHULUAN

Penyesuaian terhadap perkembangan teknologi sering kali menjadikan permasalahan dalam kehidupan sehari-hari. Penerapan perubahan dengan segala kebijakan menimbulkan pro dan kontra diberbagai kalangan dan daerah di Indonesia. Salah satu media informasi yang sangat populer bahkan setiap kebanyakan orang atau mayoritas memiliki televisi sebagai media informasi maupun hiburan. Semenjak TVRI mengudara pada tahun 1962, televisi menjadi salah satu media hiburan dan tidak dapat dilepaskan dari keseharian di Indonesia (Permana et al., 2019). Kegiatan menonton televisi sekarang memiliki perbedaan yang signifikan. Dalam hal ini TV dinilai memberikan nilai guna karena dapat menyampaikan informasi, hiburan serta pendidikan secara efektif, cepat dan murah (Suryadi et al., n.d.). Siaran televisi di Indonesia sebelumnya menerapkan sistem siaran TV analog. Namun penyiaran dengan menggunakan signal analog ini dinilai tertinggal dengan adanya peradaban perkembangan teknologi. Oleh karena itu, Menteri Komunikasi dan Informatika (Menkominfo) mulai melakukan uji coba untuk memberhentikan sistem siaran TV analog dan menggantinya dengan *signal* TV digital melalui Peraturan Menteri Nomor 07/P/M.KOMINFO/3/2007 sehingga dengan adanya peraturan penyiaran tersebut diharapkan dapat menjadi peluang perkembangan industri penyiaran dimasa mendatang (Rasidin et al., 2020a).

Kebijakan penyiaran kearah digital yang dilakukan oleh pemerintah memungkinkan penataan spektrum frekuensi untuk meningkatkan efisiesi kinerja infrastruktur penyiaran serta mampu menghadirkan tambahan *channel* televisi (Agussetianingsih & Kasim, 2021). Selain itu, manfaat lain yang dapat dirasakan dengan teknologi penyiaran digital memberikan kualitas gambar yang lebih baik dibandingkan dengan TV analog. Tidak hanya kualitas gambar melainkan suara yang lebih jernih dan teknologi yang lebih canggih yang terdapat pada TV digital (Gultom, 2018). Dilansir dari media berita *kompas.com* manfaat atau keuntungan yang didapatkan dari migrasi TV digital dari segi kualitas siaran akan menghadirkan kualitas gambar yang sangat jelas dengan resolusi tinggi, suara jernih dan teknologi canggih, kemudian dapat menghadirkan banyak program siaran yang lebih berkualitas dan bermutu untuk masyarakat, lebih tahan terhadap cuaca di Indonesia yang kurang menentu seperti jika cuaca buruk pada TV analog akan menampilkan gambar layaknya “semut” dalam layar, tetapi jika menggunakan TV digital jangkauan sinyal lebih luas. Selanjutnya manfaat yang didapat lagi untuk TV digital akan memungkinkan mendapat beberapa channel tambahan yang sebelumnya tidak ada pada TV analog.

Uji coba penyelenggaraan TV digital di Indonesia sudah dimulai sejak tahun 2008, selanjutnya sebagai upaya pemerintah dalam beradaptasi dengan perkembangan teknologi yang begitu masif, ditetapkanlah peraturan Menkominfo No.22/PER/M.KOMINFO/11/2011 tentang Penyelenggaraan Penyiaran Televisi Digital Terrestrial Penerimaan Tetap Tidak Berbayar, pada awal tahun 2012 disusul peraturan Menkominfo No.5/PER/M.KOMINFO/2/2012 tentang standar penyiaran TV Digital Terrestrial Penerimaan Tetap Tidak Berbayar (Rasidin et al., 2020b). Kemudian, dengan ditetapkannya Undang-Undang Cipta Kerja No.11 Tahun 2022 menjadi dasar hukum dimulainya proses peralihan penyiaran digital secara bertahap kepada masyarakat di berbagai wilayah di Indonesia (Agussetianingsih & Kasim, 2021). Keluranya peraturan tersebut menuai banyak kritik dari masyarakat. Hal tersebut dapat diketahui dengan

kelas pada media sosial. Media sosial merupakan salah satu cara masyarakat untuk menyampaikan suatu pemikiran, pendapat, masukan atau aspirasi yang ada tentang hal yang dirasakan secara langsung maupun melalui media perantaranya Masyarakat dari berbagai kalangan dan daerah menyampaikan beberapa tanggapan mengenai perubahan tersebut. Ada komentar yang positif mendukung perubahan signal tersebut karena dapat dirasakan manfaatnya terhadap signal TV digital. Selain itu ada juga banyak yang mengeluhkan peralihan TV menjadi TV digital ini salah satunya karena hilangnya siaran dan sedikitnya saluran televisi yang didapatkan atau signal TV digital atau TV yang belum bisa menangkap signal digital maka membutuhkan alat bantu seperti *Set Top Box* yang menjadikan pengguna harus membeli perangkat tambahan, *Set Top Box* atau biasa disingkat STB ini semacam *converter* yang berfungsi mengubah signal televisi digital dapat diterima analog untuk bisa membacanya.

Harga STB yang cukup mahal menjadi masalah yang banyak dikeluhkan oleh masyarakat yang mengalami kesulitan ekonomi. Dilansir dari situs resmi kominfo.go.id, sebagai upaya mendukung perpindahan TV analog ke TV digital pemerintah telah menyediakan bantuan STB bagi masyarakat yang kurang mampu secara gratis. Pendistribusian STB untuk masyarakat dilakukan secara bertahap. Sedangkan masyarakat yang mampu akan tetap didorong untuk segera beralih ke televisi digital. Akan tetapi, dalam peralihan ini masyarakat membutuhkan proses seperti signal yang ditangkap tidak merata atau saluran (*channel*) yang belum bisa muncul pada televisi sehingga menyebabkan rentan waktu menunggu kebijakan supaya lebih baik dengan adanya peralihan signal digital. Peralihan ini menjanjikan beragam keuntungan tetapi dapat membawa konsekuensi yang tidak sederhana dari berbagai aspek, seperti politik, sosial dan budaya.

Banyaknya tanggapan masyarakat di sosial media khususnya di twitter dan instagram, baik itu tanggapan positif, maupun negatif terkait dengan kebijakan pemerintah menetapkan peralihan TV Digital, membuat peneliti tertarik untuk melakukan analisis sentiment. Analisis sentiment merupakan kegiatan mengekstrak, mengolah dan mempelajari data tekstual untuk mendapatkan informasi dan melihat tanggapan mengenai sebuah objek atau masalah yang bernilai negative, netral, maupun positif (Sari & Wibowo, 2019). Dengan melakukan analisis sentiment dapat membantu dalam menemukan masalah yang terjadi sehingga dapat digunakan sebagai acuan untuk mengambil kebijakan selanjutnya.

Penelitian terdahulu terkait dengan analisis sentimen telah banyak dilakukan, diantaranya penggunaan algoritma Naïve Bayes dengan seleksi fitur *Chi Square* untuk mengkaji opini masyarakat tentang migrasi TV Digital. Pada penelitian tersebut terdapat 191 opini positif dan 185 opini negatif dengan menghasilkan akurasi 96%, *precision* 93% dan *recall* 100% (Zuliana et al., 2022). Studi lain melakukan perbandingan tiga metode, yaitu KNN, Logistic Regression dan Naïve Bayes. Dalam penelitian tersebut mengkaji opini masyarakat mengenai prediksi peluang kesembuhan bagi pasien covid-19 dengan hasil perbandingan akurasi KNN 75%, Logistic Regression 70,8% dan Naïve Bayes 70%. Dimana metode KNN memiliki akurasi yang paling tinggi (Romadhon & Kurniawan, 2021).

Perbandingan metode lainnya terdapat pada penelitian terkait analisis sentiment terhadap PT PAL Indonesia yang juga menggunakan 3 metode untuk penelitiannya yaitu KNN, Naïve Bayes dan Decision Tree. Menghasilkan akurasi masing-masing metode KNN 83,38 %, Naïve Bayes 84,08% dan Decision Tree sebesar 81,09 % (Pattiiha & Hendry, 2022). Studi lain juga melakukan penelitian yang menekankan seleksi fitur pearson correlation pada analisis sentiment berbelanja online. Pada penelitian tersebut menunjukkan penerapan fitur pearson correlation menungkatkan kinerja sistem mining dan menghasilkan nilai akurasi yang lebih optimal (Romadloni & Pardede, 2019).

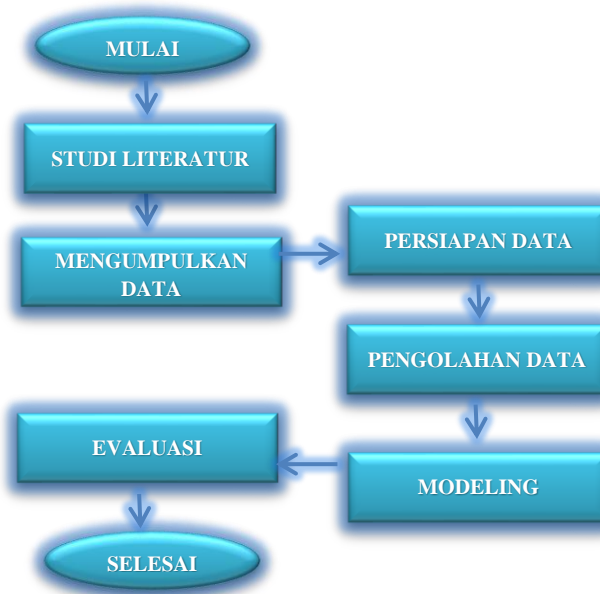
Penelitian ini bertujuan untuk mengetahui pendapat atau tanggapan masyarakat di media sosial twitter dan instagram. Data dikumpulkan dalam kurun waktu Januari 2022 - Desember 2022 dengan memanfaatkan teknologi text mining metode yaitu Support Vector Machine, K-

Nearest Neighbor dan Naïve Bayes. Selain menguji tiga metode, juga dilakukan perbandingan ketika menggunakan *future selection* pada masing-masing metode. Diharapkan penelitian ini dapat mengetahui hasil nilai akurasi yang paling baik dari masing-masing metode, selain itu penelitian ini dapat membantu mengetahui permasalahan terkait dengan regulasi peralihan televisi analog menjadi digital. Di harapkan penelitian ini dapat memberikan kontribusi sebagai salah satu pertimbangan dalam keputusan di masa yang akan datang.

METODE

Pengambilan *sampling* merupakan aspek penting dalam proses analisis sentiment untuk dijadikan objek atau sumber data dalam observasi. Teknik *sampling* pada penelitian ini menggunakan metode *Non-Random sampling*. Pengumpulan data menggunakan software RapidMiner dengan menginputkan API Twitter kemudian untuk sosial media Instagram pengambilan data dengan menggunakan DataMiner yang terdapat pada browser *extension chrome*.

Penelitian ini termasuk penelitian deskriptif kuantitatif dimana data yang diperoleh disusun dan dianalisis sesuai dengan kebutuhan peneliti. Langkah-langkah penelitian dilakukan untuk memenuhi prosedur agar hasil yang didapatkan sesuai dengan tujuan penelitian. Adapun gambaran dari prosedur dalam penelitian ini dijabarkan dalam gambar 1:



Gambar 1. Tahapan Penelitian

Adapun tahapan dalam penelitian ini dapat dijabarkan sebagai berikut: (1) Studi literatur: peneliti membaca beberapa penelitian yang telah di lakukan sebelumnya, membandingkan satu dengan yang lain bertujuan untuk mengetahui dan mempertimbangkan hasil dari setiap penelitian serta berupaya untuk melakukan penelitian yang sebelumnya belum pernah dilakukan; (2) Mengumpulkan data: data diambil atau *crawling* dari media sosial twiter dan instagram. Pengambilan data dilakukan secara bertahap sehingga didapatkan data sebanyak 1177 data; (3) Mengolah data: tahapan ini mengubah struktur dari data kedalam format yang sesuai untuk dilakukan mining; (4) Persiapan data: dalam tahapan ini merupakan tahapan yang penting yaitu mengurangi atribut yang dinilai kurang berpengaruh terhadap klasifikasi data seperti waktu, lokasi, nama akun, menghapus data yang rangkap atau sama dan lain sebagainya. Data ini masih mentah dan kotor sehingga melalui tahapan persiapan data akan menjadikan data

yang berkualitas untuk mempermudah proses klasifikasi. Setelah itu memberikan label pada data tersebut yang berisikan komentar tersebut negatif atau positif; (5) Modeling: penggunaan *pearson correlation* yang menentukan nilai yang dihasilkan berkaitan dengan data tersebut bernilai positif atau negatif. Nilai yang dihasilkan pada *Pearson's* terletak pada $[-1;1]$, untuk nilai -1 yang berarti korelasi negatif sempurna karena satu variabel meningkat, yang lainnya menurun, sedangkan +1 berarti korelasi positif sempurna dan 0 yang berarti tidak ada korelasi linier antara kedua variabel tersebut (Isaac Obilor et al., 2018); dan (6) Evaluasi: pada penelitian ini metode evaluasi yang digunakan adalah nilai akurasi.

Cross validation dalam tools yang digunakan memuat tiga jenis klasifikasi untuk menguji efektifitas metode yaitu, *Support Vector Machine*, *K-Nearest Neighbor* dan *Naïve Bayes*, *K-Nearest Neighbour* (KNN). SVM merupakan salah satu algoritma supervised learning yang cukup populer yang digunakan dalam melakukan ekstraksi informasi (Herianto, 2018) Cara kerja SVM dengan menemukan hyperplane yang mempunyai margin maksimum. Dalam implementasinya, data perlu di pisahkan dalam dua kelas yaitu kelas positif dan kelas negative. SVM dapat memproses dataset yang berdimensi tinggi dengan menggunakan kernel trik. Algoritma kedua yang akan digunakan adalah Algoritma Naïve Bayes. Algoritma Naïve Bayes menggunakan konsep dasar Theorema Bayes dimana dalam melakukan klasifikasi menggunakan perhitungan probabilitas. Algoritma Naïve Bayes memprediksi peluang dimasa depan berdasarkan pengetahuan dimasa lalu. Pada sebuah dataset setiap baris diasumsikan sebagai vector yang memiliki label/kelas. Keuntungan menggunakan metode Naïve Bayes adalah hanya membutuhkan sejumlah kecil data pelatihan untuk memperkirakan parameter (Pattihha & Hendry, 2022). Algoritma ketiga menggunakan *K-Nearest Neighbour* (KNN). Algoritma KNN merupakan algoritma yang dapat digunakan untuk melakukan klasifikasi berdasarkan data pembelajaran yang memiliki jarak paling dekat dengan objek tersebut. Pada penerapannya, algoritma ini melakukan penyimpanan fitur dan klasifikasi dari data pembelajaran. KNN diklasifikasikan berdasarkan mayoritas atau kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi. Prinsip dasar dari algoritma KNN adalah mempertimbangkan teks *K neighbor* dari jarak terdekat (paling mirip) dalam pelatihan data (Romadhon & Kurniawan, 2021).

Setelah dilakukan simulasi terhadap penerapan ketiga metode, selanjutnya akan dilakukan teknik *future selection*. *Feature Selection* merupakan teknik pemilihan fitur yang relevan yang dapat digunakan untuk mendapatkan bagian fitur yang paling baik. Keunggulan *future selection* diantaranya ukuran dataset yang lebih kecil dan kebutuhan komputasi yang lebih rendah. Tujuan dari penggunaan *feature selection* ialah memilih fitur yang terbaik dari kumpulan data fitur sehingga diharapkan dapat menghasilkan nilai akurasi yang lebih optimal.

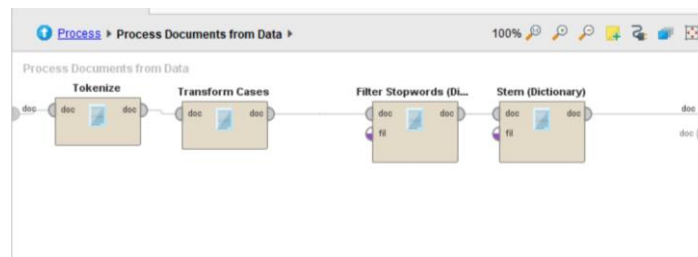
HASIL DAN PEMBAHASAN

Pada hasil dan pembahasan ini terdapat beberapa hasil dari banyak tahapan yang telah dilakukan seperti halnya dari pengambilan data sampai pengujian data dan pembahasan dari hasil ujicoba data. Pada dataset diambil variabel atau atribut yang diujikan yaitu Teks atau komentar kemudian setelah data didapatkan dilakukan pemberian label pada setiap komentar. Penandaan label pada komentar ini menunjukkan bahwa komentar tersebut positif atau negatif. Dari data yang didapatkan 1177 data yang kebanyakan dari data tersebut mempunyai jumlah label negatif lebih banyak dari pada komentar dengan label positif. Berikut contoh sebagian data yang digunakan untuk penelitian.

Tabel 1. Dataset Komentar Peralihan Sinyal TV

No.	Indikator	Nilai
1	Tinggal di tambun bekasi kenapa tidak dapat metrotv dan trans ya tv digitalnya? Padahal tv sendiri sudah dvb-t2 dan pakai set top box juga	Negatif
2	Tapi memng jadi lebih bagus siarannya, tidak ada semut-semutnya	Positif

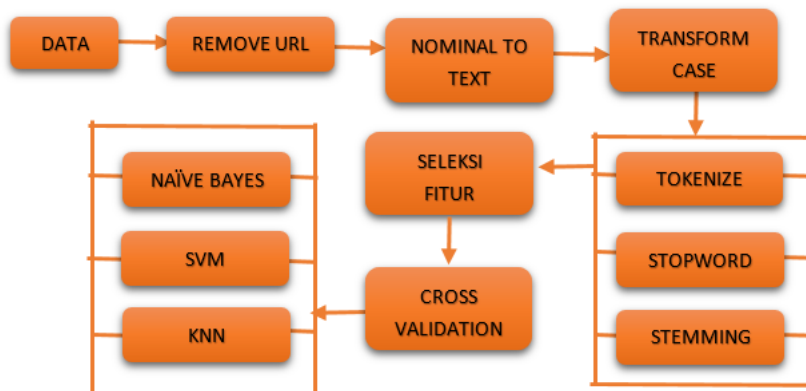
Data yang terdapat pada tabel 1 adalah hasil dari pembersihan data diantaranya menghilangkan komentar yang sama. Kemudian masih diperlukan tahap persiapan data yang terdapat langkah-langkah mempersiapkan data yaitu dengan melakukan *remove url*, *transform case*, *tokenize*, *stopword* dan *stemming*.



Gambar 2. Process Document From Data

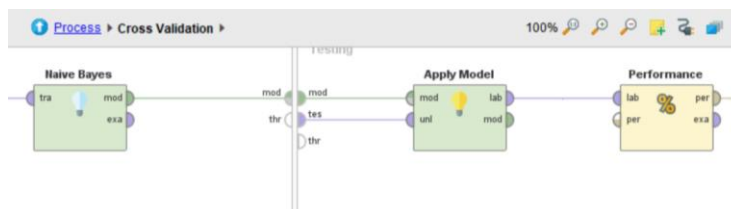
Proses yang pertama yaitu, data yang ada terkadang dalam komentar terdapat link atau URL sehingga pada proses yang pertama menghilangkan link atau URL pada setiap komentar. Sehingga komentar yang ada sudah bersih dari link atau URL. Proses selanjutnya adalah *Nominal To Text* adalah mengubah karakter angka menjadi huruf dalam setiap komentar. Kemudian untuk proses selanjutnya terdapat *Transform Case* yaitu, data yang sudah di ubah menjadi huruf semua dan tidak terdapat URL akan diubah menjadi huruf kecil semua. Dalam proses ini dilakukan seperti halnya bagian dari *text mining* yang secara luas dapat diartikan sebagai proses menggali informasi dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen dari *data mining*. Dimana proses tersebut dilanjutkan dengan tahapan *tokenizing* yaitu proses memecah dokumen menjadi kumpulan kata. Dalam proses ini dilakukan dengan menghilangkan tanda baca dan memisahkan kata tersebut dengan spasi. Kemudian untuk proses *stopword* dilakukan tahapan menghilangkan kata yang tidak penting, karena data komentar berbahasa Indonesia maka digunakan *stopword* dengan bahasa Indonesia hal ini dapat diketahui dari *stoplist* yang merupakan sebuah hasil parsing deskripsi yang didalamnya terdapat kata-kata tidak penting. Kemudian proses selanjutnya adalah *stemming*, proses ini merupakan bagian yang tidak dapat dipisahkan dari *Information Retrieval* yang tidak banyak algoritma dikhususkan untuk *stemming* bahasa Indonesia dengan berbagai keterbatasan didalamnya.

Setelah melalui tahapan mempersiapkan data selanjutnya data dapat diproses atau diolah untuk mendapatkan hasil. Dalam proses tahapan mengolah data dapat dilihat pada gambar 3 dibawah ini dengan urutan sesuai dengan proses yang dilakukan.



Gambar 3. Alur Pengujian Data

Dalam pengujian data pertama dengan menggunakan algoritma Naïve Bayes tanpa menggunakan *feature selection*. Kemudian yang kedua *feature selection* menggunakan N-Gram. Proses tersebut dilakukan sebelum melalui proses *cross validation*. Sehingga dibawah ini merupakan proses *cross validation* dimana didalamnya disesuaikan dengan algoritma yang diujikan.



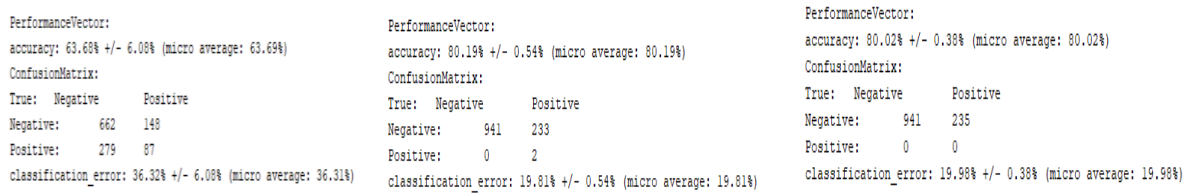
Gambar 4. Cross Validation

Setelah melalui uji coba *cross validation* pada proses gambar 4 dan dengan diujikan pada masing-masing algoritma ketiga metode yaitu *Naïve Bayes*, *Support Vector Machine*, dan *K-Nearest Neighbor*. Sehingga menghasilkan akurasi seperti yang terdapat pada gambar dibawah ini.

<p>PerformanceVector: accuracy: 61.22% +/- 4.88% (micro average: 61.22%) ConfusionMatrix: True: Negative Positive Negative: 611 126 Positive: 330 109 classification_error: 38.78% +/- 4.88% (micro average: 38.78%)</p>	<p>PerformanceVector: accuracy: 80.10% +/- 0.58% (micro average: 80.10%) ConfusionMatrix: True: Negative Positive Negative: 940 233 Positive: 1 2 classification_error: 19.90% +/- 0.58% (micro average: 19.90%)</p>	<p>PerformanceVector: accuracy: 79.93% +/- 0.53% (micro average: 79.93%) ConfusionMatrix: True: Negative Positive Negative: 940 235 Positive: 1 0 classification_error: 20.07% +/- 0.53% (micro average: 20.07%)</p>
---	---	---

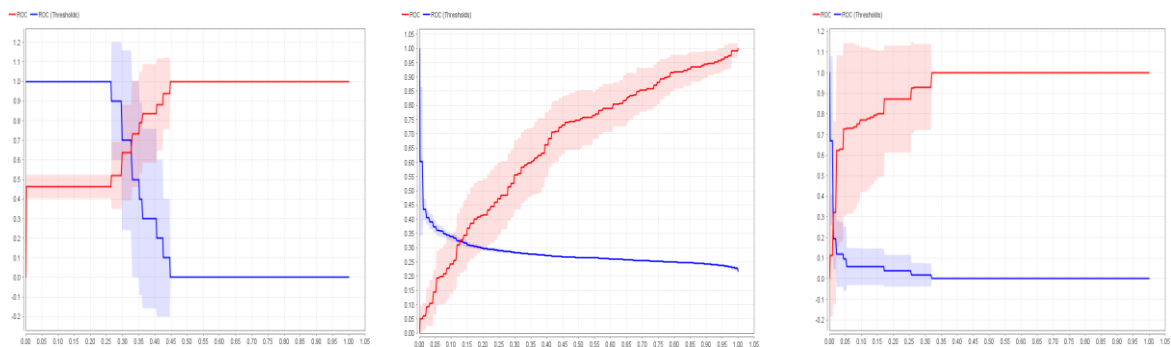
Gambar 5. Hasil akurasi dari metode *Naïve Bayes*, *Support Vector Machine*, dan *K-Nearest Neighbor* tanpa *Feature Selection*

Pada gambar 5 adalah hasil akurasi tanpa menggunakan *Feature Selection* dari setiap metode, dimana gambar 5 sebelah kiri adalah hasil akurasi *Naïve Bayes* dengan hasil 61,22 % dengan *classification error* sebanyak 38,78 %. Kemudian pada bagian tengah adalah hasil dari metode *Support Vector Machine* dengan hasil 80,10% dengan *classification error* sebesar 19,90%. Sedangkan pada metode *K-Nearest Neighbor* memperoleh akurasi 79,93% dengan *classification error* sebesar 20,07%.



Gambar 6. Hasil akurasi dari *Naïve Bayes*, *Support Vector Machine*, dan *K-Nearest Neighbor* dengan *Feature Selection*

Setelah melakukan uji tanpa menggunakan *Feature Selection*, selanjutnya digunakan uji coba setiap metode dengan menggunakan *Feature Selection* yang hasilnya terdapat pada gambar 6 pada metode *Naïve Bayes* menghasilkan akurasi sebesar 63,68 % dengan *classification error* 36,32%, kemudian untuk metode *Support Vector Machine* terdapat pada bagian tengah dengan hasil akurasi 80,19% dan *classification error* 19,81%. Sedangkan metode *K-Nearest Neighbor* pada gambar 6 bagian kanan mendapatkan hasil akurasi sebesar 80,02% dengan *classification error* 19,98%.



Gambar 7. Kurva ROC dari *Naïve Bayes*, *Support Vector Machine*, dan *K-Nearest Neighbor*

Dalam melakukan uji coba dengan ketiga metode terdapat hasil AUC (*Optimistic*) dinyatakan dengan hasil kurva ROC yang terdapat pada gambar 7 menghasilkan *micro average* pada metode *Naïve Bayes* 0,812 tidak terdapat perubahan baik sebelum dan sesudah menggunakan *future selection*, kemudian untuk hasil dari metode *Support Vector Machine* tanpa menggunakan *future selection* 0,666 setelah menggunakan *future selection* menjadi 0,678. Sedangkan dengan metode *K-Nearest Neighbor* mengalami kenaikan sebelumnya 0,930 menjadi 0,933 dengan menggunakan *future selection*. Dari hasil percobaan diatas didapatkan hasil perbandingan sebagai berikut.

Tabel 2. Hasil Pengujian Data dengan Ketiga Metode

Algoritma	Akurasi	Akurasi (<i>Future Selection</i>)	AUC	AUC + FS
NB	61,22 %	63,68 %	0,812	0,812
SVM	80,10 %	80,19 %	0,666	0,678
KNN	79,93 %	80,02 %	0,930	0,933

Dari tabel 2 dapat dideskripsikan bahwa presentase hasil pengolahan data didapatkan dengan akurasi *Naïve Bayes* 61,22%, *SVM* 80,10 % dan *KNN* sebesar 79,93%,. Kemudian jika ditambahkan dengan *feature selection* mendapatkan hasil *Naïve Bayes* 63,68%, *SVM* 80,19% dan *KNN* 80,02%. Pada metode *Naïve Bayes* dalam penelitian ini memiliki kekurangan dalam masalah kemungkinan (*Probabilitas*) terutama dalam menemukan kata-kata dalam data

pengujian tertentu. Namun, percobaan dengan Naïve Bayes memiliki kecepatan lebih tinggi pada saat diaplikasikan. Kemudian pada metode *Support Vector Machine* memiliki akurasi tertinggi dibandingkan dengan Naïve Bayes dan KNN. Sehingga dalam penelitian ini *Support Vector Machine* menghasilkan model klasifikasi yang baik, namun tidak meningkatkan akurasi secara signifikan dengan penambahan *feature selection* yaitu 0,09%. Demikian pula dengan KNN menghasilkan penambahan nilai akurasi dengan *feature selection* dengan nilai yang sama dengan *Support Vector Machine*. Namun, dalam penelitian ini KNN cukup mudah diimplementasikan hanya dengan mengatur satu parameter k.

KESIMPULAN DAN SARAN

Hasil akurasi yang didapatkan dalam pengujian data komentar media sosial mengenai peralihan TV analog menjadi TV digital mengalami kenaikan terhadap adanya *feature selection* terhadap masing-masing metode yaitu, *Support Vector Machine*, *K-Nearest Neighbor* dan *Naïve Bayes*. Dari pengujian tersebut dapat dijelaskan bahwa algoritma *Support Vector Machine* merupakan yang terbaik jika dibandingkan dengan kedua metode *K-Nearest Neighbor* dan *Naïve Bayes*. Hal ini dapat diketahui dari hasil uji coba sebelum menggunakan *Feature Selection* metode SVM mencapai akurasi tertinggi dengan hasil 80,10%. Kemudian setelah menggunakan *Feature Selection* hasil akurasi bertambah meskipun tidak terlalu banyak, dengan penambahan 0,09% sehingga hasil akurasi optimal mencapai 80,19%. Namun jika dilihat dari perbandingan sebelum menggunakan *Feature Selection* dan sesudahnya metode Naïve Bayes yang memiliki kenaikan yang paling banyak yaitu 2,46%. Pada ketiga metode tersebut mengalami kenaikan dengan *Feature Selection*.

Penelitian ini diharapkan dapat dikembangkan kembali untuk penelitian selanjutnya dengan mencoba menggunakan *feature selection* yang lain seperti *Chi Square* dan *Weight By Correlation* supaya dapat mengetahui hasil yang dapat dipertimbangkan.

DAFTAR PUSTAKA

- Agussetianingsih, B., & Kasim, A. (2021). Peran Desain Kebijakan: Digitalisasi Penyiaran Televisi di Indonesia. *Kolaborasi: Jurnal Administrasi Publik*, 7(2), 167-186.
- Gultom, A. D. (2018). Digitalisasi Penyiaran Televisi di Indonesia. *Buletin Pos Dan Telekomunikasi*, 16(2), 91-100.
- Herianto, H. (2018). Penerapan Text-Mining Untuk Mengidentifikasi Pengguna Twitter Terhadap Fenomena Peran DPR RI. *Jurnal Sains & Teknologi Fakultas Teknik Universitas Darma Persada*, 8(2), 36-44.
- Obilor, E. I., & Amadi, E. C. (2018). Test for significance of Pearson's correlation coefficient. *International Journal of Innovative Mathematics, Statistics & Energy Policies*, 6(1), 11-23.
- Pattiiha, F. S., & Hendry, H. (2022). Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia. *JURIKOM (Jurnal Riset Komputer)*, 9(2), 506-514. <http://dx.doi.org/10.30865/jurikom.v9i2.4016>
- Permana, R. S. M., Abdullah, A., & Mahameruaji, J. N. (2019). Budaya menonton televisi di Indonesia: dari terrestrial hingga digital. *ProTVF: Jurnal Kajian Televisi dan Film*, 3(1), 53-67. <https://doi.org/10.24198/ptvf.v3i1.21220>

- Rasidin, M., Witro, D., Yanti, B. Z., Purwaningsih, R. F., & Nurasih, W. (2020). The Role of Government In Preventing The Spread Of Hoax Related The 2019 Elections In Social Media. *Diakom: Jurnal Media Dan Komunikasi*, 3(2), 127-137. <https://doi.org/10.17933/diakom.v3i2.76>
- Rasidin, M., Witro, D., Yanti, B. Z., Purwaningsih, R. F., & Nurasih, W. (2020b). The Role of Government in Preventing The Spread Of Hoax Related The 2019 Elections In Social Media. *Diakom: Jurnal Media Dan Komunikasi*, 3(2), 127-137. <https://doi.org/10.17933/diakom.v3i2.76>
- Romadhon, M. R., & Kurniawan, F. (2021). A comparison of naive Bayes methods, logistic regression and KNN for predicting healing of Covid-19 patients in Indonesia. In *2021 3rd east Indonesia conference on computer and information technology (eiconcit)*, 41-44.
- Romadloni, N. T., & Pardede, H. F. (2019). Seleksi Fitur Berbasis Pearson Correlation Untuk Optimasi Opinion Mining Review Pelanggan. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(3), 505-510. <https://doi.org/10.29207/resti.v3i3.1189>
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd. Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 10(2), 681-686.
- Suryadi, D., Imansyah, F., & Marpaung, J. (2022). Identifikasi Penerimaan Sinyal Antena Digital Untuk Televisi Menggunakan Metode Sinpo. *Jurnal Teknik Elektro Universitas Tanjungpura*, 1(1), 1-9.
- Zuliana, V., Garno, G., & Maulana, I. (2022). Analisis Sentimen Program Migrasi TV Digital Menggunakan Algoritma Naive Bayes dengan Chi Square. *Jurnal Informasi dan Komputer*, 10(2), 90-95.

How to cite:

Romadloni, N. T., & Septiyanti, N. D. (2023). Optimasi *Feature Selection* Pada Komentar Media Sosial Terhadap Peralihan TV Digital Menggunakan Naïve Bayes, Support Vector Machine dan K-Nearest Neighbor. *DECODE: Jurnal Pendidikan Teknologi Informasi*, 3(2), 151-160. <http://dx.doi.org/10.51454/decode.v3i2.121>