

ANALISIS SENTIMEN PROGRAM MIGRASI TV DIGITAL MENGGUNAKAN ALGORITMA NAIVE BAYES DENGAN CHI SQUARE

Virgaria Zuliana¹, Garno², Iqbal Maulana³
Universitas Singaperbangsa Karawang¹²³

Jl. HS. Ronggo Waluyo, Puseurjaya, Kec. Telukjambe Timur, Kab. Karawang¹²³
Email : virgaria.zuliana18128@student.unsika.ac.id¹, garn0@staff.unsika.ac.id²,
iqbal.maulana@staff.unsika.ac.id³

ABSTRAK

Saat ini televisi menempati posisi nomor 2 sebagai sumber informasi setelah media sosial. Sistem siaran TV analog akan diganti dengan TV digital berdasarkan rencana yang dikeluarkan oleh Kementerian Komunikasi dan Informatika di Indonesia. Media sosial berguna untuk berbagi pemikiran dan opini tentang peristiwa, produk dan lainnya, contohnya terhadap migrasi TV digital yang sedang berlangsung. Keunggulan TV digital termasuk teknologi yang unggul dan kejernihan gambar yang jernih dan tajam. Beberapa orang berpendapat bahwa mereka puas dengan transisi ke TV digital, sementara yang lain sebaliknya. Sehingga peneliti menjadi tertarik dengan dua opini tersebut dan tertarik untuk melakukan analisis sentimen masyarakat yang mengenai program migrasi siaran TV digital di media sosial Twitter karena adanya dua tanggapan tersebut. Penelitian menggunakan algoritma *Naive Bayes* dan melakukan seleksi fitur dengan *Chi Square* pada proses penelitian untuk mengkaji perbedaan opini publik tentang migrasi ke siaran TV digital. Hasil klasifikasi dengan 191 data sentimen positif dan 185 data sentimen negatif menghasilkan tingkat akurasi 96%, *precision* 93% dan *recall* 100%.

Kata kunci : *text mining*, siaran tv digital, *naive bayes*, *chi square*

ABSTRACTS

Currently, television occupies the number 2 position as a source of information after social media. The analog TV broadcast system will be replaced with digital TV based on a plan issued by the Ministry of Communication and Information Indonesia. Social media is useful for sharing thoughts and opinions about events, products and more, for example on the ongoing digital TV migration. The advantages of digital TV include superior technology and clear, crisp picture clarity. Some people argue that they are satisfied with the transition to digital TV, while others are the opposite. So that researchers are interested in these two opinions and are interested in analyzing public sentiment regarding the migration program for digital TV broadcasts on Twitter social media because of these two responses. The research uses the *Naive Bayes* algorithm and performs feature selection with *Chi Square* in the research process to examine differences in public opinion about migration to digital TV broadcasts.. The results of the classification with 191 positive sentiment data and 185 negative sentiment data resulted in 96% accuracy, 93% precision and 100% recall.

Keywords: *text mining*, digital tv broadcast, *naive bayes*, *chi square*

1. PENDAHULUAN

Televisi merupakan media informasi umum yang berguna karena disampaikan secara efektif dan cepat. Katadata Insight Center (KIC) dan Kementerian Komunikasi dan Informatika

(Kominfo) melakukan survei mengenai “Ragam Sumber Informasi yang Paling Banyak Diakses Masyarakat” hasil dari survei menunjukkan bahwa, setelah media sosial, televisi berada pada posisi 2 sebagai media sumber informasi di

masyarakat Indonesia dengan minat tertinggi (1).

Sistem siaran televisi yang diterapkan di Indonesia saat ini menggunakan sistem siaran TV analog dan sistem siaran TV digital. Namun dengan perkembangan teknologi, penggunaan TV analog di Indonesia dinilai tertinggal (2). Maka Kementerian Komunikasi dan Informatika (Kemkominfo) telah merilis program migrasi TV analog ke TV digital serta penghentian sistem TV analog.

TV digital memberikan kualitas gambar yang lebih baik daripada TV analog, dengan suara yang lebih jernih dan teknologi yang lebih canggih. Ada beberapa masyarakat yang mengungkapkan tanggapan dan opini positif di media sosial, salah satunya pada Twitter yang mendukung dan puas setelah melakukan migrasi ke TV digital. Selain itu ada pula masyarakat yang memberikan tanggapan dan opini negatif yang mengeluhkan di media sosial Twitter terkait hilangnya siaran dan sedikitnya siaran pada TV digital. Twitter adalah media sosial tempat membagikan ide, gagasan, ekspresi dan memberikan pendapat berbentuk teks singkat, foto atau video yang disebut *tweet* terhadap suatu produk, layanan, isu politik, ataupun hal-hal viral dan lainnya (3).

Adanya opini positif dan negatif yang diberikan oleh masyarakat di media sosial Twitter membuat peneliti tertarik untuk melakukan analisis sentimen mengenai program migrasi TV analog ke TV digital untuk dilakukan klasifikasi. Analisis sentimen menurut Berry dalam Rahutomo et al., (4) yaitu proses yang dilakukan secara otomatis dalam mempelajari, mengambil, dan mengolah data tekstual untuk memperoleh informasi dan melihat tanggapan atau opini mengenai sebuah masalah atau objek menuju opini positif atau negatif.

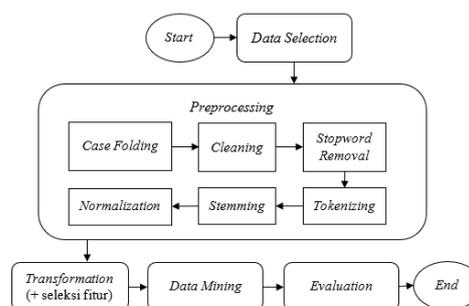
Pattiiha & Hendry, dalam penelitian perbandingan metode *KNN*, *Naive Bayes*, dan *Decision Tree* untuk melakukan analisis sentimen dari data *tweet* terkait opini pada PT PAL Indonesia, menghasilkan dalam menggunakan metode *Naive Bayes* memiliki tingkat akurasi tertinggi sebesar 84,08%, sedangkan untuk metode *K-NN* adalah 83,38% dan *Decision Tree* adalah 81,09 (5). Pada

penelitian lainnya menerapkan algoritma *Naive Bayes* dan penggunaan seleksi atribut *Chi Square* dalam mengklasifikasi loyalitas terhadap pelanggan GRAB menghasilkan akurasi dengan hanya menggunakan *Naive Bayes* sebesar 92,73% sedangkan ketika prosesnya menggunakan *Naive Bayes* dengan ditambah fitur seleksi *Chi Square* hasil akurasinya menjadi 99,51% (6). Dari hasil penelitian sebelumnya algoritma *Naive Bayes* lebih unggul dibandingkan *Decision Tree* dan *KNN*, serta penggunaan seleksi fitur *Chi Square* pada algoritma *Naive Bayes* menghasilkan akurasi yang meningkat..

Penelitian ini memiliki tujuan untuk melakukan analisis sentimen untuk mengklasifikasikan dengan menggunakan algoritma *Naive Bayes* dengan *Chi Square* sebagai seleksi fiturnya. Algoritma *Naive Bayes* menurut Hozairi et al., (7) yaitu algoritma sudah teruji memiliki nilai akurasi dan kecepatan tinggi jika menggunakannya ke *database* yang memiliki informasi besar. Seleksi fitur dengan menggunakan metode *Chi Square* diimplementasikan dengan *Select Kbest* pada *library sklearn* yang bekerja dengan menghitung nilai pada masing-masing fitur lalu menerapkan fungsi *Chi Square* dan kemudian dilakukan pengurutan nilai fitur teratas berdasarkan pada nilai *k* yang ditentukan (8).

2. METODE PENELITIAN

Beberapa tahapan dalam penelitian ini dilakukan untuk melakukan klasifikasi terhadap program migrasi siaran TV digital, beberapa tahapan tersebut dapat dilihat pada Gambar 1 yang merupakan alur penelitian.



Gambar 1. Alur Penelitian

Penelitian dimulai dengan melakukan proses *data selection* kemudian dilakukan tahap *preprocessing*, tahap *transformation*, dan *data mining* serta tahap terakhir yaitu *evaluation*. Pada tahap *preprocessing* juga meliputi beberapa tahap antara lain *cleaning*, *case folding*, *tokenizing*, *stopword removal*, *stemming* dan *normalization*. Berikut penjelasan mengenai tahapan pada penelitian ini (9) (10):

2.1 Data Selection

Penelitian ini melakukan *data selection* dimulai dengan mengumpulkan data dengan menggunakan teknik tertentu seperti teknik *crawling* dan dengan kata kunci tertentu, setelah data didapatkan maka dilanjutkan dengan melakukan seleksi dan pelabelan data.

2.2 Preprocessing

Tahap *preprocessing* melakukan proses pembersihan data dan membentuk data yang telah dikumpulkan agar data memiliki struktur yang efektif untuk digunakan pada tahap selanjutnya. Tahap *preprocessing* meliputi beberapa tahapan lain yaitu:

1. Cleaning

Membersihkan data untuk menghapus angka, *emoticon*, *hashtag*, *url*, *username*, ataupun simbol karena tidak memiliki keperluan.

2. Tokenizing

Mengubah teks kalimat dalam data menjadi kata per kata untuk proses yang lebih mudah pada proses transformasi.

3. Stopword Removal

Membersihkan data untuk menghapus kata yang tidak memiliki pengaruh sehingga pada proses klasifikasi menjadi optimal.

4. Stemming

Proses menghapus kata imbuhan yang ada pada awla dan akhir kata pada data yan digunakan.

5. Case Folding

Melakukan perubahan data teks menjadi teks dengan huruf kecil.

6. Normalization

Proses memperbaiki kata pada data yang memiliki kesalahan dalam penulisan.

2.3 Transformation

Proses mengubah struktur pada data awal menjadi yang sesuai dengan kriteria metode data mining yang digunakan pada tahap selanjutnya. Seperti mengubah data teks menjadi data numerik dengan pembobotan menggunakan TF-IDF agar dapat melakukan proses selanjutnya dengan algoritma lain. Setelah itu pada tahap ini dapat dilakukan proses pemilihan data seperti seleksi fitur, salah satunya dengan menggunakan *Chi Square*.

2.4 Data Mining

Tahap *data mining* melakukan proses melakukan perubahan data mentah sehingga jadi informasi yang berguna. Pada penelitian ini dilakukan klasifikasi dengan menggunakan algoritma tertentu dengan kalimat sentimen dikelompokkan pada kelas positif dan negatif. Salah satu algoritma yang dapat digunakan dalam proses klasifikasi yaitu *Naive Bayes*.

2.5 Evaluation

Melakukan evaluasi pada performa dari algoritma yang digunakan, seperti dengan melakukan implementasi menggunakan *Confusion Matrix* maka hasil evaluasi akan berupa nilai akurasi, presisi dan *recall*. Berdasarkan nilai tersebut maka akan mudah dipahami pengetahuan atau informasi yang sebelumnya telah dilakukan.

3. HASIL DAN PEMBAHASAN

3.1. Data Selection

Penelitian analisis sentimen mengenai opini publik terhadap migrasi siaran TV digital menggunakan data *tweet* dari pengguna twitter. Kata kunci yang digunakan yaitu “siaran tv digital”. Data dikumpulkan dalam kurun waktu 1 Februari 2022 sampai 31 Juli 2022.

Menghasilkan data asli sebanyak 3543 data *tweet* dan dilakukan seleksi guna menghapus data duplikat sehingga data menjadi 1815. Data kemudian dilakukan seleksi dengan bantuan pakar bahasa indonesia sehingga menghasilkan 376 data dengan 191 data sentimen positif dan 185 data sentimen negatif.

	username	tweet	value
0	zonalayakhuni @Moonxx	@UNSFess_sama-samaa aku ga perna...	positive
1	SuryadiGatot @UpdateBoiabola Nah	Untung Ada TVRI yg Udh Dig...	negative
2	Thopix_village0 @detikcom	Ujung ujung nya memberatkan warga ku...	negative
3	musttheguh @detikcom	Bisa nonton TV tanpa menggunakan STB...	negative
4	ecceuuuuuuuuu @hyang_wisnu @kemkominfo @jokowi @PlateJohnny	...	positive
...
371	arinuzulita	Tv digital di jamin gambar bersih tak pakai in...	positive
372	alesxenyxone @TVRI	Nasional Katanya siaran tv digital gratis...	negative
373	WardiSamme218 @PS_AmandaJKT48	Wardi juga nonton #MotoGP #Fie...	positive
374	H412RY	Katanya siaran tv digital lebih baik tapi moto...	negative
375	rani_yurama @hyang_wisnu @kemkominfo @PlateJohnny @jokowi	...	positive

376 rows x 3 columns

Gambar 2. Data Setelah Proses *Data Selection*

3.2. Preprocessing

Tahapan *preprocessing* melewati beberapa tahap dalam membentuk data sehingga siap digunakan pada tahap *data mining*.

1. Cleaning

Pada proses ini melakukan proses menghapus tag *username*, angka, *hashtag*, *url* dan tanda tanya pada atribut *tweet*. Hasil dari proses *cleaning* disimpan pada kolom baru dengan nama *remove_http*. Pada Gambar 3 berikut menunjukkan hasil pada proses *cleaning*.

	tweet	value	remove_user	remove_http
122	@hyang_wisnu @PlateJohnny @jokowi @Ayrastrisyo	positive	Dengan siaran format digital sia...	Dengan siaran format digital sia...
219	@hyang_wisnu @PlateJohnny @jokowi @Ayrastrisyo	positive	Mantap migrasi tv digital makin m...	Mantap migrasi tv digital makin m...
227	@hyang_wisnu @kemkominfo @PlateJohnny	positive	Selamat menikmati siaran tv digita...	Selamat menikmati siaran tv digita...
232	@hyang_wisnu @kemkominfo @jokowi @PlateJohnny	positive	Tv digital memangab ketesak merata...	Tv digital memangab ketesak merata...
233	@AdeBayono @PlateJohnny @kemkominfo @jokowi @...	positive	Omnibus Law sangat mendukung progr...	Omnibus Law sangat mendukung progr...
...
266	Wowmahal ya yg sorfornikat... in ga ada n...	negative	Wowmahal ya yg sorfornikat... in ga ada n...	Wowmahal ya yg sorfornikat... in ga ada n...
277	Yang sudah menggunakan televisi dengan Sistem...	positive	Yang sudah menggunakan televisi dengan Sistem...	Yang sudah menggunakan televisi dengan Sistem...
281	digital lebih menguntungkan masyarakat karena...	positive	digital lebih menguntungkan masyarakat karena...	digital lebih menguntungkan masyarakat karena...
110	lebih stabil dan tahan terhadap gangguan. Ayo...	positive	lebih stabil dan tahan terhadap gangguan. Ayo...	lebih stabil dan tahan terhadap gangguan. Ayo...
34	mi stb evercross tidak bisa terima siaran digi...	negative	mi stb evercross tidak bisa terima siaran digi...	mi stb evercross tidak bisa terima siaran digi...

376 rows x 4 columns

Gambar 3 Hasil Proses *Cleaning*

2. Case Folding, Tokenizing, Stopword Removal, Stemming

Tahap *case folding*, *tokenizing*, *stopword removal* dan *stemming* dilakukan bersamaan dengan mengubah semua kasta menjadi huruf kecil, memisahkan kata perkata, menghapus kata yang tidak diperlukan, serta menghilangkan imbuhan pada kata. Proses ini dilakukan dengan bantuan *library nltk*, menggunakan *stopwords*, *tokenize*, dan

stemming menggunakan *library* Sastrawi. Berikut pada Gambar 4 merupakan hasil dari prosesnya.

	remove_user	remove_http	tweet_clean
	Dengan siaran format digital sia...	Dengan siaran format digital sia...	{sai, format, digital, sai, tv, agam, yg, ut...
	Mantap migrasi tv digital makin m...	Mantap migrasi tv digital makin m...	{mantap, migrasi, tv, digital, agam, sai, bu...
	Selamat menikmati siaran tv digita...	Selamat menikmati siaran tv digita...	{selamat, nikmat, sai, tv, digital, jernih}
	Tv digital memangab ketesak merata...	Tv digital memangab ketesak merata...	{tv, digital, toak, rata, sai, tv, smogg}
	Omnibus Law sangat mendukung progr...	Omnibus Law sangat mendukung progr...	{omnibus, law, dukung, program, transkomsa}

	Wowmahal ya yg sorfornikat... in ga ada n...	Wowmahal ya yg sorfornikat... in ga ada n...	{wowmahal, ya, yg, sorfornikat, ga, nait, gta}
	Yang sudah menggunakan televisi dengan Sistem...	Yang sudah menggunakan televisi dengan Sistem...	{televisi, sistem, bas, digital, report, atai, ...}
	digital lebih menguntungkan masyarakat karena...	digital lebih menguntungkan masyarakat karena...	{digital, untung, masyarakat, sai, tv, lektur...}
	lebih stabil dan tahan terhadap gangguan. Ayo...	lebih stabil dan tahan terhadap gangguan. Ayo...	{stabil, tahan, ganggu, ayo, aih, sai, tv, it}
	mi stb evercross tidak bisa terima siaran digi...	mi stb evercross tidak bisa terima siaran digi...	{mi, stb, evercross, tidak, bisa, terima, siaran, digital, ko...}

Gambar 4. Hasil *Case Folding, Tokenizing, Stopword Removal, & Stemming*

3. Normalization

Tahap selanjutnya dilakukan normalisasi guna memperbaiki kata salah ketik atau singkatan dengan menggunakan kamus normalisasi. Pada Tabel 1 berikut merupakan contoh pada kamus normalisasi.

Tabel 1. Contoh Kamus Normalisasi

Before	After
blank	kosong
block	terhalang
bnyak	banyak
kesel	kesal
ntn	nonton
trjangkau	terjangkau
trpaksa	terpaksa

Hasil pada proses tersebut terdapat dikolom baru dengan nama *normalized*, seperti yang tertera pada Gambar 4 berikut.

	remove_http	tweet_clean	normalized
	Dengan siaran format digital sia...	{sai, format, digital, sai, tv, agam, yg, ut...	{sai, format, digital, sai, tv, agam, yang, ut...
	Mantap migrasi tv digital makin m...	{mantap, migrasi, tv, digital, agam, sai, bu...	{mantap, migrasi, tv, digital, agam, sai, bu...
	Selamat menikmati siaran tv digita...	{selamat, nikmat, sai, tv, digital, jernih}	{selamat, nikmat, sai, tv, digital, jernih}
	Tv digital memangab ketesak merata...	{tv, digital, toak, rata, sai, tv, smogg}	{tv, digital, toak, rata, sai, tv, smogg}
	Omnibus Law sangat mendukung progr...	{omnibus, law, dukung, program, transkomsa}	{omnibus, hukum, dukung, program, transkomsa}

	Wowmahal ya yg sorfornikat... in ga ada n...	{wowmahal, ya, yg, sorfornikat, ga, nait, gta}	{wow mahal, ya, yang, sorfornikat, ga, nait, gta}
	Yang sudah menggunakan televisi dengan Sistem...	{televisi, sistem, bas, digital, report, atai, ...}	{televisi, sistem, basis, digital, report, atai, ...}
	digital lebih menguntungkan masyarakat karena...	{digital, untung, masyarakat, sai, tv, lektur...}	{digital, untung, masyarakat, sai, tv, lektur...}
	lebih stabil dan tahan terhadap gangguan. Ayo...	{stabil, tahan, ganggu, ayo, aih, sai, tv, it}	{stabil, tahan, ganggu, ayo, aih, sai, tv, it}
	mi stb evercross tidak bisa terima siaran digi...	{mi, stb, evercross, tidak, bisa, terima, siaran, digital, ko...}	{mi, stb, evercross, tidak, bisa, terima, siaran, digital, ko...}

Gambar 5. Hasil *Normalization*

4. KESIMPULAN

Hasil pada pengujian yang telah melakukan kalsifikasi sentimen menggunakan algoritma *Naïve Bayes* dengan *Chi Square* menghasilkan nilai akurasi sebesar 96%, *precision* 93% dan *recall* 100% dengan 12 data kelas negatif diprediksi dengan benar dan 1 data kelas negatif diprediksi sebagai kelas positif. Sedangkan kelas positif telah berhasil diprediksi dengan jumlah 14 data dan tidak ada kelas positif yang diprediksi negatif.

DAFTAR PUSTAKA

- [1]. Annur CM. Survei KIC: Mayoritas Masyarakat Indonesia Mengakses Informasi di Media Sosial [Internet]. databoks.katadata.co.id. 2022 [cited 2022 May 25]. Available from: <https://databoks.katadata.co.id/datapublish/2022/01/20/survei-kic-mayoritas-masyarakat-indonesia-mengakses-informasi-di-media-sosial>
- [2]. Andarini FA. Analisis Strategi Digitalisasi Media di Era Digital PT. Media Nusantara Citra Tbk (MNCN). J Ilm Indones. 2022 Feb;7(2).
- [3]. Mahardhika YS, Zuliarso E. Analisis Sentimen Terhadap Pemerintahan Joko Widodo pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes Classifier. Sintak. 2018;2.
- [4]. Rahutomo F, Saputra PY, Fidyawan MA. Implementasi Twitter Sentiment Analysis Untuk Review Film Menggunakan Algoritma Support Vector Machine. J Inform Polinema. 2018;4(2):93–100.
- [5]. Pattiiha FS, Hendry. Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia. J Ris Komput. 2022;9(2):2407–389.
- [6]. Kisworini RY, Setiawan MA. Peningkatan Performa Naive Bayes Dengan Seleksi Atribut Menggunakan Chi Square Untuk Klasifikasi Loyalitas Pelanggan GRAB. J Informatics, Inf Syst Softw Eng Appl. 2020;2(2):69–075.
- [7]. Hozairi, Anwari, Alim S. Implementasi Orange Data Mining untuk Klasifikasi Kelulusan Mahasiswa dengan Model K-Nearest Neighbor, Decision Tree serta Naive Bayes. J Ilm NERO. 2021;6(2):133–44.
- [8]. Ratmana DO, Fajar Shidik G, Fanani AZ, Muljono, Pramunendar RA. Evaluation of Feature Selections on Movie Reviews Sentiment. 2020 Int Semin Appl Technol Inf Commun. 2020;567–71.
- [9]. Mulaab. Data Mining Konsep dan Aplikasi. Malang: Media Nusa Creative; 2017.
- [10]. Nofriansyah D. Konsep Data Mining vs Sistem Pendukung Keputusan. Yogyakarta: Deepublish; 2014.