

**ANALISIS SENTIMEN PADA CHANNEL AUTONETMAGZ
TERHADAP REVIEW MOBIL ALMAZ 2019
DENGAN METODE NAIVE BAYES CLASSIFIER DAN LEXICON BASED**

Heru Prasetyo*, Ghulam Asrofi Buntoro, Dyah Mustikasari

Prodi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Ponorogo

E-mail Kosrespondensi : heruprasetyoangan2709@gmail.com

History Artikel

Diterima: 14 Februari 2020 Disetujui: 17 Maret 2020 Dipublikasikan: 08 April 2020

Abstract

Almaz Car is one of the products of the Wuling car company. Wuling cars experienced a large increase in sales over the past 7 months in Indonesia. Wuling is able to compete in the most sales of the 9th number of 20 producers even though only 2 years Wuling products are in Indonesia. Although sales have increased, not all Indonesian people have positive or even negative comments even neutral. The public can now see videos from the most commonly used social media namely Youtube about the discussion of Wuling Almaz vehicles. In video vlogger content, users can provide a feedback through comments in the form of opinions that reinforce positive or opinions that greatly weaken negative or even neutral in the form of questions. From the various comments on the YouTube Autonnetmagz channel, techniques are needed to divide into positive, neutral or negative opinion classes. This study uses preprocessing and labeling opinions into 3 sentiment classes, namely positive, neutral and negative classes with the lexicon Based method. As for the classification using the Naive Bayes Classifier method. The data used in the form of comments about the review of Wuling almaz from Youtube Autonnetmagz channel, amounting to 1000. The results of labeling with Lexicon Based amounted to 232 for positive, neutral for 456 and negative for 312. While the classification of the Naive Bayes Classifier method resulted in an accuracy of 66.5%, 60.94% precision and 61.2% recall.

Keywords: *Sentiment Analysis, Youtube Autonnetmagz, Wuling Almaz, Lexicon Based, Naive Bayes Classifier*

Abstrak

Mobil Almaz merupakan salah satu produk dari perusahaan mobil Wuling. Mobil wuling mengalami peningkatan penjualan yang besar selama 7 bulan terakhir di Indonesia. Wuling mampu bersaing penjualan terbanyak nomor 9 dari 20 produsen meskipun baru 2 tahun produk Wuling berada di Indonesia. Meskipun penjualannya mengalami peningkatan, tidak semua masyarakat Indonesia berkomentar positif bahkan negatif bahkan netral. Masyarakat kini bisa melihat video dari media sosial yang paling umum digunakan yaitu Youtube tentang pembahasan kendaraan mobil Wuling Almaz. Dalam konten video *vlogger*, pengguna dapat memberikan sebuah *feedback* melalui komentar berupa opini yang menguatkan *positif* ataupun opini yang sangat melemahkan *negatif* atau bahkan netral berupa pertanyaan. Dari berbagai ragam komentar di channel Youtube *Autonetmagz* diperlukan teknik untuk membagi ke dalam kelas opini positif, netral maupun negatif. Penelitian ini menggunakan *preprocessing* dan melabeli opini kedalam 3 kelas sentimen yaitu kelas positif, netral dan negatif dengan metode *lexicon Based*. Sedangkan untuk klasifikasinya menggunakan metode *Naive Bayes Classifier*. Data yang digunakan berupa komentar tentang ulasan Wuling Almaz dari channel Youtube *Autonetmagz* yang berjumlah 1000. Hasil dari pelabelan dengan *Lexicon Based* berjumlah 232 untuk positif, untuk netral berjumlah 456 dan negatif berjumlah 312. Sedangkan klasifikasi metode *Naive Bayes Classifier* menghasilkan akurasi 66,5%, presisi 60,94% dan recall 61,2%.

Kata Kunci : *Analisis Sentimen, Youtube Autonetmagz, Wuling Almaz, Lexicon Based, Naive Bayes Classifier*

Prasetyo, Heru. *Analisis Sentimen pada Channel Autonetmagz Terhadap Review Mobil Almaz 2019 dengan Metode Naive Bayes Classifier dan Lexicon Based.* KOMPUTEK : Jurnal Teknik Universitas Muhammadiyah Ponorogo, 4(1), 2020: 58-70

© 2020 Universitas Muhammadiyah Ponorogo. All rights reserved

ISSN 2614-0985 (Print)

ISSN 2614-0977 (Online)

A. PENDAHULUAN

Sosial media adalah salah satu alat komunikasi teknologi yang berkembang sangat pesat, dengan sosial media kita bisa bertukar informasi melalui pesan suara, gambar atau video. Salah satu sosial media adalah youtube, youtube merupakan salah satu situs yang banyak dikunjungi oleh para pengguna internet di dunia. Youtube menyediakan berbagai macam video mulai dari video klip sampai film, serta video yang dibuat para pengguna youtube sendiri. Banyak orang juga terkenal dari situs ini dengan hanya meng-upload video mereka di youtube. Bukan hanya itu saja situs youtube juga dapat digunakan sebagai sarana untuk memasarkan produk. (Nur, Rohman, & Husna, 2015)

Youtube channel *riview* mobil banyak sekali mulai dari varian lama sampai yang terbaru. Channel ini banyak ditonton dari berbagai kalangan remaja. Contoh salah satunya adalah channel *Autonetmagz Riview* mobil yang saat ini menjadi tren. *Autonetmagz* adalah *channel* otomotif yang memiliki pengikut atau *subscriber* pada bulan Juli 2019 ini 916 ribu *subscriber* menjadikannya saluran otomotif Indonesia terbesar. Dimana dalam video tersebut akan

menjelaskan berbagai kelebihan serta kekurangan produk mobil, seperti dari segi desain mobil, kualitas, performa mesin, serta fitur-fitur apa saja yang di berikan dalam mobil tersebut. Dalam konten video *vlogger*, pengguna dapat memberikan sebuah *feedback* melalui komentar berupa opini yang menguatkan *positif* ataupun opini yang sangat melemahkan *negatif*.

Komentar opini pengguna youtube dapat di gunakan sebagai indikator untuk melihat kecenderungan pengguna terhadap Mobil Wuling Almaz, sehingga data komentar tersebut dapat menjadi sumber sentimen masyarakat serta bisa digunakan secara efisien untuk studi analisis.

Dalam perusahaan sentimen analisis digunakan sebagai pertimbangan produk di masyarakat apakah di terima atau tidak dan di manfaatkan untuk menentukan strategi meningkatkan kualitas dari produk tersebut.

Penjualan mobil pada saat ini bersaing sangat pesat salah satunya produk mobil Wuling. Produk ini sangat bersaing dalam penjualan dengan produk mobil lainnya, namun ada fakta terbaru penjualan mobil Wuling mengalami peningkatan yang besar selama 7 bulan terakhir. Wuling *Motors* Indonesia mencatat penjualan sebanyak

13.808 unit sepanjang bulan Januari – September 2019. (Tempo.com)

Berdasarkan uraian - uraian di atas Permasalahan yang di timbulkan dari channel *Autonetmagz* adalah *riview* produk mobil Wuling Almaz mempengaruhi minat beli masyarakat sebagai konsumen. Pada *vlogger* ini memiliki dampak opini, karena dengan adanya *riview* produk memberikan informasi - informasi menjelaskan kelebihan dan kekurangan teknologi yang di sematkan pada mobil tersebut. Maka dari itu, penulis membuat penelitian pengaruh *channel Autonetmagz Riview* Mobil Wuling Almaz terhadap minat beli masyarakat sebagai konsumen, analisis ini dilakukan untuk mengetahui seberapa besar opini *channel Autonetmagz meriview* produk mobil Wuling Almaz terhadap minat beli masyarakat sebagai kosumen. Penelitian ini dilakukan dengan menggunakan metode *Naive Bayes Classifier* dan *Lexicon Based*.

Tujuan yang ingin di capai dari penulisan dalam penelitian ini adalah:

- a. opini Sentimen *channel Autonetmagz Riview* produk mobil Wuling Almaz.
- b. akurasi analisis sentimen dengan metode *Naive Bayes Classifier dan Lexicon Based* pada data komentar youtube.

B. KAJIAN LITERATUR

1. AUTONETMAGZ.com (Vlogger)

Autonetmagz adalah Majalah otomotif *online*. *Channel* ini banyak dilihat dari berbagai kalangan. Contoh salah satu *vloggernya* adalah *channel Autonetmagz Riview* mobil yang saat ini sudah memiliki 916rb *subscriber*/pelanggan. Dimana dalam video tersebut akan menjelaskan berbagai kelebihan serta kekurangan produk seperti dari segi desain mobil, kualitas, performa mesin, serta fitur-fitur unik yang di sematkan dalam mobil tersebut.

2. Metode Naive Bayes Classifier

Naive Bayes classifier merupakan suatu metode klasifikasi yang menggunakan perhitungan probabilitas. Konsep dasar yang digunakan pada *Naive bayes classifier* adalah *Teorema Bayes* yang dinyatakan pertama kali oleh Thomas Bayes. (LING, N. KENCANA, & OKA, 2014)
Persamaan pada metode *naive bayes classifier* sebagai berikut :

$$p(c|X) = \frac{p(x|c)p(c)}{p(x)}$$

Keterangandari formula di atas sebagaiberikut:

$P(C|X)$: Probalitas bersarat (*conditional probability*) suatu hipotesis C terjadi jika di berikan (*envindence*) X terjadi

$P(X|C)$: Probabilitas sebuah bukti X terjadi akan mempengaruhi hipotesis C

$P(C)$: Probabilitas awal (*prior*) hipotesis C terjadi tanpa memandang bukti apapun.

$P(X)$: Probabilitas awal (*prior*) bukti X terjadi tanpa memandang hipotesis/ bukti yang lain

3. Lexicon Based

Lexicon based merupakan penentu *sentiment* kalimat sehingga kalimat dapat di klasifikasikan kedalam kelas-kelas. Dimana setiap kata pada kamus tersebut telah memiliki skor polaritas yang diberi nilai dari -1 (untuk kelas negatif) sampai dengan +1 (untuk kelas positif). Pada *library Text Blob developer* dapat menggunakan *property sentiment polarity* untuk melihat skor sentimen dari suatu kata atau kalimat. (Nafan et al., 2019)

Setelah di ketahui kata yang mengandung positif, *negative*, dan netral di dalam sebuah kalimat, selanjutnya dihitung bobot nilai yang terkandung dalam kalimat tersebut yang dilakukan dengan menjumlahkan nilai yang terkandung dalam kalimat tersebut yang di lakukan dengan menjumlahkan nilai kata opini. Jika nilai jumlah opini dalam kalimat tersebut = 0, maka nilai *sentiment* dari kalimat tersebut

adalah netral, jika nilai opini < 0 dalam kalimat tersebut adalah *negative*, dan terakhir jika nilai tersebut > 0 maka termasuk *positive*. (Buntoro, 2017)

C. PELAKSANAAN DAN METODE

1. Pengambilan Data

Pada penelitian analisis *sentiment* data diambil dari komentar youtube vlogger *autonetmagz.com*. Pengambilan data dilakukan dengan mengambil data komentar yang berjumlah 1000 dengan rentan waktu bulan Maret sampai Oktober 2019 youtube.

2. Preprocessing

Tahap selanjutnya setelah data telah didapatkan maka masuk kedalam proses *preprocessing*. *Preprocessing* data opini atau pendapat dalam penelitian ini memiliki beberapa tahap .

a. Case Folding

Proses huruf tidak konsisten dalam menggunakan huruf kapital, maka dari itu seluruh huruf yang di gunakan pada proses analisis ini disetarakan dengan mengubah huruf kapital menjadi huruf yang standar yaitu mengubah seluruh huruf menjadi huruf kecil.

b. Menghilangkan Emoticon atau Karakter
Menghilangkan emoticon pada kalimat komentar untuk mempermudah dalam proses analisis komentar.

Mengantisipasi dari kesalahan pengetikan emoticon yang tidak sesuai pada kalimat.

c. **Normalisasi Kalimat**

Proses untuk membetulkan kalimat-kalimat kata yang menggunakan tambahan, pergantian, dan perubahan huruf, sehingga tidak sesuai dengan kamus bahasa Indonesia, untuk pengejaran kalimat menjadi berubah. Contoh dari normalisasi kalimat adalah (Suara + ae, suara + nya) (Jalan 2 menjadi Jalan-jalan).

d. **Filtering**

Tahap filtering menghilangkan kata yang tidak memiliki makna seperti kata penghubung seperti “yang”, “dan”, “ke”, “dengan” dan sebagainya di dalam kalimat agar mempermudah proses penelitian. Tahap ini juga dinamakan stopword, proses filtering ini dapat membantu dalam proses penelitian.

e. **Tokenizing**

Tokenizing atau di sebut dengan parsing adalah proses dimana sebuah kalimat di pisahkan menjadi kata perkata dengan memanfaatkan sepasi sebagai pemisahannya. Pada tokenisasi kalimat di pecahkan menjadi kata perkata atau satu kata. Contohnya kalimat sebagai berikut : “ Pantas” saja sih kalo di adu

akomodasi tapi kalo sektor mesin sepertinya almaz masih ketinggal” hasil tokenizing menjadi “Pantas”, saja, sih, kalo, di, adu, akomodasi, tapi, kalo, sektor, mesin, sepertinya, almaz, masih, ketinggal,”

3. **Wordcloud**

Wordcloud merupakan gambaran visual dari kumpulan data teks. Penelitian ini menggunakan wordcloud sebagai gambaran kata dari komentar Autonetmagz yang membahas review mobil wuling almaz. Dari wordcloud sendiri, dapat melihat kata yang sering digunakan dalam pembahasan review mobil wuling almaz.

4. **Pelabelan Lexicon Based**

Dalam pemaparan sebelumnya lexicon based sendiri bisa digunakan untuk proses pelabelan data dengan cara menghitung skor sentimen. Lexicon based perlu digunakan untuk melabeli setiap komentar. Dalam penelitian (Buntoro, 2017) dipaparkan untuk menghitung setiap kata berdasarkan kamus positif dan negatif dalam kalimat tersebut, dengan menjumlahkan nilai opini. Jumlah nilai opini untuk sentimen positif bernilai 1 atau lebih, sedangkan netral dalam kalimat tersebut untuk yang bernilai = 0, begitupun sebaliknya untuk negatif dalam kalimat tersebut bernilai = -1 atau lebih.

5. Proses Klasifikasi

Pada penelitian analisis sentimen ini proses klasifikasi menggunakan metode naive bayes classifier. Penelitian ini juga menggunakan aplikasi untuk mempermudah proses klasifikasi yaitu Weka.

Cross validation adalah pengujian yang dilakukan untuk memprediksi pada *error rate*. Data *training* di bagi beberapa bagian dengan menggunakan perbandingan yang sama kemudian *error rate* dihitung setiap bagianya, Selanjutnya menghitung seluruh rata – rata *error rate* mendapatkan *error rate* secara keseluruhan.

Pada penelitian ini menggunakan metode pengujian *10 fold cross validation* yang akan mengulang pengujian 10 kali, pengukurannya nilai rata – rata dari 10 nilai pengujian, menunjukkan *10 fold cross validation* pilihan terbaik untuk mendapatkan validasi yang baik dan akurat.

Tabel 3.1. Skema 10 Fold Cross Validation

Uji ke	Data Set									
1	■	■	■	■	■	■	■	■	■	■
2	■	■	■	■	■	■	■	■	■	■
3	■	■	■	■	■	■	■	■	■	■
4	■	■	■	■	■	■	■	■	■	■
5	■	■	■	■	■	■	■	■	■	■
6	■	■	■	■	■	■	■	■	■	■
7	■	■	■	■	■	■	■	■	■	■
8	■	■	■	■	■	■	■	■	■	■
9	■	■	■	■	■	■	■	■	■	■
10	■	■	■	■	■	■	■	■	■	■

Keterangan :

Hitam = K-subset (data *testing*)

Putih = data *training*

6. Confusing matrix

Confusion matrix adalah metode biasanya digunakan untuk memperhitungkan konsep dan akurasi data mining atau sistem keputusan. pada konsep perhitungan akurasi data mining dan sistem pendukung keputusan. Pengukuran kerja menggunakan *confusion matrix*, terdapat 4 istilah representasi hasil dari proses klasifikasi. Keempat istilah tersebut adalah *True Positif* (TP), *True Negatif* (TN), *False Positif* (FP), *False Negative* (FN). Nilai *True negatif* merupakan jumlah data negatif terdeteksi dengan baik, *false positif* merupakan data negatif yang terdeteksi sebagai data positif. *True positif* merupakan data positif terdeteksi baik. *False negative* kebalikan dari true positif, hingga dapat di positif, tetapi terdeteksi sebagai data negatif. *False Negative* (FN) ialah berupa data positif namun terdeteksi sebagai data negatif. Precision merupakan data atau dokumen yang dapat berdasarkan informasi yang kurang

Namun pada penelitian ini menggunakan tambahan kelas Netral sehingga Confusion Matrik yang digunakan adalah 3x3. Penelitian ini juga memiliki istilah untuk kelas Netral diantara lain (FNT) *False Netral* dan (TNT) yaitu *True Netral* pada

penelitian (Iskandar, Suprpto, 2014) menggunakan Confusion Matrix 3x3.

Tabel 3.2. Tabel Kenyataan

Kenyataan	Prediksi		
	Positif	Netral	Negatif
Positif	True Positif (TPa)	False Netral (FNTa)	False Negatif (FNa)
Netral	False Positif (FPb)	True Netral (TNTb)	False Negatif (FNb)
Negatif	False Positif (FPc)	False Netral (FNTc)	True Negatif (TNC)

Dalam proses klasifikasi biner, presisi bias dibuat sama dengan nilai dari prediksi positif. Untuk aturan rumus precision adalah

$$\text{Precision} = \frac{TPa}{TPa+FPb+FPc} * 100\% \text{ (Iskandar, Suprpto, 2014)}$$

Recall merupakan sebuah data penghapusan yang berhasil di dapatkan dari data relevan dengan Query. Dalam proses klasifikasi biner, recall biasa dikenal sebagai sensitivitas. Perolehan data relevan yang diambil ialah menyetujui dengan Query dapat dilihat dengan recall. Untuk aturan dari recall sebagai berikut:

$$\text{Recall} = \frac{TPa}{TPa+FNTa+FNa} * 100\% \text{ (Iskandar, Suprpto, 2014)}$$

Akurasi adalah jumlah prediksi yang benar. Untuk aturan dari perhitungan akurasi sebagai berikut:

$$\text{Akurasi} = \frac{TPa+TNTb+TNC}{TPa+TNTb+TNC+FPb+FPc+FNTa+FNTc+FNa+FNb} *$$

100%

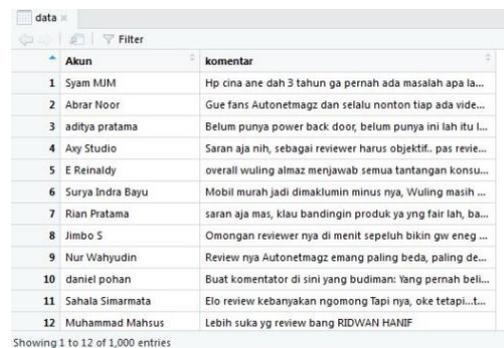
(Iskandar, Suprpto, 2014)

D. HASIL DAN PEMBAHASAN

Pengumpulan Data

Pada tahap pengumpulan data ini peneliti mengumpulkan data komentar dari *channel Autonetmagz Riview Mobil Wuling Almaz*. Pengumpulan ini dilakukan secara manual. Data yang terkumpul berjumlah 1000 hanya dengan data komentar berbahasa Indonesia. Data yang diambil dari bulan Maret sampai bulan Oktober tahun 2019. Selanjutnya data yang telah didapat disimpan di excel dengan format .xlsx. setelah data komentar tersebut disimpan pada excel, kemudian data tersebut akan diinputkan pada R Studio dengan sintaks atau perintah sebagai berikut:

```
library(readr)
data <- read_excel("D:/heru/data.xlsx")
```



Gambar 4.1 Data Diinput Pada R Studio

Preprocessing

Pada tahap *preprocessing* ini dilakukan dengan beberapa tahap sebagai berikut:

Case Folding

Pada tahap *case folding* dalam penelitian ini dilakukan dengan cara mengubah huruf besar yang ada pada komentar tersebut menjadi huruf kecil semua. Pada R Studio *case folding* dijalankan dengan perintah sebagai berikut:

Pertama dokumen dirubah ke dokumen corpus dengan target adalah kolom komentar.

```
pro <- Corpus(VectorSource(data$komentar))
lalu komentar dirubah menjadi huruf seragam
yaitu huruf kecil
pro <- tm_map(pro, tolower)
```

Tabel 4.1. Tabel Hasil Proses Case Folding

Input	Output
Pantas" saja	pantas" saja sih
sih kalo di adu	kalo di adu
akomodasi	akomodasi tapi
tapi kalo	kalo sektor
sektor mesin	mesin
sepertinya	sepertinya
almaz masih	almaz masih
ketinggal	ketinggal

Menghilangkan Angka, Emotikon Dan Tanda Baca

Pada tahap ini dilakukan atau diperlukan menghilangkan angka, emotikon dan tanda baca agar komentar benar bersih dan tidak mengubah makna komentar agar mempermudah proses pembagian ke kelas opini positif negatif dan

netral. Perintah untuk menjalankan tahap ini dilakukan dengan cara sebagai berikut:

```
pro <- tm_map(pro, removePunctuation)
pro <- tm_map(pro, removeNumbers)
```

Tabel 4.2. Tabel Hasil Proses Menghilangkan Angka, Emotikon Dan Tanda Baca

Input	Output
pantas" saja sih	pantas saja sih
kalo di adu	kalo di adu
akomodasi tapi	akomodasi tapi
kalo sektor	kalo sektor
mesin	mesin
sepertinya	sepertinya
almaz masih	almaz masih
ketinggal	ketinggal

Normalisasi Kalimat

Pada tahap normalisasi kalimat ini diperlukan untuk memperbaiki kata-katayang tidak sesuai KKBI dan selanjutnya akan mengubahnya sesuai KBBI. Proses ini dilakukan pada R Studio dengan perintah sebagai berikut:

```
pro <- tm_map(pro, gsub, pattern="kalo",
replacement="kalau") mengubah kata kalo
menjadi kalau
```

Tabel 4.3. Tabel Hasil Proses Normalisasi Kalimat

Input	Output
pantas saja sih	pantas saja sih
kalo di adu	kalau di adu
akomodasi tapi	akomodasi tapi
kalo sektor	kalau sektor

mesin	mesin sepertinya
sepertinya	almaz masih
almaz masih	ketinggal
ketinggal	

Filtering

Pada tahap filtering ini dilakukan dengan menghilangkan kata yang tidak memiliki makna seperti kata penghubung seperti “yang” “dan” dan sebagainya di dalam kalimat agar mempermudah untuk proses penelitian. Dalam penelitian ini menggunakan kamus Stopward yang umum digunakan berjumlah 758 kata. Dilakukan pada R Studio dengan menggunakan perintah sebagai berikut: (Aris tri jika H)

```

berkas <-
file("D:/heru/kamus/stopwords.txt", open =
"r")
stop <- readLines(berkas)
close(berkas)
#tambahan kamus stopwords
stopword = c(stop, "yg", "pliss", "gw",
"klo", "jga", "jg", "dgn", "tpi", "ttp",
"tp", "krn", "lgi", "jd", "dr", "dri",
"doank", "koq", "dong", "gua", "gue",
"nya", "sih", "aja", "kalo", "karna",
"dlm", "lbh", "sbg", "cma", "spt", "sdh",
"utk", "mnrt", "kyk", "udh", "byk", "trs",
"pst", "msh", "blm", "skrg", "gede",
"nih", "mah", "tuh", "deh", "yah",
"kyknya", "bnyk", "dah", "doang", "ane",
"donk", "yaa", "kyak", "bro", "gitu",
"emang", "ngga")
pro <- tm_map(pro, removeWords, stopword)

```

Tabel 4.4. Tabel Hasil Proses *Filtering*

Input	Output
pantas saja sih kalau di	adu akomodasi
adu akomodasi tapi	sektor mesin
kalau sektor mesin	almaz ketinggal
sepertinya almaz masih	
ketinggal	

Tokenisasi

Pada tahap tokenisasi dalam penelitian ini memecah kalimat menjadi perkata dan menghitungnya kata tersebut sering muncul dan membangun TDM. TDM sendiri adalah term document *matrix* adalah bentuk *matrix* matematika yang menampilkan frekuensi kata yang terbanyak di dalam *corpus* dan menjadikannya *wordcloud*.

Tabel 4.5. Tabel Hasil Proses Tokenisasi

Input	Output
adu akomodasi	adu, akomodasi,
sektor mesin	sektor, mesin,
almaz ketinggal	almaz,
	ketinggal,

Pada tahap proses tokenisasi dilakukan dengan perintah sebagai berikut:

```

tdm <- TermDocumentMatrix(pro)
a <- as.matrix(tdm)
b <- sort(rowSums(a), decreasing =
TRUE)

```

```
c <- data.frame(word = names(b), freq
= b)
```

Tahap selanjutnya adalah membangun Term Documen Metrix (TDM), TDM itu sendiri berarti bentuk matriks matematika yang menggambarkan frekuensi kata yang sering muncul dalam *corpus*, dalam penelitian ini menampilkan frekuensi kata dari 1000 komentar menghasilkan 2684 kata yang ditunjukkan pada Gambar berikut:

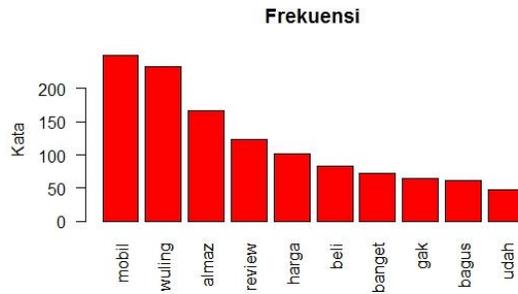
word	freq
mobil	249
wuling	233
almaz	166
review	123
harga	101
beli	83
banget	73
gak	65
bagus	62
udah	48
china	46
keren	44

Showing 1 to 12 of 2,684 entries.

Gambar 4.2 Frekuensi TDM

Dari 1000 komentar tersebut terdapat 2684 kata, namun untuk 10 kata teratas ditampilkan ke dalam diagram. Dengan menggunakan perintah sebagai berikut:

```
diagram <- barplot(c[1:10,]$freq, las =
2, names.arg = c[1:10,]$word, col =
"red", main = 'Frekuensi', ylab = "Kata")
```



Gambar 4.3. Frekuensi Kata

Dari Gambar 4.3. kata “mobil” sampai kata “udah” paling banyak digunakan karena terkait membahas review mobil wuling almaz sendiri. Setelah semua proses *preprocessing* terlewati maka mendapatkan sebuah data yang akan disimpan untuk proses pelabelan komentar. Perintah tersebut dijalankan dengan sintaks sebagai berikut ini:

Catatan dengan data bernama pro adalah data yang sudah di proses.

```
Frame<-
data.frame(text=unlist(sapply(pro, ` `
)),stringsAsFactors=F)
```

Lalu simpan data berupa format csv dan diberi nama bersih

```
write.csv(frame,
"D:/heru/bersih.csv")
```

text
1 hp china ga merk niat banget investasi indonesia alham...
2 fans autonetmagz nonton video barunya kali benarben...
3 power back door palsu ato tar dikasi sgala fasilitas harg...
4 saran reviewer objektif pas review tesla model x setting l...
5 overall wuling almaz tantangan konsumen mesin cvt tur...
6 mobil murah dimaklumin minus wuling pemain dunia m...
7 saran mas klau bandingin produk ya yng fair bandingin ...
8 omongan reviewer menit sepuluh bikin eneg lihatnya ga...
9 review autonetmagz beda detail lengkap tunggutungg...
10 komentator budiman beli produk wuling coba angkat t...
11 elo review ngomong oke tetapitetapiharga bersahabat ...
12 suka review bang ridwan hanif

Showing 1 to 12 of 1,000 entries

Gambar 4.4. Hasil *Preprocessing*

Wordcloud

Setelah tahap preprocessing, komentar dapat ditampilkan ke dalam bentuk *wordcloud*. *Wordcloud* sendiri berarti bentuk visual kata yang berdasarkan jumlah kata yang terdapat pada dokumen. Dalam penelitian ini mengambil kata yang sering muncul minimal 10 dan maksimalnya tidak ada batasan. Proses *wordcloud* sendiri dilakukan dengan perintah sebagai berikut:

```
wordcloud(words = c$word, freq = c$freq,
min.freq = 10, max.words = Inf,
random.order = F, rot.per = 0.35,
colors=brewer.pal(8, "Dark2"))
```



Gambar 4.5. Worcloud Komentar

Pelabelan Menggunakan Metode Lexicon Based

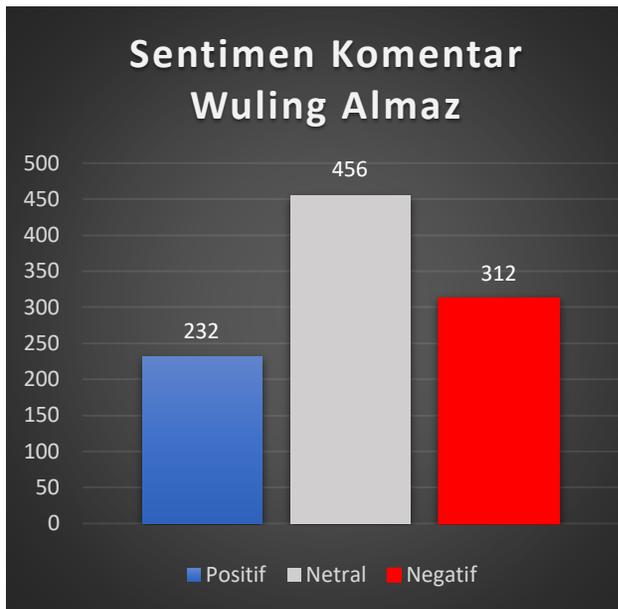
Pelabelan dalam penelitian ini menggunakan data yang telah melewati proses preprocessing dengan metode lexicon based dimana komentar akan dilabeli kedalam kelas positif, negatif dan netral yang mendasari dalam pembagian kelas sentimen adalah skor yang didapatkan apabila komentar <0 maka akan masuk kedalam kelas negatif. Untuk skor >0 masuk ke dalam kelas positif dan $=0$ masuk kelas netral. Dalam penelitian ini menggunakan R studio dan kamus positif, negatif untuk pemberian skor. Berikut ini gambar pelabelan data komentar yang berhasil dibagi kedalam kelas masing masing.

score	text	sentimen
1	hp china ga merk niat banget investasi indonesia alham...	Positif
2	fans autonetmagz nonton video barunya kali benarben...	Negatif
3	power back door palsu ato tar dikasi sgala fasilitas harg...	Positif
4	saran reviewer objektif pas review tesla model x setting l...	Positif
5	overall wuling almaz tantangan konsumen mesin cvt tur...	Positif
6	mobil murah dimaklumin minus wuling pemain dunia m...	Netral
7	saran mas klaw bandingin produk ya yng fair bandingin ...	Negatif
8	omongan reviewer menit sepuluh bikin eneg lihatnya ga...	Negatif
9	review autonetmagz beda detail lengkap tungguttungg...	Netral
10	komentator budiman beli produk wuling coba angkat t...	Positif
11	elo review ngomong oke tetapitengah harga bersahabat ...	Positif
12	suka review bang ridwan hanif	Netral

Showing 1 to 12 of 1,000 entries

Gambar 4.6. Pelabelan Data Komentar

Setelah data komentar dilabeli maka data tersebut dijumlahkan sesuai kelas positif, negatif dan netral.



Gambar 4.7. Sentimen Komentar Wuling Almaz

Hasil pelabelan tersebut didapatkan komentar positif berjumlah 232, komentar netral berjumlah 456 dan komentar negatif berjumlah 312. Hasil tersebut akan masuk proses klasifikasi menggunakan metode naive bayes.

Klasifikasi Naive Bayes

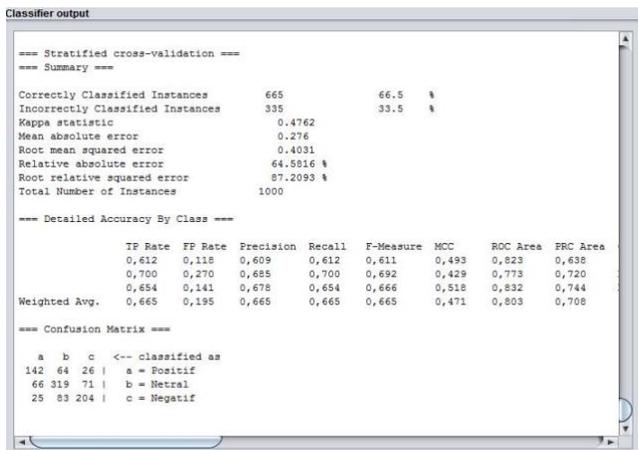
Setelah data dilabel dengan metode Lexicon Based yang didapati dari 1000 komentar menjadi bagian - bagian kelas sentimen positif, netral dan negatif dari proses tersebut menghasilkan berupa 232 sentimen positif, 456 sentimen netral dan 312 sentimen negatif. Selanjutnya data tersebut masuk ke proses klasifikasi dimana penelitian ini menggunakan metode Naive Bayes. Dalam mengaplikasikan metode tersebut diperlukan tool yang dapat membantu proses penelitian, peneliti sendiri menggunakan Weka. Sebelum masuk proses klasifikasi, data hasil pelabelan yang berupa kata akan dirubah menjadi angka dengan TF – IDF. TF (Term Frequency) adalah frekuensi kemunculan sebuah term dalam dokumen yang bersangkutan sedangkan IDF adalah hubungan ketersediaan sebuah term dalam seluruh dokumen. Dalam weka terdapat tool tersebut dalam *StringToWordVector* yang ditunjukkan pada Gambar 4.8.

No.	1: @@class@@	2: act	3: akal	4:	5: alasan	6: all	7: almaz	8:
	Nominal	Numeric	Numeric	akomodasi	Numeric	Numeric	Numeric	
1	Netral	0.0	0.0	2.9926...	0.0	0.0	0.916...	
2	Netral	0.0	0.0	0.0	0.0	0.0	0.0	
3	Negatif	0.0	0.0	0.0	0.0	0.0	0.0	
4	Netral	0.0	0.0	0.0	0.0	0.0	0.0	
5	Negatif	0.0	0.0	0.0	0.0	0.0	0.916...	
6	Netral	0.0	0.0	0.0	0.0	3.19...	0.0	
7	Negatif	0.0	0.0	0.0	0.0	0.0	0.0	
8	Negatif	0.0	0.0	0.0	0.0	0.0	0.0	
9	Netral	0.0	0.0	0.0	0.0	0.0	0.0	
10	Netral	0.0	0.0	0.0	0.0	0.0	0.0	

Gambar 4.8. StringToWordVector

Dalam Gambar 4.8 tersebut dijelaskan kata “akomodasi” terdapat baris ke-1 dengan nilai 2.9926... sedangkan lainnya 0.0 ini menunjukkan kata “akomodasi” dari baris 1 sampai 10 hanya terletak pada baris ke-1. Setelah TF-IDF dilakukan maka proses selanjutnya adalah klasifikasi menggunakan metode Naive Bayes.

Dalam penelitian ini data diuji menggunakan 10 fold cross validation, data akan dibagi menjadi dua yaitu 9/10 bagian adalah data training dan sisanya 1/10 bagian adalah data testing dan dilakukan sampai 10 kali dan diacak. Hasil proses klasifikasi Naive Bayes menggunakan weka ditunjukkan pada Gambar 4.9.



Gambar 4.9. Hasil Klasifikasi Naive Bayes

Gambar 4.9. menunjukkan data klasifikasi yang benar dari metode Naive Bayes berupa 45 teridentifikasi benar dari 82 sentimen positif, 64 teridentifikasi benar dari 114 sentimen netral dan 68 yang teridentifikasi benar dari 104. Setelah didapati klasifikasi tersebut perlu

evaluasi pengujian akurasi, presisi dan recal dari metode tersebut menggunakan Confusion Matrix.

Evaluasi Hasil

Setelah melakukan klasifikasi menggunakan Naive Bayes maka tahap akhir dalam penelitian ini adalah melakukan evaluasi hasil dari metode yang digunakan dan juga untuk menarik kesimpulan dari semua proses penelitian. Evaluasi yang dipilih dalam penelitian ini adalah Confusion Matrix. Confusion Matrix ini bisa digunakan untuk menghitung akurasi, presisi dan recall. Hasil yang diperoleh dari metode Naive Bayes dimasukkan ke dalam tabel Confusion Matrik yang ditunjukkan Tabel 4.6.

Tabel 4.6. Confusion Matrix

Prediksi	Naive Bayes		
	Positif	Netral	Negatif
Positif	142	64	26
Netral	66	319	71
Negatif	25	83	204

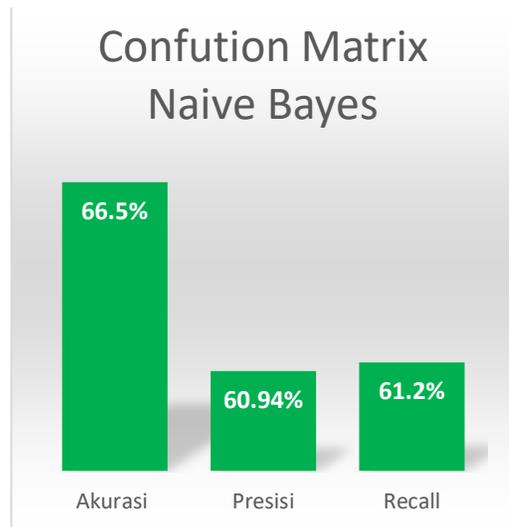
Data hasil confusion matrik selanjutnya menghitung akurasi, presisi dan recall guna mengetahui metode naive bayes untuk data komentar review wuling almaz.

$$\text{Akurasi} = \frac{142+319+204}{142+64+26+66+319+71+25+83+204} * 100\% = 66,5\%$$

$$\text{Presisi} = \frac{142}{142+66+25} * 100\% = 60,94\%$$

$$\text{Recall} = \frac{142}{142+64+26} * 100\% = 61,2\%$$

Dari hasil metode confusion matrix diatas, selanjutnya divisualisasi ke dalam bentuk diagram yang dapat dilihat pada Gambar 4.10.



Gambar 4.10. Diagram Confusion Matrix Naive Bayes

Dari diagram tersebut dijelaskan hasil dari akurasi metode naive bayes mencapai 66,5% sedangkan presisi 60,94% dan recall 61,2%. Menurut (Arikunto, 2013) hasil ukur dapat dikelompokkan menjadi 3 kategori yaitu: Baik (76%-100%), Cukup baik (56%-75%), dan Kurang (<55%). Dari pernyataan tersebut maka untuk akurasi naive bayes dalam klasifikasi 1000 komentar ke dalam 3 kelas sentimen Youtube Autonetmagz review Wuling Almaz memiliki akurasi 66,5% sehingga dikategorikan Cukup baik.

E. PENUTUP

Kesimpulan

Hasil dari penelitian yang telah dilakukan sebelumnya maka dapat disimpulkan sebagai berikut :

- Dari penelitian didapati banyak user berkomentar netral yang diikuti negatif ini menunjukkan penonton channel youtube memberi statement atau opini tentang review mobil wuling almaz yang tidak mengarah kepositif ataupun negatif. Namun ada juga berkomentar positif dan negatif tentang mobil wuling almaz.
- Hasil analisis sentimen yang dilakukan melalui preprocessing dan pelabelan menggunakan algoritma *Lexicon Based* didapati dari 1000 komentar yang masuk kedalam kelas positif berjumlah 232, untuk netral berjumlah 456 dan negatif berjumlah 312.

Kesimpulan selanjutnya dari klasifikasi metode Naive Bayes didapati memiliki akurasi 66,5%, presisi 60,94% dan recall 61,2%, dimana data input berjumlah 1000 komentar menggunakan metode naive bayes memiliki akurasi 66,5% yang dikategorikan Cukup .

Saran

Setelah kesimpulan maka saran yang bisa diberikan untuk penelitian selanjutnya sebagai berikut :

- a. Semoga dalam penelitian selanjutnya dapat menggunakan data komentar dalam jumlah yang banyak lagi.
- b. Dari hasil klasifikasi sendiri metode naive bayes cukup namun dapat juga menggunakan algoritma yang lain sebagai alternatif.

Dalam penelitian selanjutnya di harapkan rentang waktu data yang digunakan dalam penelitian diperpanjang lebih dari 8 bulan.

DAFTAR PUSTAKA

- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal Maret*, 1(1), 32–41.
- ling, j., n. kencana, i. p. e., & oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3(3), 92.
- Nafan, M. Z., Amalia, A. E., Industri, T., Informatika, P., Teknologi, I., & Purwokerto, T. (2019). *Kecenderungan Tanggapan Masyarakat terhadap Ekonomi Indonesia berbasis Lexicon Based Sentiment Analysis*. 3(4), 268–273.

Nur, J., Rohman, A., & Husna, J. (2015). Situs youtube sebagai media pemenuhan kebutuhan informasi : Sebuah survei terhadap mahasiswa program studi ilmu perpustakaan universitas Diponegoro.

Iskandar, Suprpto, (2014) Algoritma, A., & Naïve, C. D. A. N. Perbandingan akurasi klasifikasi tingkat kemiskinan . (xxx), 1–8.

https://www.academia.edu/26318128/preprocessing_text_untuk_meminimalisir_kata_yang_tidak_berarti_dalam_proses_text_mining

<https://github.com/masdevid/ID-OpinionWords>.

<https://github.com/riochr17/Analisis-sentimen-ID/tree/master/kamus>.

<https://medium.com/@mandes95/belajar-data-science-langkah-awal-mengenal-r-dan-rstudio-198ec2246f78>

<https://medium.com/@mandes95/belajar-data-science-memahami-layout-rstudio-d3d46f9f955c>.