

Perbandingan Naïve Bayes dan KNN Dalam Klasifikasi Tweet BBM Subsidi

Doddy Ircham Pambudi¹, Sulastri²

¹Sistem Informasi – Unisbank Semarang, doddyirchamp@gmail.com

²Sistem Informasi – Unisbank Semarang, sulastri@edu.unisbank.ac.id

Jalan Tri Lomba Juang Semarang, Telp. (024) 8451976

ARTICLE INFO

Article history:

Received Desember 10 2022

Received in revised form Desember 14 2022

Accepted Desember 29 2022

Available Online Juli 2023

ABSTRACT

The government that is running at this time is also not spared from public comments on Twitter, especially regarding the increase in subsidized fuel. There are at least 4 impacts felt by the community when subsidized fuel prices increase, namely a decrease in people's purchasing power, an increase in basic prices, an increase in the unemployment rate and an increase in the poverty rate. This study aims to implement the Naïve Bayes Classifier and KNN algorithms in classifying a tweet of an increase in subsidized fuel so that it can be identified as belonging to a class with positive or negative sentiments. The data used in this research are 560 tweets. The data is divided into 2, namely 500 training data from tweet data and 60 test data from tweet data stored in xlsx format. The results of the accuracy with the Naïve Bayes Classifier algorithm is 85% while the KN algorithm is 86.8% so it can be concluded that the KNN method is better than the Naïve Bayes Classifier method in classifying tweets of increases in subsidized fuel.

Keywords: Subsidized BBM, Naive Bayes, KNN

1. Pendahuluan

Kehadiran media dengan segala kelebihanannya telah menjadi bagian hidup manusia. Perkembangan jaman menghasilkan beragam media, salah satunya media sosial. Media sosial merupakan media di internet yang memungkinkan pengguna untuk mewakili dirinya maupun berinteraksi, bekerja sama, berbagi, berkomunikasi dengan pengguna lain, dan membentuk ikatan sosial secara virtual. Media sosial merupakan media digital tempat realitas sosial terjadi dan ruang-waktu para penggunanya berinteraksi. Nilai-nilai yang ada di masyarakat maupun komunitas juga muncul bisa dalam bentuk yang sama atau berbeda di internet. Pada dasarnya, beberapa ahli yang meneliti internet melihat bahwa media sosial di internet adalah gambaran apa yang terjadi di dunia nyata, seperti plagiarisme [1]. Saat ini banyak media sosial yang digunakan oleh masyarakat Indonesia diantaranya Facebook, Twitter, Instagram, Path, Line, dan masih banyak lagi. Media sosial yang akan dibahas dalam penelitian ini adalah media sosial Twitter.

Twitter adalah tempat yang tepat untuk berbagi ide, bank gagasan, tempat untuk mengumpulkan informasi, untuk menginspirasi pikiran, atau untuk melihat apa yang teman lakukan. Twitter sebagai salah satu situs *microblogging* yang mengalami banyak perkembangan, terhitung pada 21 Maret 2022 Twitter genap memasuki usianya yang ke 16. Media sosial ini secara global memiliki

Received Desember, 10, 2022; Revised Desember, 14 2022; Accepted Desember, 29, 2022

332 juta pengguna bulanan dengan 500 juta *tweet* (kicauan) dikirim setiap hari dan 200 miliar kicauan dalam setahun [2]. Banyak pengguna twitter yang melakukan posting ekspresi dan pendapatnya terhadap sebuah produk, layanan, isu politik atau hal – hal yang sedang viral.

Pada pemerintahan di seluruh dunia, twitter dimanfaatkan sebagai sarana untuk lebih dekat dengan masyarakat. Hal tersebut dapat memberikan wawasan yang mendalam tentang apa yang diinginkan oleh masyarakat. Oleh sebab itu, tren yang berkembang saat ini, pemerintah mencoba untuk bergerak lebih dekat ke model *citizen centric*, di mana prioritas dan jasa akan didorong sesuai dengan kebutuhan masyarakat daripada kemampuan pemerintah [3]. Tren ini berkembang didasari oleh fenomena Arab Spring beberapa tahun yang lalu, bagaimana pemerintah dapat berdampak jika mengabaikan sentimen masyarakat. Tren tersebut memaksa pemerintah untuk berpikir ulang dan merancang kembali kebijakannya dalam berinteraksi dengan masyarakat. Untuk membangun pengetahuan tentang apa yang masyarakat inginkan bukanlah pekerjaan yang mudah, mengingat jumlah informasi yang dihasilkan oleh media sosial sangat besar. Untungnya, analisis sentimen atau *opinion mining* dapat berguna untuk kasus tersebut.

Pemerintahan Indonesia juga melakukan pendekatan ke masyarakat menggunakan media sosial seperti Twitter. Hal ini didasari oleh himbuan Menpan tentang pemanfaatan media sosial untuk pemerintah [4]. Indonesia menjadi salah satu negara dengan pengguna Twitter terbesar di dunia. Berdasarkan laporan We Are Social, jumlah pengguna Twitter di Indonesia mencapai 18,45 juta pada 2022. Jumlah tersebut setara dengan 4,23% dari total pengguna Twitter di dunia yang mencapai 436 juta. Jumlah pengguna Twitter di Indonesia pada 2022 juga naik 31,3% dibandingkan setahun sebelumnya yang sebanyak 14,05 juta [5].

Pemerintahan yang sedang berjalan pada saat inipun tidak luput dari komentar publik dalam twitter terutama dalam kenaikan BBM subsidi. Setidaknya ada 4 dampak yang dirasakan masyarakat ketika harga BBM subsidi mengalami kenaikan yaitu turunya daya beli masyarakat, kenaikan harga pokok, menaikkan angka pengangguran dan angka kemiskinan bertambah [6]. Dengan banyaknya pengguna Twitter yang menyampaikan opini-opini tersebut dapat dimanfaatkan untuk mencari sebuah informasi. Namun dalam pemanfaatannya membutuhkan analisis yang tepat sehingga informasi yang dihasilkan dapat membantu banyak pihak untuk mendukung suatu keputusan atau pilihan. Salah satu metode yang dapat digunakan untuk menganalisis opini – opini *tweet* adalah analisa sentimen..

Analisa sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini [7]. Dalam penelitian ini analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek termasuk ke dalam sentimen positif atau negatif menggunakan algoritma *Naive Bayes Classifier* dan *K-Nearest Neighbor* (KNN). Metode *Naive Bayes Classifier* merupakan pengklasifikasian statistik yang bisa digunakan dalam memprediksi probabilitas keanggotaan suatu *class*. *Naive Bayes Classifier* memiliki akurasi dan kecepatan yang sangat tinggi saat diaplikasi ke dalam database dengan data yang besar [8]. Metode KNN adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru (*testing data*) dengan kasus lama (*training data*) yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada [9]

Penelitian oleh Devita dkk (2018) melakukan perbandingan antara metode *Naive Bayes Classifier* dengan metode *K-Nearest Neighbor* dengan hasil metode *Naive Bayes* memiliki kinerja yang lebih baik dengan tingkat akurasi 70%, sedangkan metode *K-Nearest Neighbor* memiliki tingkat akurasi yang cukup rendah yaitu 40% [10]. Pada penelitian oleh Suryono dkk (2018) terkait dengan analisis sentimen menggunakan algoritma *Naive Bayes Classifier* berhasil didapatkan hasil tingkat akurasi pada percobaan pertama sebesar 62.98%, percobaan kedua sebesar 64.95%, percobaan ketiga sebesar 66.36%, dan percobaan keempat sebesar 66.79%. Dari hasil klasifikasi didapat tingkat persentase sentiment positif sebesar 28%, sentiment negatif sebesar 20% dan sentiment netral sebesar 52% [11]. Penelitian oleh Rozi dkk (2018) dari hasil uji akurasi klasifikasi yang dilakukan oleh aplikasi menghasilkan nilai akurasi tertinggi pada setiap kategori positif, negatif, netral

masing-masing sebesar 82%, 92%, 80% dengan jumlah data latih 200 *tweet* negatif, 200 *tweet* positif, dan 200 *tweet* netral.

Penelitian ini bertujuan untuk mengimplementasi algoritma *Naïve Bayes Classifier* dan KNN dalam mengklasifikasikan sebuah *tweet* kenaikan BBM subsidi sehingga dapat diketahui termasuk ke dalam kelas bersentimen positif atau negatif dan menganalisis nilai akurasi dari algoritma *Naive Bayes Classifier* dan KNN.

2. Metode Penelitian

2.1. Metode Pengumpulan Data

Dalam penelitian ini, peneliti melakukan beberapa metode untuk memperoleh data atau informasi dalam menyelesaikan permasalahan untuk menganalisa sentimen *tweet* kenaikan BBM subsidi. Metode yang dilakukan tersebut antara lain :

1. Studi Kepustakaan

Metode pengumpulan data dengan metode kepastakaan dilakukan dengan pengumpulan jurnal atau buku sebagai sumber pustaka yang berkaitan dengan materi penulisan khususnya analisis sentimen menggunakan algoritma *Naive Bayes Classifier* dan KNN seperti Prosiding Seminar Nasional Geotik, Jurnal Informatika Polinema, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer dan lain sebagainya.

2. Pengumpulan Data *Tweet*

Data yang diperoleh merupakan sumber data yang diambil secara langsung dari twitter yang berhubungan dengan kenaikan BBM subsidi. Data *tweet* kenaikan BBM subsidi diperoleh dengan cara *crawling* (mengunduh secara otomatis) dengan menggunakan bahasa R. Untuk mengunduh *tweet* kenaikan BBM subsidi dengan bahasa R dibutuhkan *library* twitter, API key dan acces token dari twitter yang telah didaftarkan. *Script crawling* yang digunakan untuk mendapatkan *tweet* kenaikan BBM subsidi diperlihatkan seperti gambar 1.

```
tweet <- search_tweets(q = "kenaikan bbm", n = 1000)
df <- apply(tweet, 2, as.character)
write.xlsx(df, file = "myworkbook2.xlsx", sheetName = "DATA", append = TRUE)
```

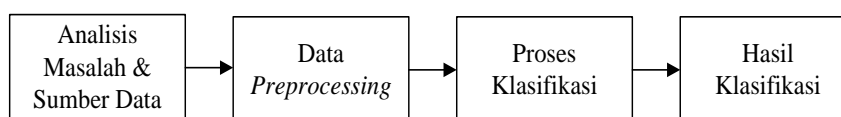
Gambar 1. Script *Crawling*

Data *tweet* kenaikan BBM subsidi yang dihasilkan dari script *crawling* diperlihatkan seperti gambar 2. Data *tweet* kenaikan BBM subsidi dari hasil *crawling* didapatkan data sebanyak 560 *tweet*. Data *tweet* kenaikan BBM subsidi kemudian dibagi menjadi 2 kelompok data yaitu data latih sebanyak 500 data *tweet* yang akan digunakan sebagai melatih model *Naive Bayes Classifier* dan KNN sedangkan data uji dibagi sebanyak 60 data *tweet* yang akan digunakan untuk menguji akurasi model *Naive Bayes Classifier* dan KNN. Data *tweet* kenaikan BBM subsidi yang telah didapat akan disimpan dan disajikan dalam format Microsoft Excel *xlsx*

tweet_id	text	sentiment
919	Perkembangan terbaru #PBBM: Rasio Twitter @BBM	FALSE
918	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
917	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
916	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
915	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
914	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
913	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
912	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
911	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
910	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
909	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
908	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
907	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
906	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
905	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
904	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
903	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
902	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
901	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE
900	Keberhasilan di kota ini bukan hanya sekedar BBM lagi...	FALSE

Gambar 2. Data *Tweet*

2.2. Tahap Klasifikasi



Gambar 3. Tahap Klasifikasi

Algoritma analisis sentimen *tweet* kenaikan BBM subsidi menggunakan algoritma *Naive Bayes Classifier* dan KNN diperlihatkan seperti gambar 3. Gambaran penyelesaian proses klasifikasi *tweet* kenaikan BBM subsidi sebagai berikut:

A. Analisis Masalah dan Sumber Data

Permasalahan yang terjadi dalam klasifikasi *tweet* kenaikan BBM subsidi yaitu:

- a. Banyak pengguna twitter yang melakukan posting ekspresi, opini atau pendapatnya terhadap kenaikan BBM subsidi. Opini-opini tersebut dapat dimanfaatkan untuk mencari sebuah informasi.
- b. Belum adanya sistem yang dapat melakukan klasifikasi *tweet* kenaikan BBM subsidi positif atau negatif.

Untuk mempermudah dalam klasifikasi *tweet* kenaikan BBM subsidi digunakan *text mining* dengan algoritma *Naive Bayes Classifier* dan KNN. Sumber data pada penelitian ini diambil secara langsung dari twitter yang berhubungan dengan kenaikan BBM subsidi. Data yang telah didapat akan disimpan dan disajikan format Microsoft Excel *xlsx*. Setelah data didapatkan, langkah selanjutnya ialah melakukan penentuan klasifikasi secara manual terhadap data tersebut. Data *tweet* akan melewati tahap *preprocessing* kemudian masuk ke tahap klasifikasi. Dalam penelitian ini, *tweet* akan diklasifikasikan menjadi dua klasifikasi yaitu klasifikasi *tweet* positif dan klasifikasi *tweet* negatif.

B. Data Preprocessing

Pada tahap ini, *tweet* yang telah dikumpulkan akan melalui beberapa tahap pemrosesan teks yang terdiri dari *case folding*, *tokenizing* dan *stopword removal*. Tujuan dari pemrosesan teks adalah supaya data yang didapat akan lebih terstruktur agar lebih mudah untuk dilakukan pengolahan data.

C. Proses Klasifikasi

Tahap selanjutnya adalah proses pengklasifikasian *tweet* yang akan diproses menggunakan metode algoritma *Naive Bayes Classifier* dan KNN dengan menghitung probabilitas dari tiap *tweet*.

D. Hasil Klasifikasi

Pada tahap ini adalah menentukan hasil dari klasifikasi menggunakan metode algoritma *Naive Bayes Classifier* dan KNN dari proses klasifikasi, apakah *tweet* termasuk dalam klasifikasi *tweet* positif atau negatif.

3. Hasil dan Analisis

3.1. Membaca Data Tweet

```
df <- read_excel("E:/Rstudio/doddy/data.xlsx")
glimpse(df)
```

Gambar 4. Membaca Data Tweet Excel

Gambar 4. merupakan *script* yang digunakan untuk memuat data *tweet* kenaikan BBM subsidi dalam format excel dengan tipe *.xlsx*. Dalam penelitian ini data yang digunakan adalah data excel yang berisi kumpulan data positif atau negatif yang telah diberi kategori yang nantinya akan digunakan untuk proses pelatihan dan pengujian. Dalam penelitian ini data *tweet* positif atau negatif yang digunakan adalah data *tweet* yang diambil dengan menggunakan teknik *crawling*.

Gambar 5 adalah tampilan dari hasil membaca data excel data.xlsx yang berisi 560 data *tweet* kenaikan BBM subsidi dengan 2 variabel yaitu KATEGORI dan TEXT. Variabel KATEGORI

berisi kategori *tweet* kenaikan BBM subsidi yaitu positif dan negatif, sedangkan variabel TEXT berisi kumpulan *tweet* kenaikan BBM subsidi

	Kategori	TEXT
1	NEGATIF	@PartaiSocmed Hahaha... Kompensasi kenaikan BBM ke...
2	POSITIF	BLT UMKM untuk 12 Juta Pelaku usaha merupakan langk...
3	NEGATIF	Setelah pesta meriah yg menurut Yaqut membawa kebah...
4	POSITIF	@Anggaradimas69 Dukung pemerintah minimalisir dam...
5	POSITIF	Nur Aziz Serahkan Bantuan BLT BBM Dampak Inflasi Kena...
6	POSITIF	RT @catatankaqihati: Pemerintah melalui KemenKopUK...
7	POSITIF	RT @AlshaibaShihab: Pencairan BLT UMKM 2022 kembali...
8	POSITIF	RT @forjakeu: BLT Alih Subsidi BBM telah diterima oleh ...
9	POSITIF	RT @riego_see: Pengalihan subsidi dan kompensasi BB...
10	NEGATIF	@hasymmah Kita tetap kritis kok, Mas...δ<U+009F> <U...
11	POSITIF	Pengalihan subsidi dan kompensasi BBM menjadi BLT BB...
12	NEGATIF	Rambak Dong mohon pengertiannya ya. Margin kami se...

Showing 1 to 12 of 560 entries

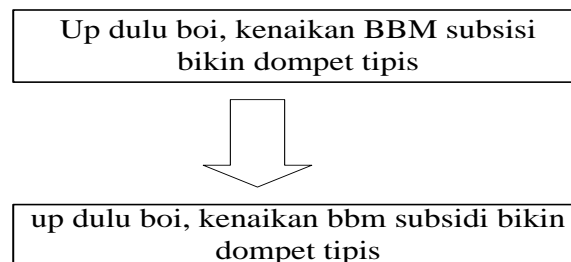
Gambar 5. Hasil Membaca Data *Tweet*

3.2. Text Processing

Proses data *text preprocessing* yang dilakukan yaitu

a. Case Folding

Pada tahap ini akan merubah semua huruf menjadi huruf kecil atau *lowercase*. Proses diawali dengan memeriksa setiap karakter *tweet* kenaikan BBM subsidi dari awal sampai akhir karakter, jika ditemukan karakter yang menggunakan huruf kapital (*uppercase*), maka huruf tersebut akan diubah menjadi huruf kecil (*lowercase*). Proses *case folding* dapat dilihat seperti gambar 6.

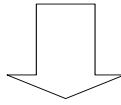


Gambar 6. Case Folding

b. Tokenizing

Pada tahap ini memotong sebuah *tweet* kenaikan BBM subsidi menjadi bagian-bagian, yang disebut dengan token dan pada saat bersamaan juga menghapus karakter khusus yang ada didalam *tweet*, angka dan tanda baca diantaranya @, hashtag (#), titik (.), koma (,), dan tanda baca lainnya karena tidak memiliki pengaruh apapun terhadap nilai klasifikasi sehingga komponen-komponen tersebut harus dibuang. Proses penghapusan karakter khusus yang ada didalam *tweet* diawali dengan memeriksa *tweet* dari proses *case folding* apakah karakter khusus, angka atau tanda baca, jika ada maka akan secara otomatis dihilangkan. Proses penghapusan karakter khusus dapat dilihat seperti gambar 7.

up dulu boi, kenaikan bbm subsidi bikin
domet tipis

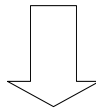


up dulu boi kenaikan bbm subsidi bikin
domet tipis

Gambar 7. Proses Penghapusan Karakter Khusus

Proses selanjutnya adalah memisahkan atau memotong *tweet* menjadi kata yang dipisahkan oleh tanda baca titik (.), koma (,), dan spasi. Proses diawali dengan mengecek *tweet* setelah di lakukan proses penghapusan karakter khusus, jika terdapat tanda baca titik (.), koma (,), dan spasi maka *tweet* akan dipisahkan atau dipotong berdasarkan tanda baca tersebut. Proses *tokenizing* dapat dilihat seperti gambar 8.

up dulu boi kenaikan bbm subsidi bikin domet tipis



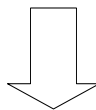
up | dulu | boi | kenaikan | bbm | subsidi | bikin | domet | tipis

Gambar 8. *Tokenizing*

c. *Stopword Removal*

Setelah melewati tahap *tokenizing* maka tahap selanjutnya adalah tahap *stopword removal*. *Stopword* adalah kosakata yang bukan merupakan ciri (kata unik) dari suatu dokumen. Misalnya “di”, “oleh”, “pada”, “sebuah”, “karena” dan lain sebagainya. Tahap *stopword removal* adalah proses menghapus atau menghilangkan kata-kata umum seperti kata sambung, kata depan, kata ganti dan kata yang tidak memiliki makna. Proses diawali dengan melakukan pencocokan ke dalam database *stopword*, jika terdapat kata-kata yang ada pada database maka kata tersebut akan dihilangkan. Proses *stopword removal* dapat dilihat seperti gambar 9.

up | dulu | boi | kenaikan | bbm | subsidi | bikin | domet | tipis



boi | kenaikan | bbm | subsidi | domet | tipis

Gambar 9. *Stopword Removal*

RStudio belum menyediakan *stopword* dalam format bahasa Indonesia sehingga dibutuhkan *stopword* dari luar Rstudio untuk melakukan penghapusan kata sambung seperti gambar 10.

```
stopwordID <- "E:/Rstudio/doddy/stoplist.txt"
cStopwordID<-readLines(stopwordID,warn=FALSE)
```

Gambar 10 *Custom Stopword*

Environment		History	Connections
Global Environment			
Data			
corpus	Large VCorpus (560 elements, 2.3 Mb)		
df	560 obs. of 2 variables		
Values			
cStopwordID	chr [1:758] "ada" "adalah" "adanya" "adapun" "agak"...		
stopwordID	"E:/Rstudio/doddy/stoplist.txt"		

Gambar 11. Hasil *Custom Stopword*

Gambar 11 menunjukkan *custom_stopwords* yang diimportkan memiliki 758 daftar kata. Daftar kata inilah yang akan dihilangkan jika ditemukan dalam *tweet* kenaikan BBM subsidi positif atau negatif. Untuk menjalankan tahap *text processing* script yang digunakan gambar 12.

```
corpus.clean <- tm_map(corpus, content_transformer(tolower))
corpus.clean <- tm_map(corpus.clean, removePunctuation)
corpus.clean <- tm_map(corpus.clean, removeNumbers)
corpus.clean <- tm_map(corpus.clean, removeWords,cStopwordID)
corpus.clean <- tm_map(corpus.clean, stripWhitespace)
```

Gambar 12. *Script Text Processing*

Data *tweet* akan dibagi menjadi data latih dan data uji. Dari 560 data *tweet*, data latih yang dipakai berjumlah 500 data *tweet* sedangkan data uji berjumlah 60 data *tweet*..

```
df.train <- df[1:500,]
df.test <- df[501:560,]
dtm.train <- dtm[1:500,]
dtm.test <- dtm[501:560,]
dim(dtm.train)
corpus.clean.train <- corpus.clean[1:500]
corpus.clean.test <- corpus.clean[501:560]
```

Gambar 13. *Script Pembagian Data*

3.3. Klasifikasi *Naive Bayes*

Untuk melatih model data menggunakan fungsi *Naive Bayes Classifier* dari *packages 'e1071'*. Karena *Naive Bayes Classifier* mengevaluasi produk probabilitas, maka memerlukan beberapa cara untuk menetapkan probabilitas non-nol ke kata-kata yang tidak terjadi dalam sampel dengan menggunakan *laplace = 1*. Gambar 14 merupakan *script* untuk model *Naive Bayes Classifier*.

```
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)
system.time( classifier <- naiveBayes(trainNB, df.train$class, laplace = 1) )
predict(classifier, newdata=testNB)
system.time( pred <- predict(classifier, newdata=testNB) )
table("Predictions"=pred, "Actual" = df.test$class )
conf.mat <- confusionMatrix(pred, df.test$class)
conf.mat
```

Gambar 14. Model *Naive Bayes Classifier*

```

~| ~|
[1] NEGATIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF
[11] POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF
[21] POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF NEGATIF POSITIF NEGATIF
[31] POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF
[41] POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF
[51] POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF
Levels: NEGATIF POSITIF

```

Gambar 15. Hasil Klasifikasi *Naive Bayes Classifier*

Gambar 15 merupakan hasil klasifikasi data uji menggunakan algoritma *Naive Bayes Classifier*. Dari pengujian 60 data *tweet* menggunakan algoritma *Naive Bayes Classifier* memberikan hasil klasifikasi positif sebanyak 53 *tweet* dan klasifikasi *tweet* negatif sebanyak 7 *tweet*.

Proses perhitungan akurasi dari hasil klasifikasi *tweet* menggunakan algoritma *Naive Bayes Classifier* pada data uji akan dicocokkan dengan sentimen yang telah diketahui sebelumnya.

	Reference	
Prediction	NEGATIF	POSITIF
NEGATIF	1	3
POSITIF	6	50

Gambar 16. Hasil *Confusion Matrix Naive Bayes Classifier*

Dari gambar 16 hasil pencocokan data sentimen klasifikasi algoritma *naive bayes* dengan sentimen sebenarnya menghasilkan :

- True* Negatif menghasilkan 1 *tweet*
- False* Negatif menghasilkan 3 *tweet*
- False* Positif menghasilkan 6 *tweet*
- True* Positif menghasilkan 50 *tweet*

```

Accuracy : 0.85
95% CI : (0.7343, 0.929)
No Information Rate : 0.8833
P-value [Acc > NIR] : 0.8434

```

Gambar 17. Akurasi Klasifikasi *Naive Bayes Classifier*

Pada gambar 17 akurasi dari algoritma *Naive Bayes Classifier* dalam melakukan klasifikasi *tweet* dalam penelitian ini sebesar 0.85= 85%.

3.4. Klasifikasi KNN

Untuk melatih model data menggunakan fungsi KNN dari *packages* 'class. Gambar 18 merupakan *script* untuk model KNN.

```

mat.df <- as.data.frame(data.matrix(dtm), stringsAsFactors = FALSE)
mat.df <- cbind(mat.df, df$KATEGORI, row.names = NULL)
colnames(mat.df)[ncol(mat.df)] <- "type"
train <- sample(nrow(mat.df), ceiling(nrow(mat.df) * .89))
test <- (1:nrow(mat.df))[- train]
cl <- mat.df[, "type"]
modeldata <- mat.df[!, colnames(mat.df) %in% "type"]
knn.pred <- knn(modeldata[train, ], modeldata[test, ], cl[train])
knn.pred

```

Gambar 18. Model KNN


```

[1] NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF POSITIF
[11] NEGATIF NEGATIF NEGATIF POSITIF NEGATIF NEGATIF NEGATIF POSITIF NEGATIF NEGATIF
[21] NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF POSITIF NEGATIF NEGATIF NEGATIF NEGATIF
[31] NEGATIF NEGATIF NEGATIF POSITIF POSITIF POSITIF NEGATIF POSITIF NEGATIF POSITIF
[41] NEGATIF NEGATIF POSITIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF NEGATIF
[51] POSITIF NEGATIF POSITIF NEGATIF POSITIF POSITIF POSITIF POSITIF POSITIF POSITIF
[61] POSITIF

```

Gambar 19. Hasil Klasifikasi KNN

Gambar 19 merupakan hasil klasifikasi data uji menggunakan algoritma KNN. Dari pengujian 60 data *tweet* menggunakan algoritma KNN memberikan hasil klasifikasi positif sebanyak 25 *tweet* dan klasifikasi *tweet* negatif sebanyak 36 *tweet*.

Proses perhitungan akurasi dari hasil klasifikasi *tweet* menggunakan algoritma KNN pada data uji akan dicocokkan dengan sentimen yang telah diketahui sebelumnya.

		Actual	
Predictions	NEGATIF	POSITIF	
NEGATIF	35	7	
POSITIF	1	18	

Gambar 20. Hasil *Confusion Matrix* KNN

Dari gambar 20 hasil pencocokan data sentimen klasifikasi algoritma KNN dengan sentimen sebenarnya menghasilkan :

- True* Negatif menghasilkan 35 *tweet*
- False* Negatif menghasilkan 7 *tweet*
- False* Positif menghasilkan 1 *tweet*
- True* Positif menghasilkan 18 *tweet*

Accuracy : 0.8689
 95% CI : (0.7578, 0.9416)
 No Information Rate : 0.5902
 P-Value [Acc > NIR] : 2.146e-06

Gambar 21. Akurasi Klasifikasi KNN

Pada gambar 21 akurasi dari algoritma KNN dalam melakukan klasifikasi *tweet* dalam penelitian ini sebesar 0.868= 86.8%.

4. Kesimpulan

Penelitian ini telah berhasil mengimplementasikan algoritma *Naïve Bayes Classifier* dan KNN dalam mengklasifikasikan sebuah *tweet* kenaikan BBM subsidi sehingga dapat diketahui termasuk ke dalam kelas bersentimen positif atau negatif. Hasil akurasi dengan algoritma *Naïve Bayes Classifier* adalah sebesar 85% sedangkan algoritma KN adalah sebesar 86.8% sehingga disimpulkan metode KNN lebih baik daripada metode *Naïve Bayes Classifier* dalam melakukan klasifikasi *tweet* kenaikan BBM subsidi.

Daftar Pustaka

- [1] R. Nasrullah, *Media Sosial Perspektif Komunikasi, Budaya, Sosioteknologi*, Bandung: Simbiosis Rekatama Media, 2016.
- [2] A. Maulana, "Twitter Rahasiakan Jumlah Pengguna di Indonesia," 23 March 2016. [Online]. Available: <https://www.cnnindonesia.com/teknologi/20160322085045-185-118939/twitter-rahasiakan-jumlah-pengguna-di-indonesia>. [Accessed 5 December 2022].

- [3] R. Arunachalam and S. Sarkar, "The new eye of government: Citizen sentiment analysis in sosial media," *In Sixth International Joint Conference on Natural Language Processing*, p. 23, 2015.
- [4] Permenpan, *Pedoman Pemanfaatan Media Sosial Instansi Pemerintah*, Jakarta, 2012.
- [5] M. A. Rizaty, "Pengguna Twitter di Indonesia Capai 18,45 Juta pada 2022," 10 August 2022. [Online]. Available: <https://dataindonesia.id/digital/detail/pengguna-twitter-di-indonesia-capai-1845-juta-pada-2022>. [Accessed 5 December 2022].
- [6] S. D. Caesaria, "Dosen UM Surabaya: 4 Dampak Kenaikan BBM Subsidi untuk Rakyat," 7 September 2022. [Online]. Available: https://www.kompas.com/edu/read/2022/09/07/091831771/dosen-um-surabaya-4-dampak-kenaikan-bbm-subsidi-untuk-rakyat?page=all#google_vignette. [Accessed 5 December 2022].
- [7] B. Liu, *Sentiment Analysis And Opinion Mining*, Morgan dan Claypool Publisher, 2012.
- [8] B. Santosa, *Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu, 2015.
- [9] Kusri and E. T. Luthfi, *Algoritma Data Mining*, Yogyakarta: Andi, 2015.
- [10] R. N. Devita, H. W. Herwanto and A. P. Wibawa, "Perbandingan Kinerja Metode Naive Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. V, no. 4, pp. 427-434, 2018.
- [11] S. Suryono, E. Utami and E. T. Luthfi, "Analisis Sentiment Pada Twitter Dengan Menggunakan Metode Naive Bayes Classifier," *Geotik*, pp. 9-15, 2018.
- [12] I. F. Rozi, E. N. Hamdana and M. B. I. Alfahmi, "Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode Naive Bayes Classifier (Studi Kasus SAMSAT Kota Malang)," *Polinema*, vol. IV, no. 2, pp. 149-154, 2018.