

**LAPORAN PROYEK AKHIR
PRAKTIKUM DATA SCIENCE
ANALISIS SENTIMEN REVIEW UBER MENGGUNAKAN
METODE RANDOM FOREST**



Zahida Nur Dzakirah 123200010
Windy Febrianti Ode 123200036

**PROGRAM STUDI INFORMATIKA
JURUSAN INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”
YOGYAKARTA
2022**

KATA PENGANTAR

Puji dan syukur penulis ucapkan kehadiran Allah Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga makalah yang berjudul “Analisis Sentimen Review Uber Menggunakan Metode Random Forest ” dapat kami selesaikan dengan baik. Laporan ini ini diajukan sebagai bukti telah menjalankan tugas sarana syarat responsi dari mata kuliah Praktikum Data Science.

Pemilihan tema ini di dasari atas dasar karakteristik media yang sering digunakan dalam social media dan paling banyak digunakan oleh masyarakat. Data yang didapatkan berasal dari data komentar pengguna twitter sehingga digunakan dalam penelitian proyek penulis dengan metode random forest

Laporan ini dapat diselesaikan tepat waktu dan tidak terlepas dari bantuan dan dukungan dari berbagai pihak. Untuk itu penulis ucapkan terima kasih atas kontribusi bantuan dalam berbagai bentuk.

Penulis menyadari bahwa masih banyak kesalahan dalam penyusunan laporan ini, baik dari tatanan bahasa dan penyusunan materi. Maka dari itu penulis mengharapkan kritik dan saran yang membangun untuk menjadikan sebagai evaluasi.

Demikian semoga laporan Analisis Sentimen Review Uber ini bisa diterima sebagai ide atau gagasan yang menambah kekayaan intelektual dalam bidang kajian media. Semoga laporan saya ini dapat bermanfaat bagi pembaca dan juga penulis sendiri.

Yogyakarta, 30 November 2022

Penulis

HALAMAN PENGESAHAN

Analisis Sentimen Review Uber Menggunakan Metode Random Forest

Disusun Oleh:

Zahida Nur Dzakhirah

123200010

Windy Febrianti Ode

123200036

Telah Diperiksa dan Disetujui oleh Asisten Praktikum pada
tanggal:.....

Menyetujui,

Asiten Praktikum

Asisten Praktikum

Dio Cahyo Saputra S.Kom

Vincentius Willy Ardiyanto

123190100

1. PENDAHULUAN

1.1 Latar Belakang

Uber merupakan sebuah aplikasi yang menghubungkan pengemudi (driver) baik mobil atau motor dengan penumpang secara langsung. Aplikasi ini dikembangkan oleh perusahaan teknologi Amerika, Uber Technology, Inc. yang berbasis di San Francisco. Ide ride sharing yang dikembangkan Uber saat ini adalah Garret Camp dan Traffic Kalanic yang pada saat itu terjebak dan sulit menemukan akses layanan taxi. Sebagai sebuah aplikasi seluler, Uber menyediakan beragam layanan, mulai dari transportasi personal, pengiriman makanan, pengiriman paket, kurir, transportasi barang, hingga penyewaan kendaraan bermotor.

Layanan Uber awalnya hanya diluncurkan untuk kawasan San Francisco saja, namun pada tahun 2010 diperluas hingga ke New York. Di luar dugaan, respon publik sangat positif dan antusias dalam menggunakan aplikasi ini sebagai alternatif sistem transportasi umum bersifat personal yang sangat nyaman dengan biaya layanan yang terjangkau.

Seiring dengan waktu, Uber mengalami perkembangan yang sangat pesat. Bahkan Uber termasuk salah satu perusahaan yang mengalami pertumbuhan tercepat di dunia. Layanannya kini telah menjangkau seluruh dunia. Diperkirakan, Uber telah memiliki lebih dari 93 juta pengguna aktif di seluruh dunia. Di Amerika Serikat, Uber menguasai 71% pangsa pasar untuk layanan ride sharing dan 22% untuk layanan pengiriman makanan.

Perkembangan teknologi di berbagai negara memunculkan banyaknya aktivitas manusia yang dapat didukung dengan sarana digital, salah satunya dalam bidang transportasi. Hal ini terlihat dari bermunculannya berbagai model transportasi berbasis online pada kota-kota besar di Indonesia. Salah satunya transportasi online dengan jasa antar adalah Uber.

Pengguna atau pelanggan dapat mengorder layanan yang dibutuhkan melalui aplikasi Uber. Secara otomatis aplikasi akan memberikan rincian informasi pesanan termasuk biaya layanan. Jika pengguna melanjutkan pesanan, aplikasi akan mencari driver guna merealisasikan pesanan tersebut. Pelanggan dapat memantau pergerakan driver dengan mudah, karena aplikasi dilengkapi dengan 'peta layanan' secara real time. Setelah transaksi selesai, pelanggan dapat memberikan peringkat (rating) atas layanan driver. Driver dengan peringkat rendah bisa dinonaktifkan, karena dianggap telah memberikan layanan yang buruk. Hal ini membuat tuntutan akan mutu pelayanan aplikasi meningkat. Sehingga pengguna dapat memberikan rating dan ulasan di berbagai social media salah satunya adalah Twitter.

1.2 Rumusan Masalah

Pada permasalahan proyek kami ini penulis menggunakan metode Random Forest dimana random forest adalah kombinasi dari masing masing tree yang baik kemudian

dikombinasikan ke dalam satu model. Penggunaan metode ini memiliki nilai akurasi yang tinggi.

1.3 Tujuan

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah mengetahui performa dari metode random forest dalam klasifikasi analisis sentiment review uber.

1.4 Manfaat

- Untuk memenuhi salah satu persyaratan responsi dalam mata kuliah Praktikum Data Science (penulis)
- Dapat memahami ilmu pengetahuan serta pengimplementasian wawasan khususnya dibidang Teknik Informatika analisis sentimen.
- Mengetahui hasil akurasi dari random forest untuk memprediksi nilai negatif dan positif dari uber.

2. METODE

Metode Klasifikasi Ensemble (Random Forest) adalah algoritma pembelajaran yang membangun satu set pengklasifikasi alih-alih satu pengklasifikasi, dan kemudian mengklasifikasikan titik data baru dengan mengambil suara dari prediksi mereka. Ensemble Classifier yang paling umum digunakan adalah Bagging, Boosting, dan Random Forest (RF).

Algoritma Random Forest disebut sebagai salah satu algoritma machine learning terbaik, sama seperti Naïve Bayes dan Neural Network. Random Forest adalah kumpulan dari decision tree atau pohon keputusan. Algoritma ini merupakan kombinasi masing-masing tree dari decision tree yang kemudian digabungkan menjadi satu model. Dalam artian besar Random Forest adalah algoritma dalam machine learning yang digunakan untuk pengklasifikasian data set dalam jumlah besar. Karena fungsinya bisa digunakan untuk banyak dimensi dengan berbagai skala dan performa yang tinggi. Klasifikasi ini dilakukan melalui penggabungan tree dalam decision tree dengan cara training dataset yang Anda miliki.

Analisis Sentimen yang digunakan dalam proyek penulis adalah suatu sentiment yang diperoleh dari twitter yang merupakan salah satu media social yang paling banyak digunakan saat ini. Analisis sentiment itu sendiri merupakan Sentiment analysis merupakan salah satu bidang dari Natural Language Processing (NLP) yang membangun sistem untuk mengenali dan mengekstraksi opini dalam bentuk teks. Informasi berbentuk teks saat ini banyak terdapat di internet dalam format forum, blog, media sosial, serta situs berisi review. Dengan bantuan sentiment analysis, informasi yang tadinya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur.

Twitter merupakan salah satu sosial media yang memiliki pengaruh cukup baik dalam menaikkan engagement pada pelanggan yang dimiliki oleh sebuah brand. Salah satu cara yang dilakukan adalah menciptakan format tweet yang sesuai dengan identitas brand itu sendiri. Bentuk respon yang disediakan oleh twitter adalah retweet, favorite,

dan reply. Semakin tinggi respon yang didapatkan, maka akan semakin viral postingan tersebut dan meningkatkan engagement dari konsumen dan calon konsumen yang ada di Twitter. Namun ada 3 jenis sentimen yang terjadi dalam setiap respon, yaitu Positif, Netral, dan Negatif.

Cara kerja algoritma random forest dalam perhitungan manual pada proyek kami sebagai berikut sebagai berikut :

Tabel 1 : Data Testing

No	Class	Data Testing
1	Positif	1000
2	Negatif	1000
3	Netral	2000
Total		4000

Menghitung probabilitas kata dilakukan untuk mendapatkan term dengan nilai yang lebih penting dan dianggap relevan untuk dijadikan kata kunci. Proses pembobotan dilakukan dengan menggunakan algoritma Random Forest dalam proses perhitungan persamaan dapat dilihat pada Tabel 2.

Tabel 2 Contoh Kemunculan Term Frekuensi

Kata	Frekuensi kemunculan kata (W_k)		
	1000 Tweet - Positif	1000 Tweet - Negatif	2000 Tweet - Netral
Top	10	17	25
Succes	31	8	11
Recommend	30	20	23
Great	15	12	40

2.1 Pengumpulan Data

Proses Pengumpulan data dilakukan dengan crawling data twitter menggunakan R library Rlang, akun twitter yang digunakan dan mengambil 5000 tweet activity. Crawling data merupakan tahap dalam penelitian yang bertujuan untuk mengumpulkan atau mengunduh data dari suatu database. Pengumpulan data dari penelitian ini yaitu data yang diunduh dari server twitter berupa user dan tweet beserta atribut-atributnya.

2.2. Preprocessing Data / Pengolahan Data

Pada tahap pengolahan data dilakukan pemilihan atribut yang digunakan, pembersihan data, transformasi data. Atribut dipilih berdasarkan hasil penelitian yaitu: (1) TEMP; (2) SLP; (3) STP; (4) WDSP; dan (5) MXSPD. Kelas dari klasifikasi ini diambil dari atribut PRCP. Pembersihan data dilakukan dengan menghapus data kosong

yang ditemukan pada atribut MXSPD dan PRCP. Jumlah data kosong yang ditemukan adalah 205 data. Hasil dari pembersihan data didapatkan 1983 data yang akan diolah untuk tahap selanjutnya. Redundansi, inkonsistensi dan outlier tidak ditemukan. Fase pengolahan data selanjutnya adalah transformasi data. Tipe data dari atribut PRCP yang awalnya berupa real dengan ditambah kode pengukuran curah hujan, diubah terlebih dahulu ke dalam tipe numerik dengan menghapus kode tersebut. Selanjutnya format data numerikal diubah menjadi tipe data kategorikal, yaitu hujan atau tidak hujan. Apabila PRCP bernilai 0 maka tidak hujan, sebaliknya jika ada nilai PRCP lebih dari 0 maka Hujan. Contoh transformasi data yang dilakukan pada atribut PRCP dapat dilihat pada.

Hasil dari tahap pengolahan data ini adalah dataset yang siap untuk diproses pada algoritma random forest. Keterangan jenis atribut dari dataset yang siap diolah dapat dilihat pada.

2.3 Perhitungan Gini Index

2.3.1 Probobality perhitungan

Kemudian Gini (D1), Gini(D2), dan Gini Split dihitung sebagai berikut:

$$Gini(HOME_TYPE \leq 10) = 1 - (1^2 + 0) = 0$$

$$Gini(HOME_TYPE > 10) = 1 - \left[\left(\frac{1}{3} \right)^2 + \left(\frac{2}{3} \right)^2 \right] = 0.04452$$

$$Gini_{SPLIT} = \left(\frac{2}{5} \right) \times 0 + \left(\frac{3}{5} \right) \times 0.04452 = 0.2671$$

Lanjutkan mencari Gini Split untuk ≤ 15 , ≤ 30 , dan ≤ 31 sehingga akan mendapat seluruh Gini Split.

Tabulates Nilai Indeks Gini untuk Atribut Home_Type setiap kemungkinan Split

Gini SPLIT	Value
Gini _{SPLIT} (HOME_TYPE \leq 10)	0.2671
Gini _{SPLIT} (HOME_TYPE \leq 15)	0.4671
Gini _{SPLIT} (HOME_TYPE \leq 30)	0.3000
Gini _{SPLIT} (HOME_TYPE \leq 31)	0.4800

Tabel 2.3.1 Nilai Indeks Gini

Dari Tabel diatas didapatkan bahwa Gini SPLIT Home_Type \leq 10 adalah yang paling kecil yaitu 0.2671.

Pada penelitian proyek ini data yang diolah menggunakan metode Random Forest. Berikut ini merupakan langkah uji yang dilakukan

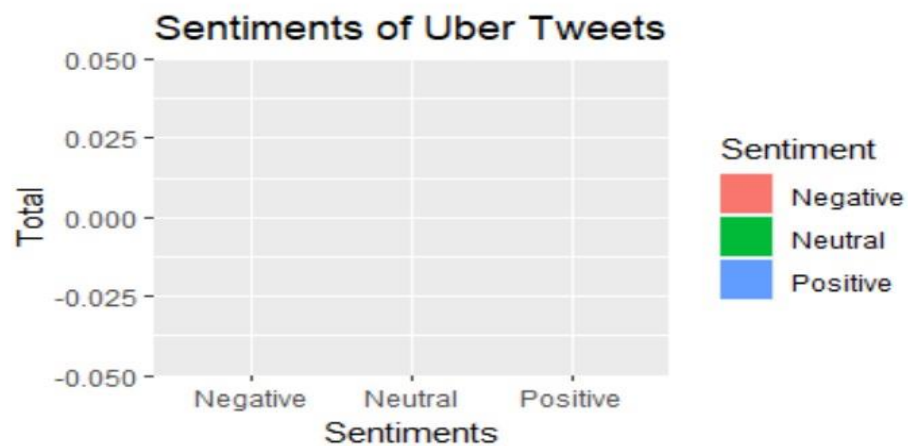
3.1 Data Original

Data berikut merupakan data kotor yang diambil dari twitter.

#	X	Y	Filter	EventId	EventCount	replyToIN	created	truncated	replyToID	id	replyToID	statusSource	screenName	retweetCount	isRetweet	retweeted	longitude
1	RT	@backdoorid	If you're uber eats driver's working just do.	FALSE	0	NA	2022-11-30 03:1451	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	Lenore99	1010	TRUE	FALSE	NA
2	RT	@davidphillips19	We've used to see such a great to.	FALSE	0	NA	2022-11-30 03:1452	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	100logging	2	TRUE	FALSE	NA
3	@jazzu101	We like to take a ride into this, we're just.	FALSE	0	Personal	NA	2022-11-30 03:1453	FALSE	1.5877881e+10	1.5877911e+10	3.036422e+10	https://www.spotify.com	User_Support	0	FALSE	FALSE	NA
4	RT	@backdoorid	If you're uber eats driver's working just do.	FALSE	0	NA	2022-11-30 03:1452	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	HMx	1010	TRUE	FALSE	NA
5	RT	@victorn2022	NEW @victorn2022 just announced.	FALSE	0	NA	2022-11-30 03:1439	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/phone	zuzen	1037	TRUE	FALSE	NA
6	RT	@victorn2022	NEW @victorn2022 just announced.	FALSE	0	NA	2022-11-30 03:1439	FALSE	NA	1.587791e+10	NA	https://mobile.twitter.com	MasterRed1	2327	TRUE	FALSE	NA
7	RT	@backdoorid	If you're uber eats driver's working just do.	FALSE	0	NA	2022-11-30 03:1438	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	Shadow504	1010	TRUE	FALSE	NA
8	RT	@backdoorid	If you're uber eats driver's working just do.	FALSE	0	NA	2022-11-30 03:1437	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/phone	JoopJopar	1010	TRUE	FALSE	NA
9	@DukeTont	You have any other apps other than Faceit.	FALSE	0	NotPublic	NA	2022-11-30 03:1427	TRUE	NA	1.587791e+10	1023221e+27	https://twitter.com/download/android	Brankilugui	0	FALSE	FALSE	NA
10	@donespider	Her uber eats driver's is a waste, man for.	FALSE	0	Personalized	NA	2022-11-30 03:1427	FALSE	1.5877881e+10	1.587791e+10	3.161370e+10	https://twitter.com/download/phone	sonodotgpt	0	FALSE	FALSE	NA
11	RT	@_Miyumi1	Let's not order a Uber for a few d.	FALSE	0	NA	2022-11-30 2022-11-30 03:1427	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	Rakulvika	9	TRUE	FALSE	NA
12	RT	@backdoorid	If you're uber eats driver's working just do.	FALSE	0	NA	2022-11-30 03:1414	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	Megistika1an	1010	TRUE	FALSE	NA
13	RT	@backdoorid	If you're uber eats driver's working just do.	FALSE	0	NA	2022-11-30 03:1357	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	satelavika	1010	TRUE	FALSE	NA
14	14.	update	Following to turn home and adding in the user.	FALSE	0	NA	2022-11-30 03:1350	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/phone	frtghn	0	FALSE	FALSE	NA
15	RT	@backdoorid	If you're uber eats driver's working just do.	FALSE	0	NA	2022-11-30 03:1341	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	Arva0077	1010	TRUE	FALSE	NA
16	16.	Delhi	Policia asks user to verify others before embarking th.	FALSE	0	NA	2022-11-30 03:1337	TRUE	NA	1.587791e+10	NA	https://twitter.com/hoc2022man	0	FALSE	FALSE	NA	
17	RT	@backdoorid	If you're uber eats driver's working just do.	FALSE	0	NA	2022-11-30 03:1347	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	ibemggy	1010	TRUE	FALSE	NA
18	RT	@backdoorid	If you're uber eats driver's working just do.	FALSE	0	NA	2022-11-30 03:1343	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/phone	Snails	1010	TRUE	FALSE	NA
19	RT	@backdoorid	If you're uber eats driver's working just do.	FALSE	0	NA	2022-11-30 03:1342	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	cam2_pricesen	1009	TRUE	FALSE	NA
20	21	@tunaru4016191	NEW @User_Support just announced.	FALSE	0	NA	2022-11-30 03:1330	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/android	0000004016191	1	TRUE	FALSE	NA
21	RT	@victorn2022	NEW @victorn2022 just announced.	FALSE	0	NA	2022-11-30 03:1338	FALSE	NA	1.587791e+10	NA	https://twitter.com/download/phone	0000004016191	1	TRUE	FALSE	NA

Gambar 3.1 Data Kotor

Berikut merupakan tampilan dari analisis sentiment



Gambar 3.2 Sentimen Uber Tweets

3.2 Cleaning

Pada tahap ini dilakukan untuk menghillangkan *URL*, *RT*, *mention*, *hastag*, *new line*, *non alphabet*, *space*, *lowecase*, *stop words*.

	tweet
1	youre uber eats driver walking just cancel food bro
2	uber proud see great turnout first tanggathon session last night thanl
3	wed like take look weve sent direct message please check inbox tha
4	new just announced partnership uber offer young voters georgia free
5	dont adds facenook uber viber snapchat facebook add showing amp
6	uber eats commercial worse mentions favorite cult
7	dont order uber eats days start getting aggressive w notifications kn
8	update listening twin flame sobbing uber
9	delhi police asks uber verify drivers onboarding check alcohol level ric
10	just newbook eridu driver come location strange you cancelled trip

Gambar 3.3 Data Cleaning

Setelah data telah di bersihkan dari beberapa kategori yang dihapus untuk mendapatkan data akurat, maka selanjutnya data yang dihasilkan adalah data terstruktur yang merupakan data yang diperoleh dari tampilan shiny atau data cleaning shiny.

	tweet
1	man masses uber cool
2	close
3	cant drive snow need uber lol
4	youre uber eats driver walking just cancel food bro
5	just tipped uber eats driver coz im pretty sure sex relationship gay rights
6	commute bolt uber daily shit super expensive
7	story complained number times
8	tony ferguson type guy call jean charles skarbowsky uber proceed drive home
9	baltimore police say investigating string robberies ride share drivers carjacked
10	one drivers abused today asked rejecting trip saidkatonbe ma

Gambar 3.4 Tampilan Data Cleaning Shiny

3.3 Data Terstruktur

Berikut merupakan tampilan dari data yang telah di cleaning sehingga menjadi data yang terstruktur

Analisis Sentimen Uber

Menggunakan Algoritma Random Forest

Data Twitter		Data Cleaned	Data Sentimen	Kategori Polaritas			
		Search: <input type="text"/>					
X	text	favorited	favoriteCount	replyToSN	created	truncated	
1	1	RT @AlwaysAkashRC: MAN OF MASSES #RamCharan Uber Cool 🇮🇳 #ManOfMassesRamCharan @AlwaysRamCharan https://t.co/E3Pw3DzWat	false	0	2022-11-30 06:15:54	false	
2	2	@uBer_Poutine So not close.	false	0	uBer_Poutine 2022-11-30 06:15:48	false	15978363
3	3	@patersonjeff You can't drive yourself in the snow? Need a Uber? Lol	false	0	patersonjeff 2022-11-30 06:15:46	false	15977698
4	4	RT @backendoc: If you're uber eats driver is walking just cancel the food bro... https://t.co/QVRL518UHV	false	0	2022-11-30 06:15:45	false	
5	5	Just tipped my Uber Eats driver coz I'm pretty sure she was in a same sex relationship. Gay rights!	false	0	2022-11-30 06:15:45	false	
6	6	RT @vthe2nd: Those of you that commute with Bolt and Uber on a daily, how do you do it ? That shit is super expensive !!!	false	0	2022-11-30 06:15:44	false	

Gambar 3.5 Data bersih

3.4 Shiny Tabel Data

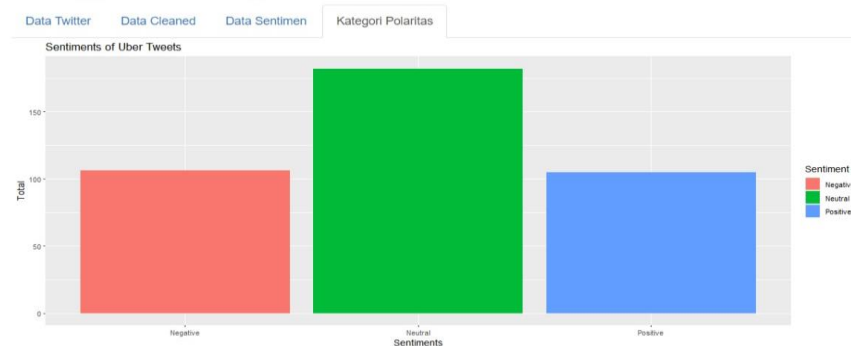
Dibawah ini merupakan tampilan worldcloud.



Gambar 3.6 Tampilan Worldcloud

Analisis Sentimen Uber

Menggunakan Algoritma Random Forest



Gambar 3.7 Tampilan Shiny Kategori

3.5 Listing Program

```
3   # Library
4   ```{r}
5   library(tidyverse)
6   library(tm)
7   library(SnowballC)
8   library(wordcloud)
9   library(ggplot2)
10  library(e1071)
11  library(caret)
12  library(randomForest)
13  ```
14
15  # Ambil data
16  ```{r}
17  # memngambil data dari file csv
18  data <- read.csv("data.csv")
19  datas <- data %>% select(tweet, sentiment)
20  ```
21
22  ```{r}
23  neutral <- length(which(datas$sentiment == "neutral"))
24  positive <- length(which(datas$sentiment == "positive"))
25  negative <- length(which(datas$sentiment == "negative"))
26  Sentiment <- c("Negative", "Neutral", "Positive")
27  Count <- c(negative, neutral, positive)
28  output <- data.frame(Sentiment, Count)
29  ggplot(data=output, aes(x=Sentiment, y=Count)) +
30    geom_bar(aes(fill = Sentiment), stat = "identity")+
31    xlab("Sentiments") + ylab("Total")+ggtitle("Sentiments of
32    Uber Tweets ")
33  ```
34
35  # Membuat corpus (Cleaning)
36  ```{r}
37  corpus <- VCorpus(VectorSource(datas$tweet))
38  corpus <- tm_map(corpus, content_transformer(tolower))
39  corpus <- tm_map(corpus, removeNumbers)
40  corpus <- tm_map(corpus, removePunctuation)
41  corpus <- tm_map(corpus, removeWords, stopwords("english"))
42  corpus <- tm_map(corpus, stemDocument)
43  corpus <- tm_map(corpus, stripWhitespace)
44  ```
45
46  # Membuat matrix dari data corpus
47  ```{r}
48  dtm <- DocumentTermMatrix(corpus)
49  dtm <- removeSparseTerms(dtm, 0.999)
50  ```
51
52  # Mencetak wordcloud
53  ```{r}
```

```

53 wordcloud(datas,min.freq = 500,colors=brewer.pal(8,
  "Dark2"),random.color = TRUE,max.words = 100)
54 ```
55
56
57 ```{r}
58 convert <- function(x) {
59   y <- ifelse(x > 0, 1, 0)
60   y <- factor(y, levels = c(0, 1), labels = c("No", "Yes"))
61   y
62 }
63
64 datanaive <- apply(dtm, 2, convert)
65
66 dataset <- as.data.frame(as.matrix(datanaive))
67 dataset$Class <- as.factor(datas$sentiment)
68 str(dataset$Class)
69 ```
70
71
72 ```{r}
73 set.seed(31)
74 split <- sample(2, nrow(dataset), prob = c(0.75, 0.25),
  replace = TRUE)
75 train_set = dataset[split == 1,]
76 test_set = dataset[split == 2,]
77
78 prop.table(table(train_set$Class))
79 prop.table(table(test_set$Class))
80 ```
81
82
83 # Random Forest Algoritma
84 ```{r}
85 rf_classifier = randomForest(x = train_set[-1210], y =
  train_set$Class, ntree = 300)
86 rf_classifier
87 ```
88
89 # Confusion matrix
90 ```{r}
91 rf_pred = predict(rf_classifier, newdata = test_set[-1210])
92 confusionMatrix(table(rf_pred,test_set$Class))
93 ```

```

Listing 3.5.1 Modelling Lanjutan

```

# Library
```{r}
library(twitteR) # scrapping
library(tm) # corpus
library(syuzhet) # labeling
library(caTools) # split data
library(dplyr) # case when

set.seed(100)
```

# Setting API Twitter
```{r}
consumer_key <- "YTkbdL6EymZylHJp6tiPxUEr"
consumer_secret <-
"0VoTq25uP4DTjnzKNaBhxcCoaSbdvAlfqsviZjd50iCinRj5C4"
bearer_token <-
"AAAAAAAAAAAAAAAAAAD9UjAEAAAAA8m3PgMs24tRhYCuJ0hQaQ5R2
moI%3DThYb3MRjCOlrHidGcDbn4ZDdOG0NtwbeZuY3IDFk0ay8Vdpla5"
access_token <- "1437206819936501760-
zvrXApLJ4MRUDd5zwSDmKXYFTbJ7eV"
access_token_secret <-
"0n6bGTCeEUyDnwaImK0Big8ZhIMtIxwF2IYvXiRPiV20"
setup_twitter_oauth(consumer_key, consumer_secret,
access_token, access_token_secret)
```

# Scraping
```{r}
Scrapping data (Ambil data dari twitter)
tweetsList <- searchTwitter('Uber', n = 1000,
retryOnRateLimit = 10e5, lang = "en")
Mengubah data list twitter menjadi data frame
tweets <- twListToDF(twList = tweetsList)
write.csv(tweets, "data_tweet.csv")
```

# Data Cleaning
```{r}
remove spam tweets
uniqueText <- unique(tweets$text)

remove retweet element
removeRetweet <- function(x) gsub("RT @\\w+: ", "", x)
cleanText <- lapply(uniqueText, removeRetweet)

#remove mention element
removeMention <- function(x) gsub("@\\w+", "", x)
cleanText <- lapply(cleanText, removeMention)

remove url element
removeURL <- function(x) gsub("http\\S+", "", x)
cleanText <- lapply(cleanText, removeURL)

```

```

remove hashtag element
removeHashtag <- function(x) gsub("#\\S+", "", x)
cleanText <- lapply(cleanText, removeHashtag)

remove new line character
removeNewLine <- function(x) gsub("\n", " ", x)
cleanText <- lapply(cleanText, removeNewLine)

remove nonalphabetical character
removeNonAlphabet <- function(x) gsub("[^A-Za-z]", "", x)
cleanText <- lapply(cleanText, removeNonAlphabet)

trim space into one space
cleanText <- lapply(cleanText, stripWhitespace)

text to lowercase
cleanText <- lapply(cleanText, tolower)

remove stop words
cleanText <- lapply(cleanText, removeWords,
stopwords("english"))
dataframe <- data.frame(tweet = unlist(cleanText))
write.csv(dataframe, "data_clean.csv")
```

# Labeling

```{r}
positiveWords <- scan("positive.txt", what = "character",
comment.char = ";")
negativeWords <- scan("negative.txt", what = "character",
comment.char = ";")

menentukan score tweet
scores <- lapply(cleanText, function(cleanText) {
 words <- unlist(str_split(cleanText, pattern = "\\s+"))
 positiveMatches <- !is.na(match(words, positiveWords))
 negativeMatches <- !is.na(match(words, negativeWords))
 score <- sum(positiveMatches) - sum(negativeMatches)
 score
})

melabeli jika score -1 maka negatif, jika score 0 maka
netral, jika score 1 maka positif
sentiment <- as.factor(ifelse(scores < 0, "negative",
ifelse(scores == 0, "neutral", "positive")))
```

# Build csv
```{r}
menyimpan data yang sudah dilabeli ke dalam file csv
dataframe <- data.frame(tweet = unlist(cleanText),
sentiment = sentiment)

```

```
write.csv(dataframe, "data_sentiment.csv")
...

```

### Listing 3.5.2 Scrapper

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

```{r}
library(shiny)
library(vroom)
library(here)
library(dplyr)
library(ggplot2)
library(RColorBrewer)
```

```{r}
ui <- fluidPage(
  headerPanel("Analisis Sentimen Uber"),
  headerPanel("Menggunakan Algoritma Random Forest"),
  mainPanel(
    tabsetPanel(
      tabPanel("Data Twitter",
DT::dataTableOutput('dataTwitter')),
      tabPanel("Data Cleaned",
DT::dataTableOutput('dataCleaned')),
      tabPanel("Data Sentimen",
DT::dataTableOutput('dataSentiment')),
      tabPanel("Kategori Polaritas",
plotOutput("sentiment"))
    )
  )
)
```

```{r}
server <- function(input, output) {
  dataTwitter <- read.csv("data_tweet.csv")
  output$dataTwitter = DT::renderDataTable({
    DT::datatable(dataTwitter, options =
list(lengthChange = FALSE))
  })

  dataCleaned <- read.csv("data_clean.csv")
  output$dataCleaned = DT::renderDataTable({
    DT::datatable(dataCleaned, options =
list(lengthChange = FALSE))
  })
}
```

```

dataSentiment <- read.csv("data_sentiment.csv")
output$dataSentiment = DT::renderDataTable({
  DT::datatable(dataSentiment, options =
list(lengthChange = FALSE))
})

neutral <- length(which(dataSentiment$sentiment ==
"neutral"))
positive <- length(which(dataSentiment$sentiment ==
"positive"))
negative <- length(which(dataSentiment$sentiment ==
"negative"))
Sentiment <- c("Negative", "Neutral", "Positive")
Count <- c(negative, neutral, positive)
plt_sentiment <- data.frame(Sentiment, Count)
plotSentiment <- function(plt_sentiment){
  ggplot(data=plt_sentiment, aes(x=Sentiment, y=Count))
+ geom_bar(aes(fill = Sentiment), stat = "identity")+
  xlab("Sentiments") +
  ylab("Total")+ggtitle("Sentiments of Uber Tweets ")
}

output$sentiment <- renderPlot({
  plotSentiment(plt_sentiment)
})
}
...

```{r}
shinyApp(ui = ui, server = server)
```

```

Listing 3.5.3 Shiny

4. KESIMPULAN SARAN

4.1 Kesimpulan

Dari penelitian proyek, penulis dapat menarik kesimpulan bahwa tingkat akurasi dari penggunaan uber lebih besar sisi netral dibanding sisi negatif di banding nilai positif. Hal ini ditunjukkan pada analisis sentiment dengan grafik yang menunjukkan bahwa nilai positif dan nilai negatif sebesar 100 dengan presentase nilai netral lebih dari angka 150. Sehingga penggunaan aplikasi uber terbilang baik dan sangat membantu masyarakat. Tentunya dengan terus meningkatkan kualitas dengan mempertimbangkan saran dan kritik yang membangun guna keefektifan juga kenyamanan pengguna di masa yang akan datang.

4.2 Saran

Proyek ini masih dalam kata kurang sempurna tetapi masih dapat menjadi literature bagi pembaca dalam pengembangan ilmu pengetahuan. Diharapkan dapat membantu

pembaca dalam memahami materi yang telah penulis tuangkan. Penulis juga mengharapkan kritikan dan saran yang membangun sebagai referensi dalam mengembangkan pengetahuan.

DAFTAR PUSTAKA

- Purwanti,T.,2021: Ekosistem Gojek jadi Andalan Masyarakat Saat Pandemi.
- Ramadinah,S., 2021 : Perbandingan Metode Klasifikasi Naïve Bayes dan SVM (Support Vector Machine) pada Analisis Sentimen Review Gojek.
- Jihad,Adna.AM., dkk. 2021 : Analisis Sentimen Terhadap Ulasan Film Menggunakan Algoritma Random Forest
- Raja, Sandhika.H., dkk. 2019 : Twitter Sentimen Gojek Indonesia dan Grab Indonesia
- Primajaya, A., dkk. 2018 : Random Forest Algorithm for Prediction of Precipitation
- Nanda,S., dkk. 2022 : Analisis Sentimen Kepuasan Pengguna Terhadap Layanan Streaming Mola Menggunakan Algoritma Random Forest