

BAB 4

HASIL PENELITIAN

4.1 RINGKASAN HASIL PENELITIAN

Dalam penelitian ini, dilakukan analisis sentimen pada *tweet* menggunakan metode Naïve Bayes Classifier dengan kata kunci "Obligasi *FR*". Data diambil pada tanggal 1 Mei 2019 - 11 Juni 2023 dengan jumlah 253 data dari *tools* snscape, 503 data dari web apifly pada tanggal 8 November 2016 - 28 Juni 2023 dan diperoleh data gabungan dalam penelitian hanya 466 data. Data yang digunakan dalam tahap *training* data sebanyak 336 data *tweet* dengan masing-masing sentimen sebanyak 168 berlabel positif dan negatif, sedangkan pada proses *testing* menggunakan data sebanyak 130 data yang belum diberi label positif dan negatif. Penelitian ini bertujuan untuk mengetahui tanggapan masyarakat mengenai bagaimana narasi terkait investasi obligasi *FR*. Penerapan analisis sentimen dilakukan menggunakan bahasa pemrograman Python dengan metode Naïve Bayes Classification dan *library* yang sudah ada seperti pandas, numpy, nltk, sastrawi, sklearn, dan pickle. Hasil analisis ditampilkan menggunakan tools Jupyter Notebook. Hasil analisis sentimen didapatkan

4.2 HASIL PREPROCESSING

Setelah melakukan pengambilan data pada *twitter* yang berjumlah 466 data yang sudah disortir, kemudian *tweet* yang sudah didapatkan akan dilakukan pembersihan data menggunakan metode preprocessing data. Data yang akan dibersihkan merupakan data *tweet* yang sudah dikumpulkan dan disaring terlebih dahulu yang ditempatkan pada atribut text. Ada beberapa proses dalam melakukan preprocessing data yang meliputi tahap cleaning data, data reduction, dan lainnya. Proses preprocessing data dilakukan secara bertahap pada data *tweet* yang sudah diambil.

4.2.1 Case Folding

Berikut hasil dari proses dilakukannya case folding dengan program yang sudah dibuat dengan menggunakan perintah `str.lowercase` untuk mengubah huruf kapital menjadi huruf kecil. Berikut hasil dari program case folding dapat ditunjukkan pada gambar 4.1.

	text	case_folding
0	Obligasi FR menarik ges	obligasi fr menarik ges
1	Ada yg mau nabung dan kemarin ngga bagian FR...	ada yg mau nabung dan kemarin ngga bagian fr...
2	@Saham_fess Obligasi FR bisa ditarik kapan aja...	@saham_fess obligasi fr bisa ditarik kapan aja...
3	W0nG dU1tNy4 4j4 b4Ny4K, nG4p41N Ut4Ng wkwkwk...	w0ng du1tny4 4j4 b4ny4k, ng4p41n ut4ng wkwkwk...
4	Aku sendiri harus bener2 memulai dari tahap 1 ...	aku sendiri harus bener2 memulai dari tahap 1 ...
...
595	Maksudnya Presiden yg bisa DIDIKTE oleh para p...	maksudnya presiden yg bisa didikte oleh para p...
596	Nyimpennya lama ga? Kalo lama taro aja di RDPU...	nyimpennya lama ga? kalo lama taro aja di rdpu...
597	Investasi obligasi dinilai lebih menguntungkan...	investasi obligasi dinilai lebih menguntungkan...
598	Obligasi Dinilai Instrumen Paling Menguntun...	obligasi dinilai instrumen paling menguntun...
599	Dari dulu sebenarnya..ada yg lebih ajaib dulu ...	dari dulu sebenarnya..ada yg lebih ajaib dulu ...

Gambar 4.1 Hasil Case Folding

4.2.2 Filtering

Selanjutnya melakukan filtering data menggunakan perintah untuk menghapus karakter-karakter khusus atau simbol-simbol tertentu, angka, tanda baca, spasi pada awal kalimat dan akhir kalimat, dan karakter yang hanya terdiri dari satu huruf dalam teks data. Hasil dari program filtering dapat dilihat pada gambar 4.2.

```

0          obligasi fr menarik ges
1  ada yg mau nabung dan kemarin ngga bagian fr...
2  fess obligasi fr bisa ditarik kapan aja kuponn...
3  wng dutny bnyk ngpn utng wkwkwk dan influence...
4  aku sendiri harus bener memulai dari tahap ya ...
   ...
461 spontanitas itu tidak bohong jadi maklumi aja ...
462 wkkww jelas utang indo dikuasai orng dalam neg...
463 global bonds ind seri usdmil dijual yg jth tem...
464 dana pensiun obligasi tabungan yuk bestie j...
465 dalam periode yang sama saham besar as mencata...
Name: filtering, Length: 466, dtype: object

```

Gambar 4.2 Hasil Filtering

4.2.3 Tokenizing

Tahap ini dilakukan untuk memisahkan teks menjadi token atau kata-kata yang nantinya akan mempermudah melakukan Stopword Removal. Berikut hasil dari program tokenizing dapat dilihat pada gambar 4.3.

```

0          [obligasi, fr, menarik, ges]
1  [ada, yg, mau, nabung, dan, kemarin, ngga, keb...
2  [fess, obligasi, fr, bisa, ditarik, kapan, aja...
3  [wng, dutny, bnyk, ngpn, utng, wkwkwk, dan, in...
4  [aku, sendiri, harus, bener, memulai, dari, ta...
...
461 [spontanitas, itu, tidak, bohong, jadi, maklum...
462 [wkkww, jelas, utang, indo, dikuasai, orng, da...
463 [global, bonds, ind, seri, usdmil, dijual, yg,...
464 [dana, pensiun, obligasi, tabungan, yuk, besti...
465 [dalam, periode, yang, sama, saham, besar, as,...
Name: token, Length: 466, dtype: object

```

Gambar 4.3 Hasil Tokenizing

4.2.4 Stopword Removal

Setelah dilakukan pemisahan teks yang sudah dilakukan oleh tokenizing. selanjutnya melakukan penghapusan kata-kata yang umum dan tidak memiliki makna. Hasil dari program stopwords removal dapat dilihat pada gambar 4.4.

```

0          [obligasi, fr, menarik, ges]
1  [, yg, mau, nabung, , kemarin, ngga, bagian,...
2  [fess, obligasi, fr, , ditarik, kapan, aja, ku...
3  [wng, dutny, bnyk, ngpn, utng, wkwkwk, , influ...
4  [aku, sendiri, , bener, memulai, , tahap, , , ...
...
461 [spontanitas, , , bohong, jadi, maklumi, aja, ...
462 [wkkww, jelas, utang, indo, dikuasai, orng, , ...
463 [global, bonds, ind, seri, usdmil, dijual, yg,...
464 [dana, pensiun, obligasi, tabungan, yuk, besti...
465 [, periode, , sama, saham, besar, as, mencatat...
Name: stopwords, Length: 466, dtype: object

```

Gambar 4.4 Hasil Stopword Removal

4.2.5 Stemming

Selanjutnya proses stemming dilakukan untuk mengubah kata-kata menjadi kata dasarnya. misalkan kata “menarik” setelah dilakukan stemming akan menghasilkan kata “tarik”. Hasil dari program stemming dapat dilihat pada gambar 4.5.

```

0          [obligasi, fr, tarik, ges]
1  [, yg, mau, nabung, , kemarin, ngga, bagi, fr,...
2  [fess, obligasi, fr, , tarik, kapan, aja, kupo...
3  [wng, dutny, bnyk, ngpn, utng, wkwkwk, , influ...
4  [aku, sendiri, , bener, mulai, , tahap, , , ...
...
461 [spontanitas, , , bohong, jadi, maklum, aja, b...
462 [wkwkw, jelas, utang, indo, asai, orng, , nege...
463 [global, bonds, ind, seri, usdmil, jual, yg, j...
464 [dana, pensiun, obligasi, tabung, yuk, bestie,...
465 [, periode, , sama, saham, besar, as, catat, t...
Name: stemmer, Length: 466, dtype: object

```

Gambar 4.5 Hasil Stemming

4.2.6 Normalizing

Tahap akhir dari proses preprocessing yaitu normalizing dimana dilakukan pengubahan kata kata yang tidak baku menjadi kata-kata yang memiliki makna dan untuk menghasilkan representasi teks yang lebih seragam dan dapat diproses dengan lebih efektif dalam tahap analisis teks selanjutnya. asil dari program normalisasi dapat dilihat pada gambar 4.6.

```

0          obligasi fr tarik ge
1  yang mau nabung kemarin ngga bagi fr ori st...
2  fes obligasi fr tarik kapan aja kupon hitung...
3  wng dutny bnyk ngpn utng wkwkwk influencer in...
4  aku sendiri bener mulai tahap isi atm ham...
...
461 spontanitas bohong jadi maklum aja beliau me...
462 wkwkw jelas utang indo asai orng negeri banya...
463 global bond ind seri usdmil jual yang jth temp...
464 dana pensiun obligasi tabung yuk bestie jangan...
465 periode sama saham besar a catat tumbuh tahu...
Name: normalizing, Length: 466, dtype: object

```

Gambar 4.6 Hasil Normalisasi

4.3 HASIL LABELING

Tahap ini dilakukan untuk memberikan nilai sentimen positif atau negatif secara manual berdasarkan pemahaman teks yang terkandung dalam suatu kalimat. teks yang dilabel positif memiliki nilai yang baik dan memihak kepada topik penelitian sedangkan teks yang dilabel negatif mempunyai makna yang membandingkan dan tidak memihak ke topik penelitian. Dari proses tersebut diperoleh data pelabelan sebanyak 336 data yang terdiri dari 168 data positif dan 168 data negatif. Hasil labeling dapat dilihat pada gambar 4.7.

no	text	kelas	label
0	1 obligasi fr tarik ge	positif	1
1	2 fe obligasi fr tarik kapan aja kupon hitung pe...	positif	1
2	3 aku sendiri bener mulai tahap isi atm hampir r...	positif	1
3	4 sucorinvest stable kmrin tara naik hari kompo...	positif	1
4	5 pikir bibit game changer banget yak semalem ab...	positif	1
...
331	332 paham obligasi fr hasil lebih bagus deposito l...	negatif	-1
332	333 cara bisnis bung besar mampu usaha ipo obligas...	negatif	-1
333	334 bilang aja mr presiden kagak ngerti ngomong gr...	negatif	-1
334	335 minat rupiah lg murah kupon bagus	negatif	-1
335	336 obligasi fr lewat bibit sih wkwk	negatif	-1

336 rows × 4 columns

Gambar 4.7 Hasil Pelabelan

Data dari pelabelan positif dan negatif dapat dilihat pada gambar 4.8.

```
df.label.value_counts()
1      168
-1     168
Name: label, dtype: int64
```

Gambar 4.8 Jumlah Label

4.4 HASIL TRAINING DATA

Setelah labeling pada data *tweet* selanjutnya melakukan proses *training* data sebanyak 336 data dengan masing-masing 168 nilai positif dan 168 negatif. Proses ini bertujuan untuk menghitung bobot kata pada setiap *tweet* dan nantinya

digunakan untuk menganalisis kinerja algoritma klasifikasi. Berikut program *training* ditunjukkan pada gambar 4.9.

```
def import_tweets():
    data_training.columns = ['No', 'kelas', 'label', 'text']
    data_training.sentiment = data_training.sentiment.replace(4,1)
    return data_training

def feature_extraction(data_training, method = "tfidf"):
    from sklearn.feature_extraction.text import TfidfVectorizer
    tfv=TfidfVectorizer(sublinear_tf=True)
    features=tfv.fit_transform(data)
    return features

def train_classifier(features, label, classifier = "naive_bayes"):
    from sklearn.metrics import roc_auc_score
    from sklearn.naive_bayes import MultinomialNB
    model = MultinomialNB()
    model.fit(features, label)
    probability_to_be_positive = model.predict_proba(features)[:,:1]

    print ("auc (train data):" , roc_auc_score(label, probability_to_be_positive))
    print ("top 10 scores: ", probability_to_be_positive[:10])

data = np.array(data_training.text)
label = np.array(data_training.label)
```

Gambar 4.9 Program *Training*

4.4.1 TF-IDF

Proses perhitungan term frequency (TF) digunakan untuk mencari nilai kemunculan sebuah kata dalam dokumen, sedangkan perhitungan inverse document frequency (IDF) digunakan untuk mencari nilai kemunculan kata dari keseluruhan dokumen. Hasil tf-idf dan bobot pada masing-masing kata dapat dilihat pada gambar 4.10.

```
features = feature_extraction(data, method = "tfidf")
print(features)

(0, 712)    0.2999146097764378
(0, 23)     0.2903360949965169
(0, 773)    0.6328503059314724
(0, 1169)   0.2194207195355393
(0, 292)    0.37377158536340627
(0, 339)    0.07345568802136354
(0, 811)    0.07215928575151433
(0, 487)    0.36069323957426513
(0, 224)    0.3109721935853038
(1, 745)    0.22648144912167875
(1, 203)    0.20626068163645925
(1, 1142)   0.17882280984802515
(1, 839)    0.20626068163645925
(1, 935)    0.18299345123741578
(1, 1096)   0.15138493805959102
(1, 105)    0.20626068163645925
(1, 908)    0.19904357733324468
(1, 849)    0.1602179598522053
(1, 1005)   0.18765583164063943
(1, 1112)   0.22648144912167875
(1, 1042)   0.11672996196794233
(1, 165)    0.17882280984802515
(1, 377)    0.22648144912167875
(1, 930)    0.1395054533531528
(1, 333)    0.14585253402861123
```

Gambar 4.10 Hasil Pembobotan TF-IDF

4.4.2 Naïve Bayes

Perhitungan naïve bayes digunakan untuk menghitung probabilitas pada setiap kata yang sudah diberikan bobot menggunakan TF-IDF dan memprediksi sejauh mana sample train tersebut cenderung memiliki sentimen positif. Hasil dari perhitungan Naïve bayes dapat dilihat pada gambar 4.11.

```
train = train_classifier(features, label, "naive_bayes")
print(train)

auc (train data): 0.9702558106575964
top 10 scores: [0.51593613 0.48773744 0.59411752 0.65231629 0.62254189 0.67134578
0.42734904 0.79994016 0.81537944 0.74059546]
None
```

Gambar 4.11 Hasil Prediksi Naive Bayes

Berdasarkan gambar 4.11 didapatkan hasil perhitungan naïve bayes classification dengan nilai positif sebesar 97% dengan nilai 10 tertinggi dari fitur yang dijalankan.

4.4.3 Klasifikasi Training Data

Proses ini bertujuan untuk mengetahui akurasi dari pemodelan pada data *training* 336 data dengan masing-masing 168 nilai positif dan negatif menggunakan confusion matrix untuk mengetahui nilai aktual dan prediksi. Hasil dari model klasifikasi dapat dilihat pada gambar 4.12.

```
nbmodel = MultinomialNB(alpha=0.1)
nbmodel = nbmodel.fit(X_train_dtm,Y_train)
Y_pred = nbmodel.predict(X_test_dtm)
print(accuracy_score(Y_test,Y_pred))

0.75

from sklearn.metrics import confusion_matrix
cm = confusion_matrix(Y_test, Y_pred)
print(cm)

[[21 10]
 [ 7 30]]

from sklearn.metrics import classification_report
report = classification_report(Y_test, Y_pred)
print(report)
```

	precision	recall	f1-score	support
-1	0.75	0.68	0.71	31
1	0.75	0.81	0.78	37
accuracy			0.75	68
macro avg	0.75	0.74	0.75	68
weighted avg	0.75	0.75	0.75	68

Gambar 4.12 Hasil Klasifikasi

Diperoleh data pada gambar 4.12 klasifikasi *training* data dengan nilai presisi mendapat nilai 75% negatif dan 75% positif, nilai recall 68% negatif dan 81% positif, dengan nilai akurasi 75%.

4.5 HASIL TESTING DAN EVALUASI

Tahap ini menggunakan data *testing* dengan jumlah 130 data yang belum diberikan label positif dan negatif, selanjutnya dilakukan penghitungan nilai akurasi untuk mengetahui perbedaan antara data *training* dan data *testing*. Hasil perhitungan klasifikasi data *testing* dapat dilihat pada gambar 4.13.

Confusion Matrix pada gambar 4.13 didapat hasil TP (True Positif) = 13,

```
nbmodel = MultinomialNB(alpha=0.2)
nbmodel = nbmodel.fit(X_train_dtm,Y_train)
Y_pred = nbmodel.predict(X_test_dtm)
print(accuracy_score(Y_test,Y_pred))

0.9230769230769231

from sklearn.metrics import confusion_matrix
cm = confusion_matrix(Y_test, Y_pred)
print(cm)

[[13  1]
 [ 1 11]]

from sklearn.metrics import classification_report
report = classification_report(Y_test, Y_pred)
print(report)
```

	precision	recall	f1-score	support
-1	0.93	0.93	0.93	14
1	0.92	0.92	0.92	12
accuracy			0.92	26
macro avg	0.92	0.92	0.92	26
weighted avg	0.92	0.92	0.92	26

Gambar 4.13 Hasil Testing

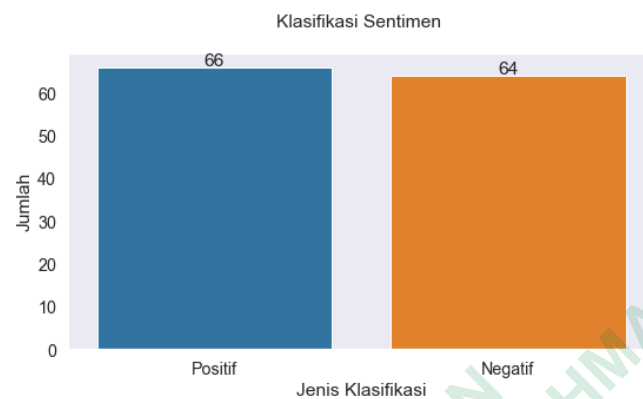
FP(Fasle Positif) = 1, TN (True Negatif) = 1, FN (False Negatif) = 11. Dengan demikian maka perbandingan hasil antara data *training* sebanyak 336 data dan data *testing* 130 data memiliki hasil pengujian yang dapat dilihat pada tabel 4.1.

Tabel 4.1 Perbandingan Hasil Akurasi Data

Akurasi	Hasil
Data Training(336)	0,75
Data Testing(130)	0,92

4.6 HASIL ANALISIS

Setelah dilakukan pengujian pada data *tweet* dengan jumlah 130 data. Hasil rincian data uji yang berpedoman pada data *training* yang sudah dilabeli menghasilkan data yang dapat dilihat pada gambar 4.14.



Gambar 4.14 Hasil Data Uji

Pada hasil pengujian tersebut didapatkan 66 data positif dan 64 data negatif. Pada langkah hasil analisis dilakukan perhitungan seluruh data pengujian dan berhasil mendapat nilai akurasi yang cukup baik sebesar 92% dan menandakan bahwa hasil pengujian memiliki nilai yang baik dalam memberikan nilai sentimen pada data *tweet* yang disediakan.

Sentimen positif pada data *tweet* obligasi *FR* membahas mengenai beberapa hal yaitu ketertarikan untuk berinvestasi ke obligasi *FR* karena memiliki keuntungan ketika sudah jatuh tempo, lebih memilih obligasi *FR* daripada investasi yang lain atau yang memilih obligasi *FR* serta investasi lain, harga lebih kompetitif obligasi *FR*. Dari banyaknya sentimen positif yang didapat, netizen *twitter* menilai obligasi *FR* berdasarkan imbal hasil yang diberikan setiap enam bulan sekali. Contoh 5 data sentimen positif dapat dilihat pada Tabel 4.2.

Tabel 4.2 Data Positif

No	Data Tweet
1	makin pjg tenor dlm suatu instrumen investasi risiko likuiditas makin tinggi jk kamu investasi dengan tenor yang lbh pjg mendapat invest pd obligasi negara non ritel seri fr fixed rate yang juga banyak pilih seri sp

2	menaruh bibit obligasi fr bisa tuh liquid nge beat inflasi yield nya gede kalau ngambil jatuh tempo nya lama
3	saya dukung investasi obligasi fr saya tidak masalah kalau pasar platform tentu transparan sampai mungkin baik jelas lebih lengkap sesuai utama investor jual jatuh tempo
4	menaruh bibit obligasi fr bisa tuh liquid nge beat inflasi yield nya gede kalau ngambil jatuh tempo nya lama
5	yang low risk aja kalau tidak sempet mantau taruh rdpu rdo coba obligasi fr

Sedangkan sentimen negatif pada *twitter* beberapa membahas tentang hal yang tidak berkaitan dengan investasi obligasi *FR* dan tidak memihak atau masih mempertanyakan Obligasi *FR*. Contoh 5 data sentimen negatif dapat dilihat pada tabel 4.3.

Tabel 4.3 Data Negatif

No	Data Tweet
1	obligasi fr eh buntut jual tol
2	bagaimana cara mbk cari info prospektus dari buah obligasi surat hutang entah obligasi biasa seri fr obligasi syariah sukuk seri pb obligasi retail ori seri ori obligasi retail syariah sukri seri sr
3	pimpin orientasi uang tidak heran kalau menteri manfaat umat khusus umat islam yang notabane besar indonesia mulai dari dana haji bangun jalan tol dengan beli obligasi seri fr yang aset adl jalan tol hingga beli obligasi bumk karya
4	bisnis ambil cth rd pasar uang pak biasa mi tempat dana kelola surat utang negara sun seperti fr seri xxx sbn ori sukri kemudian masuk obligasi swasta seperti obligasi biaya leasing obligasi perush go publik amp obligasi bumk
5	fr jadi bank pertama jual surat harga seperti sun obligasi sbn yang milik bank tsb kemudian bank usaha tagih kalau perlu paksa baca ambil alih aset kpd yang ambil pinjam bank terus kalau cara tsb rasa kurang yah bank jual beberapa aset

Berdasarkan hasil klasifikasi sentimen investasi obligasi *FR* menggunakan naïve bayes banyak opini yang diklasifikasikan kedalam sentimen positif dari pada sentimen negatif. Sentimen positif ini memiliki nilai yang baik yang nantinya bisa digunakan untuk mengetahui pandangan masyarakat terkait investasi obligasi *FR*.