

ANALISIS SENTIMEN KEPUASAN INVESTASI PADA MASA PANDEMI DENGAN METODE SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBORS

Nabila Cahyani¹, Fauziah²

Jurusan Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika^{1,2}
Universitas Nasional, Jl. Sawo Manila, Pejaten Ps. Minggu Jakarta 12520
fauziah@civitas.unas.ac.id²

Submitted August 6, 2022; Revised November 14, 2022; Accepted November 29, 2022

Abstrak

Investasi merupakan kegiatan untuk menanamkan asset kepada orang atau lembaga tertentu agar mendapatkan keuntungan dengan jangka waktu tertentu. Investasi tak hanya dilakukan untuk jangka pendek, investasi juga bisa untuk menabung yang digunakan untuk masa depan. Investasi kini menjadi trend semenjak pandemi melanda. Banyak sekali orang ikut terjun kedalam dunia ini dengan harapan mendapatkan keuntungan yang sangat menjanjikan, baik dari usia remaja maupun dewasa, mereka sudah mencoba terjun kedalam dunia investasi. Namun sebagian orang masih kurang paham tentang cara berinvestasi yang benar, sehingga resiko yang dihadapi juga cukup tinggi karena akan berdampak pada keuntungan yang didapat. Terlebih lagi pada masa pandemi saat ini inflasi terhadap nilai Indeks Harga Saham Gabungan bisa mengalami fluktuasi seiring dengan terjadinya berbagai hal. Penelitian ini bertujuan untuk meninjau bagaimana respon masyarakat dalam melakukan investasi dimasa pandemi saat ini. Penelitian menggunakan metode *Support Vector Machine* dan *K-Nearest Neighbors*. Dengan menggunakan metode *Support Vector Machine* dan *K-Nearest Neighbors* dapat dilihat SVM lebih baik dibandingkan KNN. Dengan nilai akurasi SVM sebanyak 96.84% dan KNN sebanyak 95.74% dengan selisih akurasi kedua metode 1.10%.

Kata kunci: *Investasi, K-Nearest Neighbors, Pandemi, Sentimen, Support Vector Machine,*

Abstract

Investment is an activity to invest assets in certain people or institutions in order to get profits for a certain period of time. Investment can not only be made for the short term, but also be used as saving for the future use. Investment has now become a trend since the pandemic started. Many people jump into this activity with the hope of getting very promising profits. Both teenagers and adults have tried to plunge into the investment activity. However, some people still don't understand how to invest properly, so that they must take is quite high risks that create a negative impact on the profits. During the current pandemic, it is worsened by the inflation on the value of the Composite Stock Price Index that can fluctuate along with various things. This study aims to review how the public responds in making investments during the current pandemic. The research uses the Support Vector Machine and K-Nearest Neighbors methods. Seeing the use of the Support Vector Machine and K-Nearest Neighbors methods, it can be concluded that SVM is better than KNN, with the SVM accuracy value of 96.84% and that of KNN of 95.74% with a accuracy difference of 1.10% between the two methods.

Keywords: *Investment, K-Nearest Neighbor, Pandemic, Sentiment, Support Vector Machine*

1. PENDAHULUAN

Semenjak dunia dilanda pandemi covid-19 index harga saham mengalami keterpurukan yang dapat dilihat dari abnormalnya return dan pesimisnya return yang didapat investor dimasa depan. Selain

dari pandemi covid-19 ini index harga saham juga bisa terpengaruh dari hal yang terduga atau bisa disebut dengan Black Swan Even, ekspetasi investor dalam membeli dan menjual saham dapat berpengaruh terhadap peristiwa seperti pandemi, terorisme dan bencana alam[1].

Di Indonesia sendiri pada awal Januari 2020 IHSG turun sebanyak 5,71% , 8,20% di bulan Februari 2020 dan 16,76 pada bulan Maret 2020 [2]. Selama pandemi pemerintah melakukan *social distancing* untuk mencegah penyebaran virus yang semakin meluas. Covid-19 telah menyebabkan perubahan industrial dari beberapa sektor sehingga menyebabkan naiknya permintaan yang tidak terduga[3]. Pada penelitian yang berkaitan sentiment investor di media sosial dapat mempengaruhi dalam memprediksi harga saham sehingga dapat mempengaruhi investasi. Selain itu variable ekuitas juga dapat mempengaruhi prediksi harga. Selama pandemi covid-19 pasar saham dan obligasi juga berpengaruh karena dapat mengalami fluktuasi[4]. Penelitian sebelumnya menggunakan Google Trends untuk meneliti dampak pasar saham dan ekonomi makro selama pandemi berlangsung terlihat impact yang dirasakan cukup besar di berbagai sektor seperti pariwisata , hotel, dan industri penerbangan[5].

Dengan melakukan analisis sentiment dapat dengan mudah menarik kesimpulan dari opini masyarakat tentang suatu produk atau trend yang sedang terjadi, opini ini berupa pandangan atau statement baik positif dan negative. Analisis sentiment berguna untuk pengambilan keputusan dari parameter yang dijadikan topik analisis, sehingga bisa menjadi bahan evaluasi. Proses analisis sentiment ini berkaitan pengolahan data teks menjadi pola data yang terstruktur dan konseptual sehingga data yang dihasilkan berkaitan dengan data bersih yang sudah melalui beberapa tahap seperti case folding untuk mengubah kata menjadi *lowercase*, filtering , *stopword removal*, dan *tokenizing* sehingga dataset yang diperoleh jauh lebih terstruktur untuk kemudian akan dilabeling yang nanti akan menghasilkan output positif dan negatif[6]. Yang nantinya dari labeling ini akan

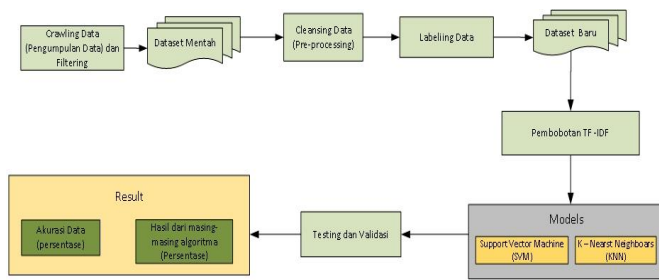
menjadi perbandingan apakah sentiment yang dihasilkan sudah sesuai dengan prediksi dari metode yang dipakai.

Penelitian terkait dengan metode SVM dapat menghasilkan nilai akurasi sebesar 87% dengan nilai positif sebesar 88%, negative sebesar 67%. Pada penelitian ini dilakukan pembobotan TF – IDF untuk memberikan nilai pada kata yang sudah diekstrak, yang nanti nilai pembobotan ini akan digunakan pada proses metode SVM nya[6]. Penelitian lainnya dengan menggunakan metode SVM dan KNN menghasilkan nilai akurasi 71,19% pada metode SVM dan KNN mendapatkan nilai 72,96% dengan mengambil data pada twitter. Dengan data pada metode SVM negative ada 122 data dan positif ada 63 data, sedangkan dengan metode KNN data negative ada 88 data dan positif ada 25 data [7]. Sementara penelitian dengan metode yang sama nilai SVM lebih tinggi dibandingkan dengan nilai KNN, dengan nilai akurasi yaitu 90,01% dan KNN dengan nilai 62,01% [8].

Penelitian ini menjelaskan bagaimana metode SVM dan KNN digunakan untuk menganalisis suatu topik berupa “Investasi dimasa pandemic”. Metode SVM digunakan untuk pengklasifikasian suatu prediksi dengan memisahkan dua fungsi kelas yang berbeda secara optimal[9]. Sedangkan metode KNN digunakan untuk mencari objek pada data train yang dekat atau mirip dengan data testing [10].

2. METODE PENELITIAN

Penelitian ini menggunakan metode SVM dan KNN dengan menggunakan data yang sudah dikumpulkan dari beberapa platform diolah menjadi dataset untuk bisa menghasilkan kesimpulan metode mana yang baik digunakan untuk studi kasus ini , sehingga alur kerjanya dapat digambarkan sebagai berikut:



Gambar 1. Flowchart Sistem Kerja Analisis Sentiment

Pada gambar 1 merupakan alur kerja sistem analisis sentiment investasi dimasa pandemi. Tahapan dimulai dengan cara mengumpulkan data yang dijadikan studi kasus, data yang terkumpul sudah difiltering dan menjadi dataset mentah karena belum di *pre-processing*. Selanjutnya tahap *pre-processing* dimana data akan dicleansing agar menjadi dataset yang baru. Dataset akan dilabeling sentiment positif dan negatif. Kemudian dataset melakukan pembobotan TF-IDF, setelah itu dataset diolah dengan metode SVM dan KNN. Hasil dari kedua metode berupa persentase sentiment yang didapat serta akurasi yang didapat dari kedua metode tersebut.

Pengumpulan Data dan Filtering

Pengumpulan data diperoleh dari beberapa platform atau media sosial seperti twitter, quora, dan sindonews yang terdiri dari data train dan data testing. Data yang sudah dikumpulkan akan menjadi sebuah dataset yang sudah melalui tahap filtering yaitu penghapusan simbol atau tanda baca, url dan elemen yang tidak diperlukan.

Cleansing Data (Pre-processing)

Tahap selanjut nya adalah *pre – procesing* yaitu tahap pemberishan data, data yang didapat kadang masih terdapat hal hal yang tidak diperlukan seperti singkatan dan kata yang kurang baku oleh karena itu dibutuhkan *pre-processing* dengan proses *pre-processing* yang digunakn seperti *case*

folding, tokenizing, dan stopwords removal[10].

Labeling

Setelah dataset melalui tahap *pre-processing* selanjutnya data kan diberi labeling. Pada tahap labeling ini akan menggunakan Python. Labeling data akan diberi nilai sentiment yang nanti nya output yang didapat berupa sentiment positif dan negatif. Sentiment yang didapat ini akan menjadi pembanding dengan metode yang akan digunakan untuk analisis sentiment ini.

Proses Text

Setelah data melalui tahap labeling, dataset selanjutnya akan diolah dengan menggunakan metode SVM dan KNN . Pada tahap ini terjadi pengklasifikasian data pada masing masing metode, dan akan menjadi pembanding apakah labeling sebelumnya sesuai prediksi atau tidak dengan metode yang digunakan.

1. Pembobotan TF-IDF

Pembobotan TF – IDF dilakukan untuk melakukan pembobotan pada model vektor , pembobotan dilakukan untuk mencari frekuensi kemunculannya kata pada kumpulan dataset. TF (*Term Frequency*) menunjukkan frekuensi kata yang muncul, sedangkan IDF(*Inverse Document Frequency*) menghitung frekuensi kemunculan kata , dengan cara mengalikan TF dengan IDF[9].

2. Support Vector Machine

Support Vector Machine merupakan metode pengkasifikasian suatu prediksi atau peramalan. Metode ini memiliki kemampuan dalam mengearlisasi data lebih baik. *Support Vector Machine* melakukan pengklasifikasian melalui training data set untuk membuat prediksi dari data yang baru. Metode ini bertujuan untuk membagi dua buah *class* pada *input space* dengan mengukur Hyperline antara dua *class* menggunakan *margin hyperline* dengan

beberapa model yang disebut kernel SVM, salah satunya mode linear [6].

Linear

$$(x_i) = x_i^T x_j + C \quad (1)$$

Dengan x_i dan x_j merupakan vektor dari data set dan C merupakan constant

3. K-Nearest Neighbors

K-Nearest Neighbors adalah proses pengklasifikasian terhadap perhitungan bobot perkata dalam data latih paling dekat dengan objek data testing pada kelompok k objek dengan menghitung jarak data latih yang dekat dengan objek yang akan diklasifikasi [10].

Rumus :

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum k_i} = 1 \quad (2)$$

Dengan $d(x, y)$ sebagai jarak antar variabel x_1 dan y_2 , x sebagai variabel dan k sebagai jumlah variabel.

Hasil Akurasi

Pada tahap data yang sudah akan diuji kembali apakah hasil yang sudah didapat sesuai atau tidak sehingga pada tahap ini kemungkinan akan terjadi perubahan prediksi dan jumlah data prediksi perkelas dari tahap labeling sebelumnya. Hasil kedua metode akan menghasilkan output yang berbeda sehingga pada tahap ini akan diperoleh mana metode yang sesuai untuk studi kasus ini.

4. HASIL DAN PEMBAHASAN

Pengumpulan data dilakukan dengan meng-crawling data twitter dengan menggunakan library tweepy dari python dan library scrapy untuk mengambil data berita, sedangkan untuk mengambil data quora dilakukan dengan aplikasi ParseHub dengan menggunakan keyword "investasi pandemi". Data yang dicrawling sudah di filtering sehingga simbol atau tanda yang

tidak diperlukan sudah dihilangkan. Data yang diperoleh berbentuk dataset yang berisi 3.640 data. Dataset yang di peroleh masih dataset mentah sehingga harus tahap *pre-processing* seperti *case folding*, *stopword removal*, dan *tokenizing*.

Tabel 1. Jumlah Data yang Diperoleh

Platform	Data
Twitter	3.568
Quora	51
Sindonews	20
	3639

Berdasarkan tabel 1 diatas menunjukkan data yang diperoleh berjumlah 3.640 data. Data twitter diambil dari periode April – Juni 2022 dengan data 3.569 data yang didapat, data Quora diambil pada periode 2021-2022 dengan data yang diperoleh sebanyak 51 data, dan data pada portal berita Sindonews mendapatkan 20 data dengan periode data yang diambil pada bulan Mei.

Tabel 2. Dataset Sebelum Pre-processing

Data
RT Kinerja Investasi Tetap Meningkat Ditengah Pandemi Covid 19 https://t.co/AOtK6sNx76
GoTo perusahaan domestik kebanggaan nasional mampu menarik investasi luar biasa dari luar negeri https://t.co/NWGbYbT01X
Investasi industri manufaktur tembus Rp 103,5 triliun pada triwulan I 2022 Semangat dan selalu yakin Indonesia https://t.co/AynAJceSVL

Pada tabel 2 terlihat data yang diambil masih mengandung elemen yang tidak diperlukan seperti url, dan RT pada twitter, sehingga perlu melakukan *pre-processing* seperti *case folding* untuk mengubah semua huruf uppercase menjadi *lowercase*, *tokenizing* pemotongan perkata dalam suatu kalimat. Dari *tokenizing* ini kata perkalimat dapat dilihat frekuensi kata yang sering digunakan, yang terakhir *stopword removal* yaitu menghapus kata yang tidak memiliki makna.

Pre –Processing



Gambar 2. Flowchart alur pre-processing

Pada gambar 2 diatas, setelah data terkumpul selanjutnya melakukan *pre-processing* data , karena data yang terkumpul masih belum terstruktur. Langkah yang diperlukan untuk melakukan preprocessing untuk membuat dataset menjadi lebih terstruktur. Pre-processing dilakukan dengan menggunakan python pada jupyter notebook, dengan menggunakan library pandas, numpy, re, emoji, dan constraction., Dengan dataset yang sudah diperoleh, selajutnya dataset tersebut akan di cleasing terlebih dahulu untuk membuat dataset lebih tersrstruktur berikut implementasinya :

```

def cleansing(Text):
    # replace tag rt
    t1 = re.sub('RT\s','',Text)
    # replace text @username
    t2 = re.sub('\B@w+', '', t1)
    # normalisasi emoji with text
    t3 = emoji.demojize(t2)
    # replace URL
    t4 = re.sub('(http|https):\/\/\/\S+', ' ',t3)
    # replace hastag
    t5 = re.sub('#+', ' ',t4)
    # singkatan
    t6 = contractions.fix(t5)
    return t8
  
```

Tahap selanjutnya *case folding* untuk mengubah huruf *uppercase* menjadi *lowercase* dengan menggunakan `df["Text"] = df["Text"].str.lower()`. Selanjutnya *tokenizing* , yaitu potongan kalimat mejadi perkata , yaitu potongan kalimat menjadi per-kata dan akan diperoleh frekuensi kata yang digunakan dalam kalimat tersebut. Dengan menggunakan library ntlk, string dan re.

```

def word_tokenize_wrapper(text):
    return word_tokenize(text)
df['tokenizing'] = df['Text'].apply(word_tokenize_wrapper)
df.head(10)
  
```

Selanjutnya tahap *stopword removal* yaitu , menghilangkan kata yang tidak penting sehingga menjadi kalimat yang lebih ringkas

```

def stopwords_removal(words):
    return [word for word in words if word not in list_stopwords]
df['stopword_removal'] = df['tokenizing'].apply(stopwords_removal)
print(df['stopword_removal'].head())
df.head()
  
```

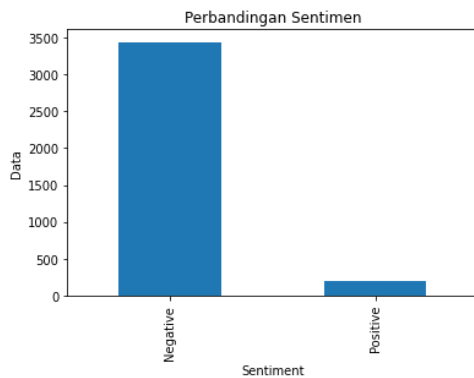
Tabel 3. Dataset Pre-processing

Text (Case Folding)	Tokenizing	Frekuensi Distribusi Tokenizing	Stopword
kinerja investasi tetap meningkat ditengah pandemi covid	['kinerja', 'investasi', 'tetap', 'meningkat', , 'ditengah', 'pandemi', 'covid']	<FreqDist with 7 samples and 7 outcomes >	['kinerja', 'investasi', 'meningkat', , 'ditengah', 'pandemi', 'covid']
['goto', 'perusahaa n', 'domestik', 'kebangga n', 'nasional', 'mampu', 'menarik', 'investasi', 'luar', 'biasa', 'dari', 'luar', 'negeri', 'orang', 'luar', 'neger']	['goto', 'perusahaa n', 'domestik', 'kebangga n', 'nasional', 'mampu', 'menarik', 'investasi', 'luar', 'biasa', 'dari', 'luar', 'negeri', 'orang', 'luar', 'neger']	<FreqDist with 14 samples and 16 outcomes >	['goto', 'perusahaa n', 'domestik', 'kebangga n', 'nasional', 'menarik', 'investasi', 'negeri', 'orang', 'neger']
investasi sektor industri pada triwulan naik secara year on year semangat dan selalu yakin indonesia	['investasi', 'sektor', 'industri', 'pada', 'triwulan', 'naik', 'secara', 'year', 'on', 'year', 'semangat', 'dan', 'selalu', 'yakin', 'indonesia']	<FreqDist with 14 samples and 15 outcomes >	['investasi', 'sektor', 'industri', 'triwulan', 'year', 'on', 'year', 'semangat', 'indonesia']

Pada tabel 3 dari data yang sudah dipre-processing. Pada *tokenizing* kalimat akan dipecah menjadi satu suku kata dan akan diperoleh frekuensi dari kata yang digunakan dalam *text*. Pada kalimat pertama ada 7 sampel kata dan kata yang dihitung ada 7. Pada kalimat ke dua ada 14 sample kata dan kata yang dihitung ada 16 karena ada kata berulang. Sedangkan kalimat terakhir ada 14 sample kata dan kata yang dihitung ada 15 kata karena ada kata yang digunakan lebih dari 1 kali. Setelah itu kalimat melalui proses *stopword* yaitu penghapusan kata yang tidak memiliki makna sehingga kalimat yang dihasilkan jauh lebih ringkas.

Labeling Text

Setelah melalui tahap *pre-processing text* selanjutnya dataset akan melalui tahap pelebelan dimana setiap kalimat akan di beri bobot yang nantinya akan menghasilkan value berupa positif, dan negatif. *Labeling text* ini akan menjadi perbandingan apakah dari hasil labeling ini sentiment yang diberikan akan diproses dengan menggunakan metode SVM dan KNN.



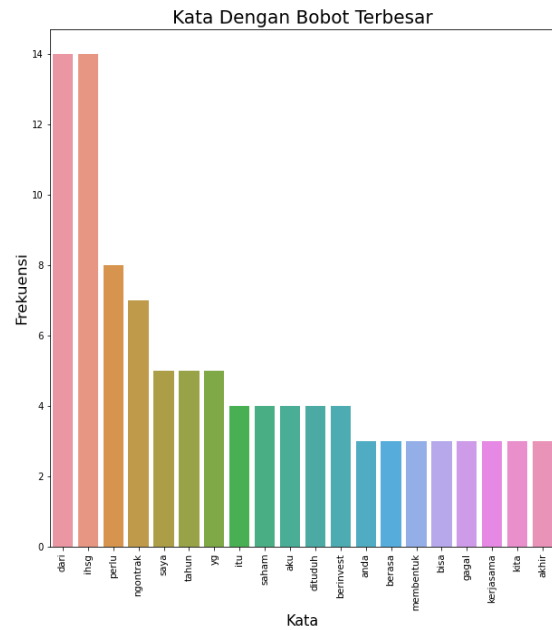
Gambar 3. Hasil Labeling

Pada gambar 3 diatas hasil dari labeling dengan menggunakan bahasa python dari 3639 data menghasilkan sentiment positive sebanyak 202 data dan negative 3437 dari labeling ini akan diproses menggunakan metode SVM dan KNN untuk diprediksi apakah sesuai atau tidak sehingga akan

terjadi perubahan sentiment dan persentase perkelasnya.

TF – IDF

Sebelum pembobotan kata TF-IDF, data akan di vectorisasi terlebih dahulu untuk mengubah *text* menjadi numerik, pada tahap ini menggunakan python dengan library *scikit-learn*.



Gambar 4 . Hasil TF – IDF

Pada gambar 4 hasil dari pembobotan TF – IDF didapat kata “ahingga” mendapat frekuensi sebanyak 137 dan “bank” mendapat frekuensi sebanyak 86 . Dan kata yang dengan frekuensi sedikit yaitu “saham” dengan 18 frekuensi. Dari hasil TF-IDF ini bisa melihat bagaimana suatu kata dapat mempengaruhi hasil dari sentiment yang didapat dari suatu data.

Evaluasi

Setelah melalui tahap labeling dan pembobotan kata, selanjutnya adalah membandingkan metode SVM dan KNN dari outputl evaluasi tersebut akan menghasilkan prediksi sentiment dan seberapa akurat metode dalam memprediksi sentiment tersebut. Data yang diuji dibagi menjadi 2 bagian dengan perbandingan

80:20 data training 80 dan data testing 20. Data dilakukan fitting metode terhadap data train lalu metode tersebut akan diukur keakuratannya dengan menghitung akurasinya pada data testing.

Hasil Support Vector Machine

Hasil analisis dengan menggunakan metode ini menggunakan Python dan menghasilkan laporan klasifikasi yang berisi *recall*, *fi-score* dan *precision* [12] :

	precision	recall	f1-score	support
Negative	0.97	1.00	0.98	688
Positive	1.00	0.42	0.60	40
accuracy			0.97	728
macro avg	0.98	0.71	0.79	728
weighted avg	0.97	0.97	0.96	728

Akurasi dari SVM : 0.9684065934065934

Gambar. 5 Hasil Metode SVM

Pada gambar 5 akurasi yang diperoleh dengan metode ini mendapat nilai 0.9684 atau 96,84% dengan perhitungan menggunakan nilai *true* dari prediksi yang sesuai yang digambarkan pada tabel 4.

Tabel 4. Prediksi dan Accuracy SVM

	Prediksi	
	Negative	Positive
Accuracy	Negative 688	Positive 23
	Positive 0	17

Pada tabel 4 prediksi negative yang sesuai ada 688 data, dan positive yang sesuai ada

17 data. Sehingga perhitungan untuk mendapatkan nilai akurasi sebagai berikut :

$$\begin{aligned} \text{Akurasi} &= \frac{\text{True positif} + \text{True negatif}}{\text{total data matrix}} \times 100 \quad (1) \\ &= \frac{688 + 17}{728} \times 100 \\ &= \frac{705}{728} \times 100 = 96,84 \end{aligned}$$

Untuk perhitungan nilai precision , recall dan fi – score dapat dijelaskan sebagai berikut :

$$\text{Precision} = \frac{\text{True positif}}{\text{True positif} + \text{False Positif}} \quad (2)$$

$$\text{Negative} = \frac{688}{688+23} = 0,97$$

$$\text{Positive} = \frac{17}{17+0} = 1,00$$

$$\text{Recall} = \frac{\text{True positif}}{\text{True positif} + \text{False negatif}}$$

$$\text{Negative} = \frac{688}{688+0} = 1,00$$

$$\text{Positive} = \frac{17}{17+23} = 0,42$$

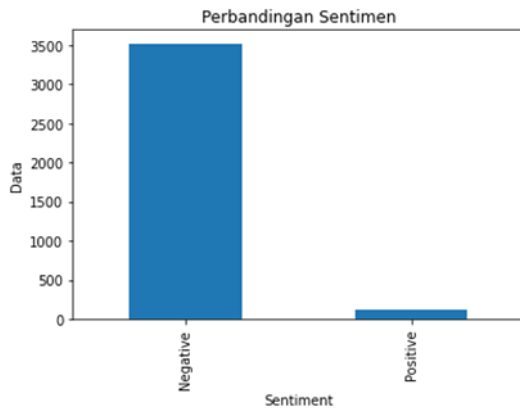
$$\text{Fi – Score} = 2 \times \left(\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \right) \quad (3)$$

$$\text{Negative} = 2 \times \left(\frac{0,97 \times 1,00}{0,97 + 1,00} \right)$$

$$2 \times \left(\frac{0,97}{1,97} \right) = 0,98$$

$$\text{Positive} = 2 \times \left(\frac{1,00 \times 0,42}{1,00 + 0,42} \right)$$

$$2 \times \left(\frac{0,42}{1,42} \right) = 0,6$$



Gambar.6 Perbandingan Sentimen SVM

Pada gambar 6 bisa dilihat kelas negative yang awalnya diprediksi sebanyak 3437 data pada metode SVM terdapat 3521 data yang negative, kelas positive yang awalnya diprediksi 202 data menjadi 118 data, dengan akurasi yang didapat sebanyak 96,84%. data yang diuji menggunakan data training dan testing dengan perbandingan 80:20

i. Analisis K – Nearest Neighbors

Hasil analisis dengan menggunakan metode ini menggunakan Python dan menghasilkan laporan klasifikasi yang berisi *recall*, *fi- score* dan *precision* [12] :

$$\text{Akurasi} = \frac{\text{True positif} + \text{True negatif}}{\text{total data matrix}} \times 100 \quad (4)$$

$$= \frac{688 + 9}{728} \times 100$$

$$= \frac{697}{728} \times 100 = 95,74$$

Untuk perhitungan nilai precision , recall dan fi – score dapat dijelaskan sebagai berikut:

$$\text{Recall} = \frac{\text{True positif}}{\text{True positif} + \text{False negatif}} \quad (5)$$

$$\text{Negative} = \frac{688}{688+0} = 1,00$$

$$\text{Positive} = \frac{9}{9+31} = 0,23$$

	precision	recall	f1-score	support
Negative	0.96	1.00	0.98	688
Positive	1.00	0.23	0.37	40
accuracy			0.96	728
macro avg	0.98	0.61	0.67	728
weighted avg	0.96	0.96	0.94	728

Akurasi dari KNN : 0.9574175824175825

Gambar 7. Hasil Metode KNN

Pada gambar 7 akurasi yang diperoleh dengan metode ini mendapat nilai 0.9574 atau 95,74% dengan perhitungan menggunakan nilai true dari prediksi yang sesuai yang digambarkan pada tabel 5.

Tabel 5. Prediksi dan Accuracy KNN

Prediksi		Negative	Positiv e
Accuracy	Negative	688	31
	Positive	0	9

Pada tabel 5 prediksi negatif yang sesuai ada 688 data, dan positif yang sesuai ada 9 data. Sehingga perhitungan untuk mendapatkan nilai akurasi sebagai berikut :

$$\text{Precision} = \frac{\text{True positif}}{\text{True positif} + \text{False Positif}} \quad (6)$$

$$\text{Negative} = \frac{688}{688+31} = 0,96$$

$$\text{Positive} = \frac{9}{9+0} = 1,00$$

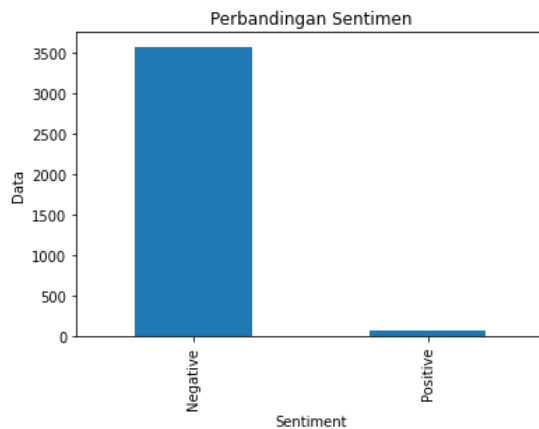
$$\text{Fi – Score} = 2 \times \left(\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \right) \quad (7)$$

$$\text{Negative} = 2 \times \left(\frac{0,96 \times 1,00}{0,96+1,00} \right)$$

$$2 \times \left(\frac{0,96}{1,96} \right) = 0,98$$

$$\text{Positive} = 2 \times \left(\frac{1,00 \times 0,23}{1,00+0,23} \right)$$

$$2 \times \left(\frac{0,23}{1,23} \right) = 0,37$$



Gambar 8. Perbandingan Metode KNN

Pada gambar 8 dengan menggunakan data *training* dan *testing* dengan perbandingan 80:20 bisa dilihat kelas negative yang awalnya diprediksi sebanyak 3437 data pada metode KNN terdapat 3572 data yang negative dengan selisih 135 data, kelas positive yang awalnya diprediksi 202 data menjadi 67 data, dengan perkelas dengan akurasi yang didapat sebanyak 95,74%.

4. SIMPULAN

Analisis sentiment kepuasan investasi dimasa pandemi yang dihasilkan dari pengumpulan data dari 3 sumber yang dijadikan dataset melalui tahap *crawling*, *filtering*, *pre-processing*, *labeling*, TD-IDF dan metode yang digunakan yaitu SVM dan KNN berdasarkan hasil dan pembahasan diatas, dengan menggunakan data training dan testing dengan perbandingan 80:20 dapat disimpulkan bahwa sentiment investasi dimasa pandemi dari 2 metode yang dipakai mendapat respon negatif lebih banyak dibandingkan positif. Bisa disimpulkan bahwa berinvestasi dimasa pandemi mendapat respon negatif, karena dampak yang timbulkan akibat pandemi ini cukup berpengaruh terhadap harga saham, sehingga keuntungan yang

didapat kurang memuaskan. Dari 2 metode yang digunakan diperoleh :

- Metode SVM terdapat 3521 data yang negative dan positive 118 data. Dengan akurasi yang didapat sebanyak 96.84 %.
- Metode KNN terdapat 3572 data negative dan 67 data positif. Dengan akurasi yang didapat 95,74%
- Dari kedua metode yang digunakan metode SVM lebih baik dibandingkan dengan metode KNN dengan selisih 1.10%.

DAFTAR PUSTAKA

- [1] D. Valle-Cruz, V. Fernandez-Cortez, A. López-Chau, and R. Sandoval-Almazán, "Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study Of The H1n1 And The Covid-19 Periods" *Cognitive Computation*, Vol. 14, No. 1, pp. 372–387, January 2021
- [2] A. P. Firdaus, "Analisis Pengaruh Sentimen Investor Terhadap Return Saham Sektoral Bei Di Masa Pandemi Covid-19", *Khazanah Intelektual* Vol. 5 No. 2, pp. 1107–1127, Agustus 2021.
- [3] E. Bulut, S. Akbulut Bekar, And G. Çizgici Akyüz, "Media Coverage Related Investor Sentiment During The Covid-19 Pandemic" *Ege Academic Review*, pp. 357–372, 2021.
- [4] T. Mezghani, M. Boujelbène, And M. Elbayar, "Impact Of Covid-19 Pandemic On Risk Transmission Between Googling Investor's Sentiment, The Chinese Stock And Bond Markets" *China Financ Review International*, Vol. 11, No. 3, pp. 322–348, June 2021.

- [5] B. Jiang, H. Zhu, J. Zhang, C. Yan, And R. Shen, "Investor Sentiment And Stock Returns During The Covid-19 Pandemic" *frontiers In psychology*, Vol. 12, pp. 1–5, July 2021.
- [6] A. Suci, R. Tami, And W. Maharani, "Kecendrungan Sentimen Dengan Pendekatan Support Vector Machine Pada Komunitas Yang Berpengaruh Di Twitter", *E-Proceeding of Engineering*, Vol. 7, No. 3, pp. 9822–9834, Desember 2020.
- [7] D. Muhidin And A. Wibowo, "Perbandingan Kinerja Algoritma Support Vector Machine Dan K-Nearest Neighbor Terhadap Analisis Sentimen Kebijakan New Normal," *String (Satuan Tulisan Riset Dan Inovasi Teknologi)*, Vol. 5, No. 2, pp. 153, Desember 2020.
- [8] F. Sodik And I. Kharisudin, "Analisis Sentimen Dengan SVM , Naive Bayes Dan KNN Untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 Pada Media Sosial Twitter," *Prisma*, Vol. 4, pp. 628–634, 2021.
- [9] E. Dwi Nurindah Sari, "Analisis Sentimen Nasabah Pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner , Naïve Bayes Classifier (NBC), Dan Support Vector Machine (SVM)", *Jurnal Sains Dan Seni ITS*, Vol. 8, No. 2, pp. 177, 2019.
- [10] R. Kosasih And A. Alberto, "Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor", *Jurnal Nasional Informatika dan Jaringan*, Vol. 6, No.1, September 2021.
- [11] R. Tinages, A. Triayudi, And I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)", *Jurnal Media Informatika Budidarma*, Vol. 4, No. 3, pp. 650-658, Juli 2020,
- [12] M. I. Petiwi, A. Triayudi, And I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes Dan Support Vector Machine", *Jurnal Media Informatika Budidarma*, Vol. 6, No. 1, p. 542-550, Januari 2022,