

**ANALISIS PENDAPAT MASYARAKAT TERKAIT INVESTASI  
ONLINE DI ERA DIGITAL MENGGUNAKAN NAIVE BAYES  
CLASSIFIERS**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat  
Sarjana Program Studi Informatika



disusun oleh

**ARYO BIMO SAPUTRO**

**19.11.2621**

Kepada

**FAKULTAS ILMU KOMPUTER UNIVERSITAS  
AMIKOM YOGYAKARTA YOGYAKARTA**

2024

**ANALISIS PENDAPAT MASYARAKAT TEKAIT INVESTASI ONLINE DI  
ERA DIGITAL MENGGUNAKAN NAIVE BAYES CLASSIFIERS**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**ARYO BIMO SAPUTRO**

**19.11.2621**

Kepada

**FAKULTAS ILMU KOMPUTER UNIVERSITAS  
AMIKOM YOGYAKARTA YOGYAKARTA**

**2024**

**PERSETUJUAN**

**SKRIPSI**

**ANALISIS PENDAPAT MASYARAKAT TEKAIT INVESTASI ONLINE DI ERA  
DIGITAL MENGGUNAKAN NAIVE BAYES CLASSIFIERS**

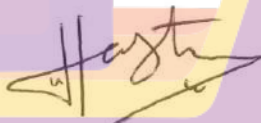
yang disusun dan diajukan oleh

**ARYO BIMO SAPUTRO**

**19.11.2621**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 24 September 2024

**Dosen Pembimbing**



**Hastari Utama, M.Cs**

**NIK. 190302230**

**PENGESAHAN**

**SKRIPSI**

**ANALISIS PENDAPAT MASYARAKAT TEKAIT INVESTASI ONLINE DI  
ERA DIGITAL MENGGUNAKAN NAIVE BAYES CLASSIFIERS**

yang disusun dan diajukan oleh

**ARYO BIMO SAPUTRO**

19.11.2621

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 2 Juli 2024

**Susunan Dewan Penguji**

**Nama Penguji**

Hastari Utama, M.Cs  
NIK. 190302230

Hendra Kurniawan, M.Kom  
NIK. 190302244

Ferian Fauzi Abdulloh, M.Kom  
NIK. 190302276

**Tanda Tangan**



Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 24 September 2024

**DEKAN FAKULTAS ILMU KOMPUTER**



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.  
NIK. 190302096

## PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

**Nama mahasiswa : Aryo Bimo Saputro**  
**NIM : 19.11.2621**

Menyatakan bahwa Skripsi dengan judul berikut:

### **ANALISIS PENDAPAT MASYARAKAT TERKAIT INVESTASI ONLINE DI ERA DIGITAL MENGGUNAKAN NAÏVE BAYES CLASSIFIERS**

Dosen Pembimbing : Hastari Utama, M.Cs

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 24-09-2024

Yang Menyatakan,



Aryo Bimo Saputro

## HALAMAN PERSEMBAHAN

Alhamdulillah, Segala Puji Bagi ALLAH SWT. Kepada-Nya kita memohon, memuji dan meminta perlindungan, serta petunjuk Kepada-Nya. Kita berlindung kepada ALLAH SWT dari kejahatan diri kita dan keburukan amal-amalkita. Dengan ini penulis mempersembahkan skripsi ini kepada semua pihak yang telah membantu dan memberikan motivasi serta dukungan yang besar sehingga skripsi ini dapat diselesaikan dengan sebaik baiknya.

Melalui skripsi ini saya persembahkan terimakasih serta syukur kepada:

1. Rasa syukur kepada Allah SWT yang telah memberikan rahmat serta karunia untuk terus mengucap syukur dan sabar dalam menyelesaikan tugasakhir skripsi.
2. Kepada keluarga penulis Bapak Bowo Listyo Novi Hastono, Ibu Heru Marsetiani yang selalu memberikan doa,cinta dan dukungan yang tidak pernah berhenti diberikan kepada penulis.
3. Hastari Utama, M.Cs selaku dosen pembimbing yang selalu memberikan saran dan masukan kepad penulis dalam menyelesaikan skripsiini.

## KATA PENGANTAR

Puji syukur kehadirat ALLAH SWT karena berkat rahmat serta hidayah- Nya penulis diberikan kesehatan dan kekuatan serta iman yang kuat sehingga dapat menyelesaikan skripsi yang berjudul **“ANALISIS PENDAPAT MASYARAKAT TERKAIT INVESTASI ONLINE DI ERA DIGITAL MENGGUNAKAN NAÏVE BAYES CLASSIFIERS”**. Oleh karena itu penulis mengucapkan syukur karena dapat menyelesaikan skripsi ini. Skripsi ini diajukan kepada Program Studi S1 Informatika, Fakultas Ilmu Komputer, serta Universitas AMIKOM Yogyakarta.

Penulis dalam menyelesaikan skripsi mendapatkan beberapa hambatan dalam berbagai hal, namun banyak pihak yang selalu membantu sehingga skripsi ini dapat diselesaikan dengan baik. Dengan begitu penulis hendak mengungkapkan terima kasih terhadap:

1. Bapak prof. Dr. M Suyanto, M.M selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradya D, M.Kom., selaku Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
4. Hastari Utama, M.Cs selaku dosen pembimbing yang selalu memberikan saran dan arahan dalam proses penulisan skripsi ini.

Yogyakarta, < 24 September 2024 >

Penulis

## DAFTAR ISI

SKRIPSI.....	i
PERSETUJUAN.....	ii
PENGESAHAN.....	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xiii
INTISARI.....	xiv
ABSTRAK.....	xv
BAB I.....	1
PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan masalah.....	4
1.3. Batasan Masalah.....	4
1.4. Tujuan penelitian.....	5
1.5. Manfaat Penelitian.....	5
1.6. Metode Penelitian.....	6
1.6.1. Studi Literatur.....	6
1.6.2. Pengumpulan Data.....	6
1.6.3. <i>Data-Preprocessing</i> .....	6
1.6.4. Metode Algoritma.....	6
1.6.5. Hasil dan Pembahasan.....	6
BAB II.....	8
TINJAUAN PUSTAKA.....	8
2.1 Tinjauan Pustaka.....	8
2.2 Landasan Teori.....	14



2.2.1.	<i>Data Mining</i> .....	14
2.2.2.	<i>Text Mining</i> .....	14
2.2.3.	Analisis Sentimen .....	15
2.2.4.	<i>Machine learning</i> .....	16
2.2.5.	Investasi .....	16
2.2.6.	Twitter .....	16
2.2.6.1.	<i>Crawling Data Twitter</i> .....	18
2.2.7.	<i>Pre-processing</i> .....	18
2.2.7.1.	<i>Cleaning Data</i> .....	18
2.2.7.2.	Case Folding .....	18
2.2.7.3.	Tokenizing .....	19
2.2.7.4.	Normalization.....	19
2.2.7.5.	<i>Stopword Removal</i> .....	19
2.2.7.6.	Stemming .....	19
2.2.8.	Labeling Data.....	20
2.2.9.	Ekstrasi Fitur dan Pembobotan Kata.....	20
2.2.9.1.	Term <i>Frequency</i> (TF).....	21
2.2.9.2.	<i>Normalization Term Frequency</i> .....	21
2.2.9.3.	Inverse Document <i>Frequency</i> (IDF).....	21
2.2.9.4.	Term <i>Frequency</i> – Inverse Document <i>Frequency</i> (TF-IDF) .....	21
2.2.10.	Multinomial Naïve Bayes Classifier .....	21
2.2.11.	K-Fold Cross Validation .....	23
2.2.12.	Evaluasi dan Visualisasi.....	23
2.2.12.1.	WordCloud.....	23
2.2.12.2.	Confusion Matrix .....	23
BAB III	.....	26
METODE PENELITIAN	.....	26

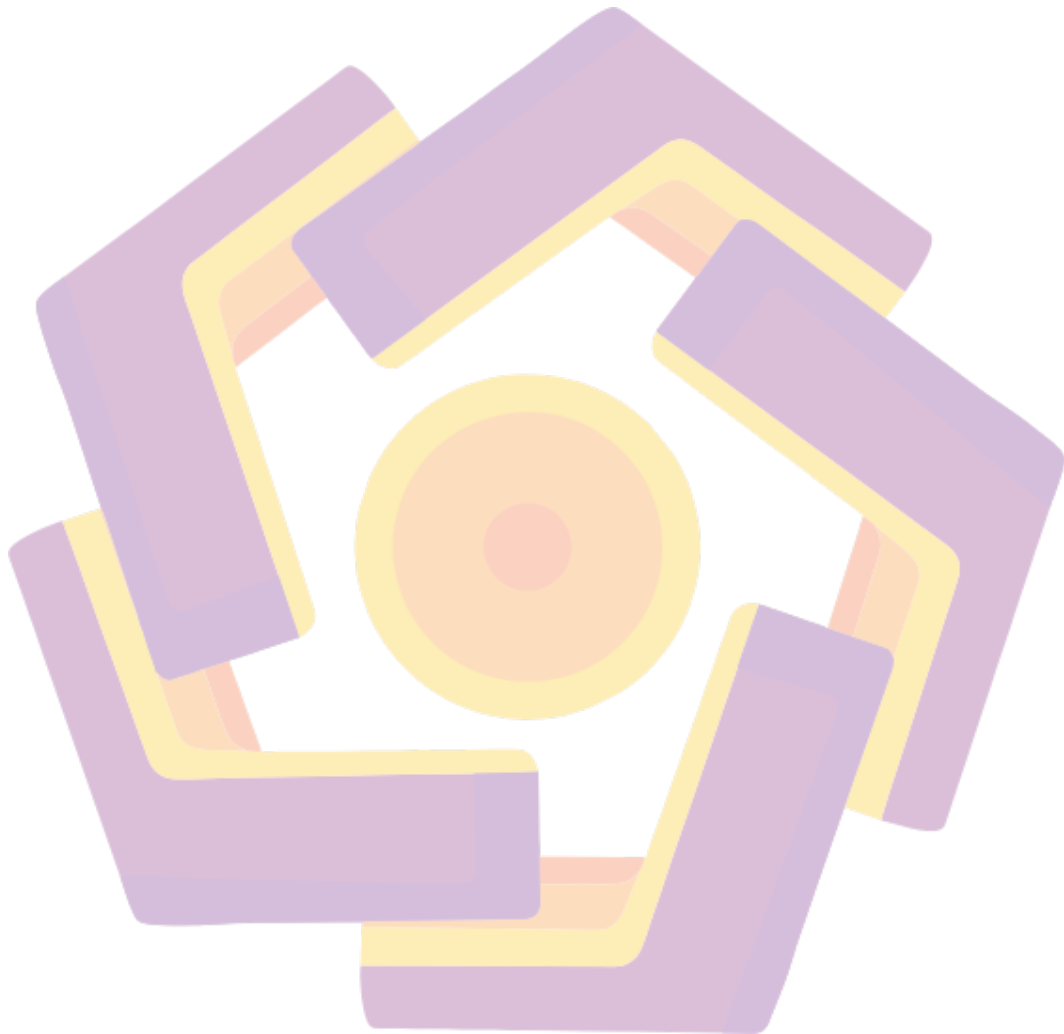
3.1. Object Penelitian .....	26
3.2. Alur Metode Penelitian.....	26
3.3. Analisis Kebutuhan Perangkat.....	28
3.3.1. <i>Hardware dan Software</i> .....	28
3.4. Uraian Metodologi.....	28
3.4.1. Menentukan Kata Kunci .....	28
3.4.2. <i>Pre-processing</i> .....	29
3.4.2.1. Normalization ASCII .....	29
3.4.2.2. Terjemah .....	29
3.4.2.3. Cleaning .....	30
3.4.2.4. Tokenizing .....	30
3.4.2.5. Normalization Words.....	31
3.4.2.6. StopWords Removal .....	31
3.4.2.7. Stemming .....	32
3.4.2.8. Labeling .....	33
3.4.3. Feature Ekstrasi dan Pembobotan ( Term Weight ) .....	33
3.4.3.1. Term Frequency .....	34
3.4.3.2. Term Normalization.....	35
3.4.3.3. Invers Document Frequency .....	37
3.4.3.4. TF-IDF .....	37
3.4.4. Classification Naïve Bayes Multinomial .....	38
3.4.4.1. Perancangan System Klasifikasi Naïve Bayes Multinomial .....	38
3.4.5. K-Fold Cross Validation .....	41
3.4.6. Evaluasi.....	42
3.4.6.1. <i>WordCloud</i> Visualisasi .....	42
3.4.6.2. <i>Confusion Matrix</i> Visual .....	43
3.4.6.2.1. <i>Accuracy</i> .....	43

3.4.6.2.2.	<i>Precision</i>	43
3.4.6.2.3.	<i>Recall</i>	44
3.4.6.2.4.	<i>F1-score</i>	44
BAB IV		45
HASIL DAN PEMBAHASAN		45
4.1.	Crawling	45
4.2.	Preprocessing	46
4.2.1.	Kode Prose <i>Cleaning</i>	46
4.2.2.	Kode Proses <i>Normalization Unicode</i>	46
4.2.3.	<i>Cleaning</i>	47
4.2.4.	Kode Proses Case Folding	48
4.2.5.	Kode Proses Tokenization	49
4.2.6.	Kode Proses <i>Normalization Word Dictionary</i>	49
4.2.7.	Kode Program Proses <i>StopWords Removal</i>	50
4.2.8.	Kode Program Proses Stemming	51
4.3.	Kode Proses Labeling	52
4.4.	Kode Proses Pembobotan Kata ( <i>Term Weighting</i> )	53
4.5.	Implementasi Classification Naïve Bayes Multinomial	54
4.6.	Kode Implementasi K-Fold Cross Validation	55
4.7.	Evaluasi Performa Model	55
4.7.1.	Evaluasi WordCloud	55
4.7.2.	Performance Model Confusion Matrix	58
BAB V		61
PENUTUP		61
5.1.	KESIMPULAN	61
5.2.	SARAN	61
DAFTAR PUSTAKA		62

## DAFTAR GAMBAR

Gambar 3. 1 Alur Metode Penelitian .....	26
Gambar 3. 2 Token Autentikasi Twitter API .....	28
Gambar 3. 3 Ilustrasi K-Fold Cross Validation .....	42
Gambar 3. 4 Visual WordCloud .....	42
Gambar 4. 1 Install library node.js.....	45
Gambar 4. 2 Tweet Kata Kunci investasi.....	45
Gambar 4. 3 Kode Proses Terjemah tweet bahasa Indonesia .....	46
Gambar 4. 4 Kode Proses Normalization Unicode .....	47
Gambar 4. 5 Kode Proses Cleaning .....	47
Gambar 4. 6 Kode Proses Cleaning White Space .....	48
Gambar 4. 7 Cleaning dan Mengatur Ulang Indexs.....	48
Gambar 4. 8 Kode Proses Case Folding .....	48
Gambar 4. 9 Kode Proses Tokenization.....	49
Gambar 4. 10 Kode Proses Normalization Word Dictionary .....	49
Gambar 4. 11 Kode Proses Removing Bracket Token .....	50
Gambar 4. 12 Kode Proses StopWords Removal .....	51
Gambar 4. 13 Kode Program Proses Stemming .....	52
Gambar 4. 14 Gambar Kode Proses Labeling.....	52
Gambar 4. 15 Kode Program Proses Term Weighting .....	53
Gambar 4. 16 Kode Split Pembagian data .....	54
Gambar 4. 17 Kode Implementasi Classification Multinomial .....	54
Gambar 4. 18 Implementasi K-Fold Cross Validation.....	55
Gambar 4. 19 Jumlah Analisis Sentimen Investasi .....	55
Gambar 4. 20 Presentase Sentimen Investasi.....	56
Gambar 4. 21 Kode Proses WordCloud Count Sentimen Investasi.....	56
Gambar 4. 22 Kode Program Ranking Frekuensi Term WordCloud .....	57
Gambar 4. 23 Hasil Evaluasi WordCloud.....	58
Gambar 4. 24 Score Akurasi Non Pembagian Score Threshold Label .....	58
Gambar 4. 25 Cross Validation Non Threshold Label.....	59

Gambar 4. 26 Score Akurasi dari threshold label 65 .....59  
Gambar 4. 27 Cross Validation menggunakan score labeling 0.65 .....59  
Gambar 4. 28 Score Model Classification data 70.....60  
Gambar 4. 29 Cross Validation menggunakan score labeling 0.70 .....60



## DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka.....	10
Tabel 3. 1 Preprocessing Normalization ASCII.....	29
Tabel 3. 2 Preprocessing Terjemah.....	30
Tabel 3. 3 Preprocessing Cleaning.....	30
Tabel 3. 4 Preprocessing Tokenizing.....	31
Tabel 3. 5 Preprocessing Normalization Words.....	31
Tabel 3. 6 Preprocessing StopWords Removal.....	32
Tabel 3. 7 Preprocessing Stemming.....	32
Tabel 3. 8 Preprocessing Labeling.....	33
Tabel 3. 9 Term Frequency.....	34
Tabel 3. 10 Term Normalization Frequency.....	35
Tabel 3. 11 Total dokumen mengandung kata i.....	36
Tabel 3. 12 Invers Dokument Frequency (IDF).....	37
Tabel 3. 13 TF-IDF Term Weighting.....	38
Tabel 3. 14 Tweet Data Train.....	38
Tabel 3. 15 Prior probabilitas.....	39
Tabel 3. 16 Data Test.....	40
Tabel 3. 17 Visual Confusion Matrix.....	43

## INTISARI

Era digital yang terus berkembang menuntut individu untuk terus belajar, terutama dalam hal investasi. Perkembangan teknologi yang mudah diakses oleh berbagai kelompok usia telah membawa dampak signifikan dalam akses informasi, di masa lalu investasi terbatas pada properti dan saham yang memerlukan interaksi langsung dengan broker. Namun, perubahan pengelolaan investasi telah terjadi di era digital, di mana masyarakat dapat mengakses informasi nilai aset secara real-time dan melakukan investasi online dengan mudah. Meskipun demikian, kemudahan ini juga membuka peluang bagi penyalahgunaan investasi.

Peneliti menganalisis mengenai sentimen yang ada pada media sosial twitter terhadap topik investasi dan saham. Proses penelitian dilakukan menggunakan metode naive bayes classifier terhadap data yang telah diambil berdasarkan kata kunci investasi dan saham. Metode Naive bayes classifier menghasilkan akurasi cukup tinggi dengan akurasi 75%. Peneliti melakukan uji coba pada parameter hybrid algoritma dengan model adabost menghasilkan akurasi 71%. Peneliti mencoba dengan mengubah paramter pada ambang tertentu dari hasil proses labeling data dengan ambang batas skor 0.65, naive bayes menghasilkan akurasi 86%. Proses pengujian naik signifikan pada pengujian terhadap kenaikan pengambil skor label diangka skor 0.70 menghasilkan akurasi sebesar 89%.

Kata kunci : Investasi, Saham, Analisis Sentimen, Naive Bayes Classifier, Twitter

## ABSTRAK

The rapidly evolving digital era demands individuals to continuously learn, especially in investment matters. The accessibility of technology across different age groups has significantly impacted information access. In the past, investments were limited to properties and stocks, requiring direct interaction with brokers. However, investment management has transformed in the digital era, where people can access real-time asset information and easily invest online. Despite these conveniences, they also present opportunities for investment misuse.

Researchers analyzed sentiment on Twitter regarding investment and stocks. The research process employed a naive Bayes classifier method on data extracted using keyWords related to investment and stocks. The naive Bayes classifier method achieved a high Accuracy of 75%. The researchers experimented with a hybrid algorithm parameter using an AdaBoost model, resulting in 71% Accuracy. Adjusting parameters with a specific threshold score of 0.65 from the data labeling process, naive Bayes achieved an Accuracy of 86%. Further testing showed significant improvement when increasing the threshold to 0.70, reaching an Accuracy of 89%.

**KeyWords:** Investment, Stocks, Sentiment Analysis, Naive Bayes Classifier, Twitter