

ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP INVESTASI KEUANGAN DI INDONESIA MENGGUNAKAN METODE NAIVE BAYES

Andreas Danny Agus Wahyudi^{1*}, Tinaliah,M.Kom²
Informatika^{1*} Universitas Multi Data Palembang²³ Palembang, Indonesia
Andreas123@mhs.mdp.ac.id ^{1*} tinaliah@mdp.ac.id²

Keywords

Investment 1
Twitter 2
Naive Bayes 3
SMOTE 4

Abstrak

Abstract: Financial investment is important for Indonesian people to prepare for their future financially. In this digital era, social media such as Twitter have become popular platforms for sharing opinions and views on various topics including financial investments. By leveraging the data available on Twitter, sentiment analysis can be used to understand user views and opinions regarding financial investments in Indonesia. The Naive Bayes method can be used to perform sentiment analysis on Twitter data by utilizing probability theory to classify tweets with positive, negative or neutral views about financial investment in Indonesia. The amount of tweet data is unbalanced, so it is necessary to do SMOTE over-sampling so that the dataset is balanced and do the testing using k-fold validation so that you can see the confision matrix and get the values for accuracy, precision, recall, and f1-score. Based on the sentiments obtained from Twitter social media, it shows that Twitter social media users have positive sentiments towards financial investment in Indonesia with a total number of positive sentiments of 426 data from a total of 1000 tweet data. Unbalanced data affects the classification results, namely an accuracy of 45% with the SMOTE up-sampling method and an accuracy of 89% without using the SMOTE up-sampling method.

Investasi 1
Twitter 2
Naive Bayes 3
SMOTE 4

Abstrak: Investasi keuangan merupakan hal yang penting bagi masyarakat Indonesia untuk mempersiapkan masa depan mereka secara finansial. Dalam era digital ini, media sosial seperti Twitter telah menjadi platform yang populer untuk berbagi opini dan pandangan tentang berbagai topik termasuk investasi keuangan. Dengan memanfaatkan data yang tersedia di Twitter, analisis sentimen dapat digunakan untuk memahami pandangan dan opini pengguna terkait investasi keuangan di Indonesia. Metode Naive Bayes dapat digunakan untuk melakukan analisis sentimen pada data Twitter dengan memanfaatkan teori probabilitas untuk mengklasifikasikan tweet yang berpandangan positif, negatif, atau netral tentang investasi keuangan di Indonesia. Jumlah data tweet tidak seimbang, sehingga perlu melakukan SMOTE over-sampling sehingga dataset tersebut seimbang dan melakukan pengujian menggunakan k-fold validation sehingga bisa melihat confision matrix serta mendapatkan nilai akurasi, precision, recall, dan f1-score. Berdasarkan dari sentimen yang didapat dari media sosial Twitter menunjukkan bahwa pengguna media sosial Twitter memiliki sentimen positif terhadap investasi keuangan di Indonesia dengan total jumlah sentimen positif sebanyak 426 data dari total 1000 data tweet. Data yang tidak seimbang mempengaruhi hasil klasifikasi yaitu akurasi sebesar

PENDAHULUAN

Investasi keuangan penting bagi masyarakat Indonesia untuk mempersiapkan masa depan. Namun, berbagai faktor seperti pandemi Covid-19, kondisi ekonomi global, dan kebijakan pemerintah memengaruhi keputusan investasi. Di era digital, Twitter menjadi platform populer untuk berbagi opini tentang investasi, sehingga analisis sentimen dapat digunakan untuk memahami pandangan masyarakat.

Fenomena investasi bodong menjadi masalah serius di Indonesia, ditandai dengan janji keuntungan besar, modal rendah, promosi di media sosial, dan kurangnya regulasi yang sah. Beberapa algoritma analisis sentimen pada data Twitter adalah Naive Bayes, Logistic Regression, SVM, dan ANN. Setiap algoritma memiliki keunggulan dan tantangan, dengan kinerja yang diukur menggunakan metrik seperti akurasi, presisi, recall, dan F1-score.

Penelitian menunjukkan Naive Bayes unggul dalam data kecil atau tidak seimbang, meski hasilnya bergantung pada kualitas data pelatihan dan fitur yang digunakan. Studi sebelumnya menunjukkan keberhasilan metode ini dalam analisis sentimen, termasuk topik saham dan investasi di Twitter, dengan akurasi berkisar antara 70-82%.

Analisis sentimen pada Twitter dapat membantu investor memahami opini masyarakat terkait investasi keuangan di Indonesia, memberikan wawasan penting dalam pengambilan keputusan. Penelitian ini akan menggunakan metode Naive Bayes untuk menganalisis sentimen pengguna Twitter terkait investasi di Indonesia. Oleh karena itu pada penelitian ini akan dilakukan analisis sentimen pengguna twitter tentang investasi keuangan di Indonesia dengan menggunakan metode *naive bayes*.

1. METODE PENELITIAN

Investasi dan Klasifikasi Data

Investasi adalah aktivitas membeli aset seperti saham, obligasi, atau properti untuk keuntungan di masa depan. Ada dua jenis utama: berbasis utang (contoh: obligasi, deposito) dan berbasis ekuitas (contoh: saham). Investasi berbasis ekuitas memiliki risiko lebih tinggi, sehingga analisis fundamental dan teknikal diperlukan sebelum mengambil keputusan. Fundamental menilai kinerja keuangan perusahaan, sedangkan teknikal menganalisis pola harga saham. Diversifikasi portofolio juga penting untuk mengelola risiko.

Klasifikasi Data dan Metode Pendukung

Klasifikasi adalah proses mengelompokkan data ke dalam kategori tertentu, misalnya sentimen positif atau negatif. Dalam analisis sentimen, ketidakseimbangan kelas sering terjadi, diatasi dengan metode seperti *SMOTE* yang membuat sampel sintetis untuk kelas minoritas.

Naive Bayes

Naive Bayes adalah metode klasifikasi berbasis probabilitas yang memprediksi peluang berdasarkan data sebelumnya. Keunggulannya meliputi kecepatan, efisiensi pada data kecil, dan kesederhanaan. Namun, metode ini mengasumsikan independensi variabel, yang kadang tidak sesuai dengan kenyataan.

Pendekatan Evaluasi

Cross-validation mengevaluasi akurasi model dengan membagi data menjadi beberapa subset untuk pelatihan dan pengujian secara bergantian. Teknik ini membantu menemukan parameter terbaik untuk model yang akurat.

Preprocessing Data

Teknik *preprocessing* meliputi:

Tokenisasi: Memecah teks menjadi kata atau token.

Stopword Removal: Menghapus kata tidak penting seperti "dan" atau "di".

Stemming: Mengubah kata berimbuhan menjadi kata dasar.

TF-IDF: Memberi bobot kata berdasarkan frekuensi dan keberadaannya dalam dokumen.

Google Colab

Google Colab adalah platform berbasis *cloud* untuk menjalankan *Python* tanpa instalasi. Cocok untuk analisis data karena fleksibilitasnya dalam berbagi dan menyimpan program.

2. HASIL DAN PEMBAHASAN

2.1 Hasil

2.1.1 Case Folding

Case Folding adalah tahapan awal dari proses *Preprocessing* data. Dalam tahapan ini dilakukan transformasi atau mengubah huruf pada data teks *tweet* dari huruf kapital menjadi huruf kecil atau *lower case*. Adapun hasil dari proses ini, terhadap data yang telah di labeli oleh pakar dapat dilihat pada Tabel 2.1.1

Tabel 2.1.1 Hasil Case Folding

Sebelum	Sesudah
Pilihan buku banyak setarus empat puluh buku yang ada di koleksi Ringkas terdiri dari banyak tema dan topik sesuai bidang yang kamu mau. Ada kesehatan politik karier bisnis mindfulness sejarah keuangan dan investasi serta masih banyak lagi	pilihan buku banyak setarus empat puluh buku yang ada di koleksi ringkas terdiri dari banyak tema dan topik sesuai bidang yang kamu mau. ada kesehatan politik karier bisnis mindfulness sejarah keuangan dan investasi serta masih banyak lagi

Dapat dilihat pada tabel tersebut bahwa huruf kapital di ubah menjadi huruf kecil atau *lower case*.

2.1.2 Tokenisasi

Tokenisasi merupakan proses yang dilakukan setelah data teks melalui tahapan *Case Folding*, tokenisasi dilakukan untuk memisahkan antara kata pada kalimat dan menghilangkan tanda baca, angka dan *emoticon*. Hasil yang didapat dari proses tokenisasi yang dimana data inputan yang digunakan untuk proses ini adalah data yang berasal dari *Case Folding*, hasilnya dapat dilihat pada tabel 2.1.2

Tabel 2.1.2 Hasil Tokenisasi

Sebelum	Sesudah
pilihan buku banyak setarus empat puluh buku yang ada di koleksi ringkas terdiri dari banyak tema dan topik sesuai bidang yang kamu mau. ada kesehatan politik karie bisnis mindfulness sejarah keuangan dan investasi serta masih banyak lagi	pilihan buku banyak setarus empat puluh buku yang ada di koleksi ringkas terdiri dari banyak tema dan topik sesuai bidang yang kamu mau ada kesehatan politik karier bisnis mindfulness sejarah keuangan dan investasi serta masih banyak lagi

2.1.3 Stopword

Stopword adalah proses yang dilakukan setelah proses *Case Folding* dan Tokenisasi, dimana *Stopword* bertujuan untuk menghilangkan kata – kata yang sering digunakan atau sering muncul dan juga tidak dibutuhkan dalam proses analisis. Pada penelitian ini proses *Stopword* sendiri dilakukan dengan bantuan menggunakan modul NLTK yang disediakan dalam bahasa pemrograman *Pyhon*. Hasil yang didapat dalam tahap *Stopword* ini dapat dilihat pada Tabel 2.1.3

Tabel 2.1.3 Hasil Stopword

Sebelum	Sesudah
pilihan buku banyak setarus empat puluh buku yang ada di koleksi ringkas terdiri dari banyak tema dan topik sesuai bidang yang kamu mau ada kesehatan politik karier bisnis mindfulness sejarah keuangan dan investasi serta masih banyak lagi	pilihan buku setarus puluh buku koleksi ringkas tema topik sesuai bidang kesehatan politik karier bisnis mindfulness sejarah keuangan investasi

2.1.4 Stemming

Stemming merupakan tahapan terakhir pada *preprocessing*, yang dimana fungsi dari *Stemming* adalah mengembalikan atau mengubah kata yang mengandung imbuhan menjadi kata dasar. Pada penelitian ini, proses *Stemming* dilakukan dengan menggunakan modul Sastrawi yang disediakan bahwa bahasa pemrograman *Pytohn*. Hasil dari *Stemming* dapat dilihat pada Tabel 2.1.4

Tabel 2.1.4 hasil Stemming

Sebelum	Sesudah
pilihan buku setarus puluh buku koleksi ringkas tema topik sesuai bidang Kesehatan politik karier	pilih buku setarus puluh buku koleksi ringkas tema topik sesuai bidang sehat politik karier bisnis mindfulness sejarah uang investasi

bisnis mindfulness sejarah keuangan investasi	
---	--

2.1.5 Pembobotan *TF-IDF*

Pembobotan *TF-IDF* adalah (singkatan dari bahasa Inggris: *term frequency-inverse document frequency*, bahasa Indonesia: frekuensi istilah-inversi frekuensi dokumen) adalah ukuran statistik yang menggambarkan pentingnya suatu istilah terhadap sebuah dokumen dalam sebuah kumpulan atau korpus. Ukuran ini sering dipakai sebagai faktor pembobot dalam pencarian informasi, penambangan teks, dan pemodelan pengguna. Nilai *TF-IDF* bertambah sebanding dengan jumlah kemunculan istilah dalam dan bergantung pada jumlah dokumen dalam korpus yang memiliki istilah tersebut. Hasil *TF-IDF* dapat di lihat pada tabel 3.1.5

Tabel 3.1.5 hasil *TF-IDF*

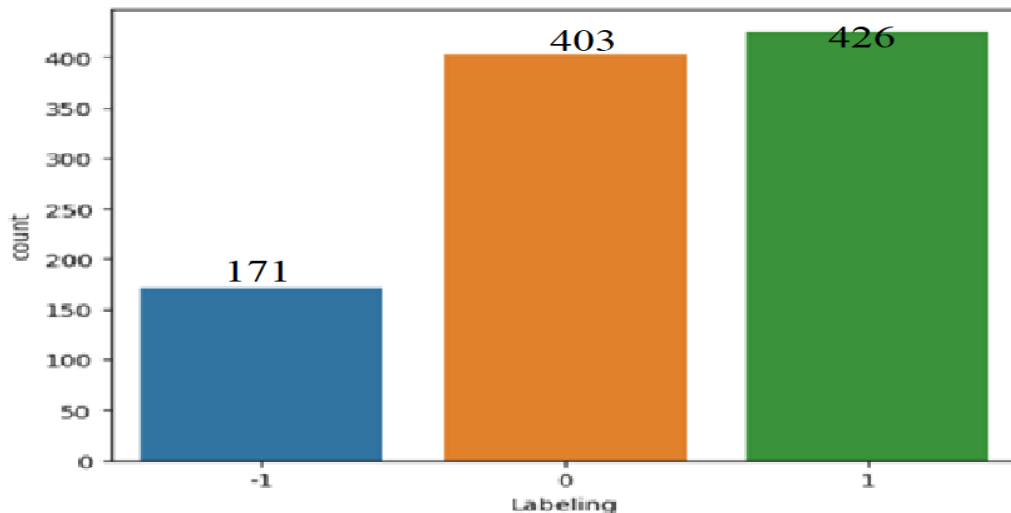
(0, 1251)	0.0421872568819928
(0, 3310)	0.04239889230470282
(0, 2805)	0.25103317210412346
(0, 1986)	0.27139001070042273
(0, 388)	0.13720038800360268
(0, 1447)	0.28754814251326627
(0, 2467)	0.259925633732065
(0, 2802)	0.18533271390460818
(0, 369)	0.14004808891828835
(0, 2863)	0.18689569291377947
(0, 3247)	0.2323031249508638
(0, 3142)	0.2129552226858094
(0, 2692)	0.28754814251326627
(0, 1578)	0.28754814251326627
(0, 2559)	0.28754814251326627
(0, 2865)	0.28754814251326627
(0, 458)	0.37066542780921635
(0, 2423)	0.18533271390460818
(1, 205)	0.2676805483501233
(1, 1548)	0.2676805483501233
(1, 1274)	0.2676805483501233
(1, 2083)	0.2676805483501233
(1, 525)	0.2676805483501233
(1, 1725)	0.2676805483501233
(1, 1058)	0.233688510645839
:	:
(997, 1251)	0.09425095754604038
(997, 3310)	0.09472377428539075
(997, 388)	0.30652071029883443
(998, 2690)	0.36162966361444304
(998, 1996)	0.341308684591437
(998, 1267)	0.2862487624772859
(998, 1046)	0.2921517774321531

(998, 2573)	0.2862487624772859
(998, 320)	0.2921517774321531
(998, 2823)	0.2862487624772859
(998, 903)	0.27183079840914703
(998, 1218)	0.2678192419248721
(998, 245)	0.23308029883372716
(998, 1738)	0.2761916085780394
(998, 1251)	0.05305603222370771
(998, 3310)	0.05332219164332558
(998, 458)	0.23308029883372716
(999, 2011)	0.4718991343646988
(999, 455)	0.3987099444062725
(999, 945)	0.41322844224441696
(999, 2321)	0.27842483909621724
(999, 2017)	0.2967858256556419
(999, 787)	0.3422655776535311
(999, 2813)	0.3987099444062725
(999, 1251)	0.07659163783882571

Berdasarkan representasi data diatas , terdapat 1000 baris (0-999) yang mewakili 1000 dokumen yang berbeda. Setiap dokumen direpresentasikan dalam bentuk vektor dengan berbagai atribut dan fitur yang dihasilkan oleh metode *TF-IDF*. Untuk mengetahui jumlah atribut dan fitur yang ada, hanya perlu melihat jumlah kolom dalam matriks. Matriks yang diberikan adalah hasil dari penerapan metode *TF-IDF* pada teks atau dokumen. Setiap entri dalam matriks tersebut diberikan dalam format (baris, kolom) nilai, di mana kolom mewakili indeks atribut atau fitur dalam matriks. Untuk mengetahui jumlah atribut atau fitur, kita perlu mencari indeks terbesar pada kolom. Dalam contoh tersebut, indeks terbesar adalah 3310, yang berarti terdapat 3311 atribut atau fitur dalam matriks tersebut.

2.1.6 Metode Naive Bayes

Pada tahap klasifikasi data, data yang dimaksud adalah data yang telah melewati proses *Preprocessing* dan Pembobotan *TF-IDF* dibagi menjadi 2 yaitu data latih dan data uji, dimana data uji dan data latih tersebut dibagi dengan perbandingan rasio data sebesar 80% untuk data latih dan 20% untuk data uji. Dapat dilihat pada gambar 2.1.6



Keterangan :

1 : positif

0 : netral

-1 : negatif

Gambar 2.1.6 Model Sentimen Positif, Netral, dan Negatif

2.2 Pembahasan

Tahapan Pembahasan dilakukan dengan hasil dari pengujian terhadap *Tweet* yang telah dilabeli oleh pakar menjadi label positif, negatif, dan netral. Pengujian ini dilakukan dengan membagi data menjadi 80% data latih dan 20% data uji dengan pengujian yaitu data tiga label (positif, netral, dan negatif). Pengujian dilakukan secara keseluruhan dan menggunakan *K-Fold Cross Validation* dengan nilai $k = 10$. Hasil dari pengujian akan dihitung menggunakan *Confusion Matrix* untuk mendapatkan nilai *Accuracy*, *Recal*, *Precision*, dan *F1-Score*.

2.2.1 Skenario Pengujian Pertama (Sebelum Menggunakan SMOTE)

Pada skenario pengujian pertama ini dilakukan pengujian dengan menggunakan data berlabel, dimana data yang digunakan sebanyak 426 data berlabel positif, 171 data berlabel negatif dan 403 data berlabel netral tanpa menggunakan metode *up-sampling* SMOTE.

Labeling	Data Latih	Data uji	Total
Positif	319	107	426
Netral	297	106	403
Negatif	134	37	171

Tabel 2.2.1 Hasil Pengujian Pertama Sebelum melakukan SMOTE

Pada tahap validasi dilakukan dengan menggunakan metode *K-Fold CrossValidation* dengan nilai $k = 10$.

Accuracy	Precision			Recall			F1-Score		
	Positif	Netral	Negatif	Positif	Netral	Negatif	Positif	Netral	Negatif
45%	53%	48%	27%	43%	49%	41%	48%	49%	32%

Tabel 2.2.2 Hasil Nilai Accuracy Sebelum Melakukan SMOTE

Pada tabel 3.7 menunjukkan pada skenario pengujian pertama hasil klasifikasi yang dilakukan oleh metode Naïve Bayes cukup baik ditunjukkan oleh akurasi terbesar yaitu 45% . Selanjutnya, dapat dilihat bahwa untuk sentimen positif kemampuannya untuk membedakan data sentimen positif dengan yang lainnya sudah baik dengan nilai *precision* terbesar yaitu 53% pada *precision*. Kemudian, kemampuannya dalam mengenali data sentimen positif sudah baik dengan nilai *recall* terbesar yaitu 43%. Selain itu untuk akurasi terkecil, dan untuk sentimen negatif memiliki nilai *precision* terkecil yaitu 27%.

2.2.2 Skenario Pengujian Kedua (Menggunakan SMOTE)

Pada skenario pengujian kedua ini dilakukan pengujian dengan menggunakan data berlabel, dimana data yang digunakan sebanyak 333 data berlabel positif, 333 data berlabel negatif dan 333 data berlabel netral dengan menggunakan metode *up-sampling* SMOTE. Pada tahap validasi dilakukan dengan menggunakan metode *K-Fold CrossValidation* dengan nilai $k = 10$.

Tabel 2.2.3 Hasil Pengujian Kedua Setelah Melakukan SMOTE

Labeling	Data Latih			Data uji			Total		
Positif	266			67			333		
Netral	266			67			333		
Negatif	266			67			333		

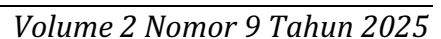
Accuracy	Precision			Recall			F1-Score		
	Positif	Netral	Negatif	Positif	Netral	Negatif	Positif	Netral	Negatif
89%	88%	88%	92%	87%	88%	93%	87%	88%	93%

Tabel 2.2.4 Hasil Accuracy Setelah Melakukan SMOTE

Pada tabel 3.9 menunjukkan pada skenario pengujian kedua hasil klasifikasi yang dilakukan oleh metode *Naïve Bayes* sangat baik ditunjukkan oleh akurasi terbesar yaitu 89%. Selanjutnya, dapat dilihat sentimen positif kemampuannya untuk membedakan data sentimen positif dengan yang lainnya sangat baik dengan nilai *precision* terbesar yaitu 88% . Kemudian, pada sentimen netral kemampuannya dalam mengenali data sentimen netral sangat baik dengan nilai *recall* sebesar 88%.

2.2.3 Visualisasi Data

Word Cloud merupakan modul yang disediakan oleh bahasa pemrograman *Python* yang dapat digunakan untuk memvisualisasikan tesk. Tesk dengan frekuensi yang berbeda-beda digambarkan dengan ukuran dan warna yang berbeda dari hasil pengumpulan data *tweet* dengan *Hashtag* #investasi, kata-kata yang muncul terkait dengan kata kunci tersebut ditampilkan pada gambar 2.1.7



Dari gambar dapat diketahui dari 1000 *Tweet* yang digunakan pada penelitian ini yang dikumpulkan dari periode 9 Februari 2023 sampai dengan 12 April 2023 menunjukkan dengan kata yang paling banyak muncul adalah positif “investasi” sebanyak 468, negatif “uang” sebanyak 194 kata, netral “uang” sebanyak 461 kata.

Gradio adalah sebuah perpustakaan (*library*) Python yang memungkinkan pengembang untuk dengan mudah membuat antarmuka pengguna (UI) interaktif untuk model pembelajaran mesin (*machine learning*) dan aplikasi analisis data lainnya. *Gradio* dapat digunakan dalam analisis sentimen untuk membuat antarmuka pengguna interaktif yang memungkinkan pengguna memasukkan teks dan melihat hasil sentimen dari teks tersebut. Berikut contoh *prototype gradio* pada gambar 2.1.8

<https://journal.hasbaedukasi.co.id/index.php/at-taklim>

3. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat ditarik kesimpulan sebagai berikut.

1. Berdasarkan dari sentimen yang didapat dari media sosial *twitter* menunjukkan bahwa pengguna media sosial *Twitter* memiliki sentimen positif terhadap investasi keuangan di Indonesia dengan total jumlah sentimen positif sebanyak 426 data dari total 1000 data tweet.
2. Data yang tidak seimbang jumlahnya yaitu 426 data berlabel positif, 171 data berlabel negatif, dan 403 data berlabel netral mempengaruhi hasil klasifikasi.
3. Dalam analisis sentimen yang dilakukan menggunakan algoritma *Naïve Bayes* dengan 3 sentimen yaitu positif, negatif dan netral, hasil performa terbesar yang didapatkan yaitu akurasi sebesar 0,78, *precision* sebesar 0,79 dan *recall* sebesar 0,78 dengan metode up-sampling SMOTE dibandingkan tanpa menggunakan metode *up-sampling* SMOTE menghasilkan akurasi sebesar 0.89, *precision* sebesar 0.92 dan *recall* sebesar 0.94.
4. Dari 1000 data *Tweet* yang digunakan pada penelitian ini yang dikumpulkan dari periode 9 Februari 2023 sampai dengan 12 April 2023 menghasilkan kata yang paling banyak muncul adalah “investasi”

4. DAFTAR PUSTAKA

- Amaliah, F., & Dwi Nuryana, I. K. (2022). Perbandingan akurasi metode lexicon based dan naive bayes classifier pada analisis sentimen pendapat masyarakat terhadap aplikasi investasi pada media twitter. *Journal of Informatics and Computer Science (JINACS)*, 3(03), 384–393. <https://doi.org/10.26740/jinacs.v3n03.p384-393>
- Anggraeni, D. T. (2022). Analisis sentimen vaksinasi booster covid-19 pada platform twitter menggunakan metode naïve bayes. *EXPERT: Jurnal manajemen sistem informasi dan teknologi*, 12(2), 113. <https://doi.org/10.36448/expert.v12i2.2812>
- Azhar, R., Surahman, A., & Juliane, C. (2022). Analisis sentimen terhadap cryptocurrency berbasis python textblob menggunakan algoritma naïve bayes. *Jurnal sains komputer & informatika (J-SAKTI)*, 6(1), 267–281.
- Guntur, M., Santony, J., & Yuhandri, Y. (2018). Prediksi harga emas dengan menggunakan metode naïve bayes dalam investasi untuk meminimalisasi resiko. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2(1), 354–360. <https://doi.org/10.29207/resti.v2i1.276>

- Haryadi, D., & Mandala, R. (2019). Prediksi harga minyak kelapa sawit dalam investasi dengan membandingkan algoritma naïve bayes, support vector machine dan k-nearest neighbor. *IT for society*, 4(1). <https://doi.org/10.33021/itfs.v4i1.1181>
- Helmiah, N., Nooraeni, R., Nulkarim, A. R., Munia, N., Susanti, A., Putra, A. P., & Taufiqurrahman, F. (2020). Penerapan metode naïve bayes dalam analisis persepsi masyarakat mengenai rencana pengesahan RUU omnibus law di bidang investasi dan ketenagakerjaan tahun 2020 di Indonesia. *Jurnal MSA (matematika dan statistika serta aplikasinya)*, 8(2), 48. <https://doi.org/10.24252/msa.v8i2.16743>
- Leonie Syafira, & Rikumahu, B. (2020). Analisis korelasi sentimen pada twitter terhadap abnormal return saham (studi kasus pada saham indeks Lq45 di twitter). *Jurnal mitra manajemen*, 4(9), 1322–1335. <https://doi.org/10.52160/ejmm.v4i9.458>
- Niyah, T. Y., & Yamasari, Y. (2021). NBC berbasis ekstraksi fitur untuk analisis sentimen masyarakat terhadap fluktuasi harga saham. *Journal of Informatics and Computer Science (JINACS)*, 3(02), 113–122. <https://doi.org/10.26740/jinacs.v3n02.p113-122>
- Pribadi, M. R., Purnomo, H. D., Hendry, Hartomo, K. D., Sembiring, I., & Iriani, A. (2022). Improving the accuracy of text classification using the over sampling technique in the case of sinovac vaccine. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2022-October*(October), 106–110. <https://doi.org/10.23919/EECSI56542.2022.9946508>
- Rafih Munir, M. S., Syafrianto, A., & Ilmu Komputer, F. (2022). Sentimen analis thread investasi pada twitter menggunakan naïve bayes. *IJCSR: The Indonesian Journal of Computer Science Research*, 1, 44–52. <https://subset.id/index.php/IJCSR>
- Rusardi, M. N., Rahayudi, B., & Adikara, P. P. (2022). Analisis sentimen masyarakat terhadap isu new normal scenario berdasarkan opini dari twitter menggunakan algoritma naïve bayes classifier. *Jurnal pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(3), 1434–1440.
- SAPUTRA, T. (2020). Analisis sentimen berita saham menggunakan algoritma naïve bayes. *Katalog.Ukdw.Ac.Id*. <http://katalog.ukdw.ac.id/id/eprint/4166>
- W, B., Riski, I., Dwi, K., Nooraeni, R., Siahaan, T., & Dhea, Y. (2019). Analisis text mining dari cuitan twitter mengenai infrastruktur di Indonesia dengan metode klasifikasi naïve bayes. *Eigen mathematics journal*, 2(2), 92–101. <https://doi.org/10.29303/emj.v1i2.36>