

**ANALISIS SENTIMEN MASYARAKAT TERHADAP
KEBIJAKAN TABUNGAN PERUMAHAN RAKYAT (Tapera)
BERDASARKAN OPINI DARI TWITTER MENGGUNAKAN
METODE NAIVE BAYES DAN SUPPORT VECTOR
MACHINE (SVM)**

SKRIPSI

Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana
Komputer



Oleh:

Regy Saputra

1703027

**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN
ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
BANDUNG
2024**

ANALISIS SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN
TABUNGAN PERUMAHAN RAKYAT (Tapera) BERDASARKAN OPINI
DARI TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN
SUPPORT VECTOR MACHINE (SVM)

Oleh:
Regy Saputra
1703027

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer pada Fakultas Pendidikan Matematika dan Ilmu
Pengetahuan Alam

© Regy Saputra
Universitas Pendidikan Indonesia
Agustus 2024

Hak Cipta Dilindungi Undang-Undang
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak
ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis

REGY SAPUTRA
1703027

ANALISIS SENTIMEN MASYARAKAT TERHADAP KEBIJAKAN
TABUNGAN PERUMAHAN RAKYAT (Tapera) BERDASARKAN OPINI
DARI TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN
SUPPORT VECTOR MACHINE (SVM)

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,



Dr. Rani Megasari, M.T.

NIP. 198705242014042002

Pembimbing II,



Ani Anisyah, S.Pd., M.T.

NIP. 920200419930811201

Mengetahui,

Ketua Departemen Pendidikan Ilmu Komputer



Dr. Muhammad Nursalman, M.T.

NIP. 1979092920060

ABSTRAK

Salah satu upaya pemerintah untuk menekan angka ketimpangan pemilikan rumah atau *backlog* yang dilaporkan mencapai 9,95 juta anggota keluarga yaitu dengan program Tabungan Perumahan Rakyat atau Tapera. Namun kebijakan ini mendapat banyak kritikan khususnya di sosial media Twitter. Banyaknya opini/pendapat yang dituliskan di twitter dapat diklasifikasikan sesuai sentimen yang ada agar mudah untuk mendapatkan kecenderungan tanggapan tersebut terhadap kebijakan Tapera ini, apakah lebih cenderung ke hal yang positif, negatif, atau netral. Pada kasus ini algoritma Naïve Bayes dan SVM adalah metode yang di gunakan untuk melakukan klasifikasi pada penelitian ini, karena kedua metode ini yang populer untuk masalah klasifikasi, Kesimpulan dari hasil penelitian ini menerangkan bahwasanya metode naïve bayes memiliki akurasi yang lebih tinggi dibandingkan SVM.

Kata Kunci: Analisis Sentimen, Naïve Bayes, Support Vector Machine

ABSTRACT

One of the government's efforts to reduce the number of home ownership inequality or backlog reported to reach 9.95 million family members is through the People's Housing Savings or Tapera program. However, this policy has received a lot of criticism, especially on social media Twitter. The many opinions/views written on Twitter can be classified according to existing sentiments so that it is easy to get the tendency of the response to this Tapera policy, whether it tends to be positive, negative, or neutral. In this case, the Naïve Bayes and SVM algorithms are the methods used to classify in this study, because these two methods are popular for classification problems. The conclusion of the results of this study explains that the Naïve Bayes method has higher accuracy than SVM.

Keywords: *Sentiment Analysis, Naïve Bayes, Support Vector Machine*

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
DAFTAR ISI	iii
DAFTAR TABEL	vi
DAFTAR GAMBAR	vii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian.....	4
BAB 2 TINJAUAN PUSTAKA	6
2.1 Tabungan Perumahan Rakyat.....	6
2.2 Machine Learning.....	7
2.2.1 Supervised Learning.....	7
2.2.2 <i>Unsupervised Learning</i>	8
2.3 Proses <i>Machine Learning</i>	8
2.4 <i>Natural Language Processing</i>	10
2.5 <i>Text Mining</i>	12
2.6 Analisis Sentimen.....	12
2.7 Media Sosial	13
2.8 Twitter	14
2.9 <i>Data Crawling</i>	15
2.10 <i>Text Preprocessing</i>	15

2.11	<i>Bidirectional Encoder Representations from Transformers (BERT)</i>	16
2.12	<i>K-Fold Cross Validation</i>	18
2.13	Klasifikasi.....	19
2.14	<i>Split Validation Data</i>	19
2.15	<i>Synthetic Minority Oversampling Technique (SMOTE)</i>	20
2.16	<i>Naïve Bayes Classifier</i>	21
2.17	<i>Support Vector Machine (SVM)</i>	23
2.18	<i>Confusion Matrix</i>	27
2.19	Cross Validation	29
BAB 3	METODOLOGI PENELITIAN	32
3.1	Alur Penelitian.....	32
3.1.1	Pengumpulan Data	32
3.1.2	Implementasi	33
3.1.3	Analisis.....	34
3.2	Perangkat Penelitian	35
BAB 4	HASIL DAN PEMBAHASAN	36
4.1	Pengumpulan Data.....	36
4.1.1	Studi Literatur	36
4.1.2	Crawling Data	36
4.2	Implementasi	38
4.2.1	Load Dataset.....	38
4.2.2	Preprocessing	39
4.2.3	Labeling.....	46
4.2.4	Split Data.....	46
4.2.5	Pengujian.....	50
4.3	Analisis.....	57

BAB 5 PENUTUP	64
5.1 Kesimpulan.....	64
5.2 Saran.....	64
DAFTAR PUSTAKA	65

DAFTAR TABEL

Table 2-1 Macam-macam fungsi kernel	27
Table 2-2 Confusion Matrix 2x2.....	28
Table 4-1 Hasil proses text cleaning	41
Table 4-2 Hasil normalization.....	43
Table 4-3 Hasil stopword removal	44
Table 4-4 Hasil stemming	45
Table 4-5 Contoh data untuk perhitungan teorema bayes.....	50
Table 4-6 Word fequencies	51

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Stratified K-Fold Cross Validation (k=5).....	18
Gambar 2.2 Hyperplane Terbaik Yang Memisahkan Kedua Class Negatif Dan Positif	26
Gambar 2.3 Persamaan akurasi	28
Gambar 2.4 Persamaan precision	28
Gambar 2.5 Persamaan recall.....	29
Gambar 2.6 Persamaan F1-score.....	29
Gambar 3.1 Alur Penelitian.....	32
Gambar 3.2 Implementasi	33
Gambar 4.1 Instalasi library twikit.....	36
Gambar 4.2 Tampilan Google Collab	37
Gambar 4.3 Crawling data dengan library twikit.....	37
Gambar 4.4 Hasil Crawling Data	38
Gambar 4.5 Load dataset.....	38
Gambar 4.6 Remove Unused Column	39
Gambar 4.7 Case Folding.....	39
Gambar 4.8 Proses text cleaning.....	40
Gambar 4.9 String kosong.....	41
Gambar 4.10 Remove Empty Tweet.....	42
Gambar 4.11 Menghapus non-Indonesia tweet.....	42
Gambar 4.12 Menghapus data ganda	43
Gambar 4.13 Normalization menggunakan microsoft excel.....	43
Gambar 4.14 Stopword Removal	44
Gambar 4.15 Stemming	45
Gambar 4.16 Tokenization.....	46
Gambar 4.17 Hasil Labeling	46
Gambar 4.18 Split data.....	47
Gambar 4.19 Perbandingan label positif dan negatif.....	47
Gambar 4.20 Oversampling dengan SMOTE	47
Gambar 4.21 Hasil data train setelah oversampling.....	48

Gambar 4.22 Pembobotan dengan BERT	49
Gambar 4.23 Hasil proses BERT	50
Gambar 4.24 Import package untuk klasifikasi naïve bayes.....	53
Gambar 4.25 Proses klasifikasi naïve bayes	53
Gambar 4.26 SVM Classification	54
Gambar 4.27 Akurasi naïve bayes	54
Gambar 4.28 Hasil uji model naïve bayes	55
Gambar 4.29 Akurasi SVM.....	55
Gambar 4.30 Hasil uji model svm.....	56
Gambar 4.31 Hasil pengukuran evaluasi performa naïve bayes	57
Gambar 4.32 Hasil pengukuran evaluasi peforma SVM	58
Gambar 4.33 Hasil cross validation pada naïve bayes	59
Gambar 4.34 Hasil cross validation pada SVM	59
Gambar 4.35 Wordcloud untuk tweet negatif	60
Gambar 4.36 Wordcloud untuk tweet positif.....	61
Gambar 4.37 Distribusi sentimen positif dan negatif.....	61
Gambar 4.38 Learning curve Naïve Bayes dan SVM.....	62

DAFTAR PUSTAKA

- Undang-Undang No. 6 tahun 2011 tentang Perumahan dan Kawasan Permukiman.
Diakses tanggal 2 Agustus dari
<https://peraturan.bpk.go.id/Details/39128/uu-no-1-tahun-2011>
- Peraturan Pemerintah No. 21 Tahun 2024 tentang Penyelenggaraan Tabungan Perumahan. Diakses tanggal 2 Agustus dari
<https://peraturan.bpk.go.id/Details/286236/pp-no-21-tahun-2024>
- Ahmad, M., Aftab, S., & Ali, I. (2017). Sentiment Analysis of Tweets using SVM. *International Journal of Computer Applications*, 177(5), 25–29.
<https://doi.org/10.5120/ijca2017915758>
- Apriani, R., Gustian, D., Program, S., Sistem, I., Putra, U. N., Indonesia, S., Raya, J., Kaler, C., 21, N., & Sukabumi, K. (2019). ANALISIS SENTIMEN DENGAN NAÏVE BAYES TERHADAP KOMENTAR APLIKASI TOKOPEDIA. In *Jurnal Rekayasa Teknologi Nusa Putra* (Vol. 6, Issue 1).
- Arnout Van Messem. (2020). *Support vector machines: A robust prediction method with applications in bioinformatics* (Vol. 43). Elsevier.
- Asghar, M. Z., Khan, A., Ahmad, S., Qasim, M., & Khan, I. A. (2017). Lexicon-enhanced sentiment analysis framework using rule-based classification scheme. *PLoS ONE*, 12(2), 1–22.
<https://doi.org/10.1371/journal.pone.0171649>
- Athira Luqyana, W., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 4704–4713. <http://j-ptiik.ub.ac.id>
- Berrar, D. (2019). *Cross validation*. 1–13.
- Cruz, N. P., Taboada, M., & Mitkov, R. (2016). A machine-learning approach to negation and speculation detection for sentiment analysis. In *Journal of the Association for Information Science and Technology* (Vol. 67, Issue 9, pp. 2118–2136). <https://doi.org/10.1002/asi.23533>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. *Jurnal*

- Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>
- Diekson, Z. A., Prakoso, M. R. B., Putra, M. S. Q., Syaputra, M. S. A. F., Achmad, S., & Sutoyo, R. (2022). Sentiment analysis for customer review: Case study of Traveloka. *Procedia Computer Science*, 216(2022), 682–690. <https://doi.org/10.1016/j.procs.2022.12.184>
- Faiq, M., Putro, A., & Setiawan, E. B. (2022). Analisis Sentimen Terhadap Kebijakan Pemerintah dengan Feature Expansion Metode GloVe pada Media sosial Twitter. *E-Proceeding of Engineering*, 9(1), 54–66.
- Fajar, R., Program, S., Rekayasa, P., Lunak, N., & Bengkalis, R. (2018). *Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter*. 3(1).
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, 161, 765–772. <https://doi.org/10.1016/j.procs.2019.11.181>
- Fowler, R. (2022). *Understanding language: An introduction to linguistics*. Routledge.
- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: an Overview*. <http://arxiv.org/abs/2008.05756>
- Haque, R., Islam, N., Tasneem, M., & Das, A. K. (2023). Multi-class sentiment classification on Bengali social media comments using machine learning. *International Journal of Cognitive Computing in Engineering*, 4(January), 21–35. <https://doi.org/10.1016/j.ijcce.2023.01.001>
- Imron, A. (2019). Analisis Sentimen Terhadap Tempat Wisata di Kabupaten Rembang Menggunakan Metode Naive Bayes Classifier. *Teknik Informatika*, 10–13. <https://dspace.uii.ac.id/handle/123456789/14268>
- Izzul Fikri, M., Shina Sabrila, T., & Azhar, Y. (2020). *Comparison of the Naïve Bayes Method and Support Vector Machine on Twitter Sentiment Analysis*. 10.
- Joseph, K., & Carley, K. M. (2016). Relating semantic similarity and semantic

association to how humans label other people. *NLP + CSS 2016 - EMNLP 2016 Workshop on Natural Language Processing and Computational Social Science, Proceedings of the Workshop*, 1–10.

<https://doi.org/10.18653/v1/w16-5601>

Kharde, V. A., & Sonawane, S. S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. In *International Journal of Computer Applications* (Vol. 139, Issue 11). <http://ai.stanford>.

Mahfudh, A. A., & Mustofa, H. (2019). Klasifikasi Pemahaman Santri Dalam Pembelajaran Kitab Kuning Menggunakan Algoritma Naive Bayes Berbasis Forward Selection. *Walisongo Journal of Information Technology*, 1(2), 101. <https://doi.org/10.21580/wjit.2019.1.2.4529>

Mas Pintoko, B., & Muslim, K. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier. *E-Proceeding of Engineering*, 5(3), 8121–8130.

Musto, C., Semeraro, G., & Giuliani, A. (2014). *A Comparison of Lexicon-based Approaches for Sentiment Analysis of Microblog Posts*. <http://aiia2014.di.unipi.it/dart/>

Neogi, A. S., Garg, K. A., Mishra, R. K., & Dwivedi, Y. K. (2021). Sentiment analysis and classification of Indian farmers' protest using twitter data. *International Journal of Information Management Data Insights*, 1(2), 100019. <https://doi.org/10.1016/j.jjime.2021.100019>

Nurhuda, F., Sihwi, S. W., & Doewes, A. (2013). *Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier*. 2(2).

Patel, A., Oza, P., & Agrawal, S. (2022). Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model. *Procedia Computer Science*, 218, 2459–2467. <https://doi.org/10.1016/j.procs.2023.01.221>

Pradany, L. N., & Fatichah, C. (2016). *ANALISA SENTIMEN KEBIJAKAN PEMERINTAH PADA KONTEN TWITTER BERBAHASA INDONESIA MENGGUNAKAN SVM DAN K-MEDOID CLUSTERING*.

Prasetyo, V. R., Benarkah, N., & Chrisintha, V. J. (2021). Implementasi Natural

- Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya. *Teknika*, 10(2), 114–121.
<https://doi.org/10.34148/teknika.v10i2.370>
- Putri, M. I., & Kharisudin, I. (2022). Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Analisis Sentimen Data Review Pengguna Aplikasi Marketplace Tokopedia. *PRISMA, Prosiding Seminar Nasional Matematika*, 22.
- Ramadhan, D., & Setiawan, E. B. (2019). Analisis Sentimen Program Acara di SCTV pada Twitter Menggunakan Metode Naive Bayes dan Support Vector Machine. ... *Telkomuniversity.Ac.Id*, 6(2), 9736–9743.
<https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/10708>
- Rustiana, D., & Rahayu, N. (2017). ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL: TWEET TWITTER MENGGUNAKAN NAÏVE BAYES. *Jurnal SIMETRIS*, 8.
- Sandy, B. C., Manongga, D., & Iriani, A. (2015). ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA BAHAN BAKAR MINYAK (BBM) PADA MEDIA ONLINE. 1–23.
- Sholekha, I., Faqih, A., & Bahtiar, A. (2022). Sentiment Analysis of Public Opinion Covid-19 Vaccine Using Naïve Bayes and Random Forest Methods. *Teknik Informatika*, 15.
- Sujadi, H. (2022). Analisis Sentimen Pengguna Media Sosial Twitter Terhadap Wabah Covid-19 Dengan Metode Naive Bayes Classifier Dan Support Vector Machine. *INFOTECH Journal*, 8(1), 22–27.
<https://doi.org/10.31949/infotech.v8i1.1883>
- Syakuro, A. (2017). Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG) Halaman Judul Skripsi Oleh : Abdan Syakuro. *Analisis Sentimen Masyarakat Terhadap E-Commerce Pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG)*, 1–89.
<http://etheses.uin-malang.ac.id/11706/>
- Tangkelayuk, A., & Mailoa, E. (2022). *Klasifikasi Kualitas Air Menggunakan*

Metode KNN, Naïve Bayes Dan Decision Tree. 9(2), 1109–1119.

<http://jurnal.mdp.ac.id>

Turban, E., Sharda, R., & Delen, D. (2011). *Decision Support and Business Intelligence Systems 9th Ed.* Pearson Education.

Turmudi Zy, A., Adji Ardiansyah, L., & Maulana, D. (2021). Implementasi Algoritma Naïve Bayes Dalam Mendiagnosa Penyakit Angin Duduk. *Jurnal Pelita Teknologi, 16(1), 52–65.*

Valero-Carreras, D., Alcaraz, J., & Landete, M. (2023). Comparing two SVM models through different metrics based on the confusion matrix. *Computers and Operations Research, 152(April 2022), 106131.*

<https://doi.org/10.1016/j.cor.2022.106131>

Wandani, A. (2021). Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes. *Jurnal Sains Komputer & Informatika (J-SAKTI, 5(2), 651–665.*

Yang, X. S. (2019). Introduction to algorithms for data mining and machine learning. *Introduction to Algorithms for Data Mining and Machine Learning, 1–173.* <https://doi.org/10.1016/C2018-0-02034-4>

Yerzi, F. S., & Sibaroni, Y. (2021). Analisis Sentimen terhadap Kebijakan Pemerintah dalam Menangani Covid-19 dengan Pendekatan Lexicon Based. *E-Proceeding of Engineering, 8(5), 11354–11366.*

<https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/download/15616/15329>