E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 1, No. 3c, Juli 2024

MISTER

Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research

Jurnal Penelitian Multidisiplin dalam Ilmu Pengetahuan, Teknologi dan Pendidikan

UNIVERSITAS SERAMBI MEKKAH

KOTA BANDA ACEH

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology and Educational Research

Journal of MISTER

Vol. 1, No. 3c, Juli 2024

Pages: 1676-1686

Penerapan Algoritma SVM untuk Deteksi Review Buatan Bot pada Ulasan Aplikasi Facebook di Google Play Store

Mikhail Shams Afzal Karim, Albi Akhsanul Hakim, Ama Maulidatul Khairah, Anggraini Puspita Sari

Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia

Article in Journal of MISTER

Available at	:	https://jurnal-serambimekkah.org/index.php/mister/index
DOI	:	https://doi.org/10.32672/mister.v1i3c.2065

How to Cite this Article

APA	:	Shams Afzal Karim, M., Akhsanul Hakim, A., Maulidatul Khairah, A., & Puspita Sari, A. (2024). Penerapan Algoritma SVM untuk Deteksi Review Buatan Bot pada Ulasan Aplikasi Facebook di Google Play Store. *MISTER: Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, *1*(3c), 1676 – 1686. https://doi.org/10.32672/mister.v1i3c.2065
Others Visit	:	https://jurnal-serambimekkah.org/index.php/mister/index

ISSN 3032-7105 ISSN 3032-601X

Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research
UNIVERSITAS SERAMBI MEKKAH
https://jurnal-serambimekkah.org/index.php/mister

e-ISSN 3032-601X & p-ISSN 3032-7105
Vol. 1 No. 3c, Juli, 2024
Doi: 10.32672/mister.v1i3c.2065
Hal. 1676-1686

Penerapan Algoritma SVM untuk Deteksi Review Buatan Bot pada Ulasan Aplikasi Facebook di Google Play Store

Mikhail Shams Afzal Karim[1], Albi Akhsanul Hakim[2], Ama Maulidatul Khairah[3], Anggraini Puspita Sari[4*]

Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Surabaya, Indonesia[1,2,3,4]

[*]Email Korespodensi: anggraini.puspita.if@upnjatim.ac.id

Diterima: 19-07-2024 | Disetujui: 21-07-2024 | Diterbitkan: 23-07-2024

ABSTRACT

App reviews on Google Play Store serve as a crucial source of information for users to select applications that align with their needs. However, there is a growing concern regarding the misuse of app reviews through the creation of fake reviews by bots. These fake reviews can mislead users and negatively impact the reputation of app developers. This research aims to detect bot-generated reviews in Facebook app reviews on Google Play Store using the Support Vector Machine (SVM) algorithm. SVM is a classification algorithm that has proven effective in various classification tasks, including fraud detection. The research utilizes a dataset that has been categorized as genuine or bot-generated texts. The results demonstrate that the SVM algorithm successfully detects bot reviews with a high F1-score of 0.8981, exhibiting no signs of overfitting. This is further supported by the consistent F1-score values across training and testing data. Implementing the SVM algorithm can contribute to enhancing the quality of app reviews and protecting users from misleading information. This research shows that the SVM algorithm can be an effective tool for detecting bot-generated reviews on Facebook application reviews on the Google Play Store.

Keywords: *Facebook; Reviews; Support Vector Machine; Bot Generated.*

ABSTRAK

Ulasan aplikasi di Google Play Store menjadi sumber informasi penting bagi pengguna dalam memilih aplikasi yang sesuai dengan kebutuhannya. Namun, terdapat potensi penyalahgunaan dengan pembuatan ulasan palsu oleh bot, yang dapat menyesatkan pengguna dan mencoreng reputasi pengembang aplikasi. Penelitian ini bertujuan untuk mendeteksi ulasan buatan bot pada ulasan aplikasi Facebook di Google Play Store menggunakan algoritma Support Vector Machine (SVM). SVM merupakan algoritma klasifikasi yang terbukti efektif dalam berbagai tugas klasifikasi, termasuk deteksi penipuan. Penelitian ini menggunakan dataset yang telah diklasifikasikan sebagai teks asli atau teks buatan bot. Hasil penelitian menunjukkan bahwa algoritma SVM mampu mendeteksi ulasan bot dengan nilai F1-score yang tinggi, yaitu 0.8981, tanpa menunjukkan tanda-tanda overfitting. Hal ini dibuktikan dengan kesamaan nilai F1-score antara data training dan data test. Penerapan algoritma SVM dapat membantu meningkatkan kualitas ulasan aplikasi dan melindungi pengguna dari informasi yang menyesatkan. Penelitian ini menunjukkan bahwa algoritma SVM dapat menjadi alat yang efektif untuk mendeteksi review yang dihasilkan bot pada ulasan aplikasi Facebook di Google Play Store.

Katakunci: Facebook; Ulasan; Support Vector Machine; Buatan Bot

e-ISSN3032-601X & p-ISSN3032-7105

PENDAHULUAN

Pesatnya perkembangan teknologi digital sehingga berdampak langsung dengan peningkatan pemanfaatan teknologi di lingkungan masyarakat (Sari, A. et al., 2022). Dunia ini telah didominasi oleh jejaring sosial online, yang digunakan sebagai saluran komunikasi publik, sehingga memiliki peran penting dalam kehidupan saat ini (Yuhdi Fahrimal, 2018). Salah satunya adalah Facebook, merupakan salah satu platform sosial media terbesar di dunia dengan penggunaan yang luas di berbagai negara. Facebook telah berkembang menjadi lebih dari jaringan sosial dan telah menjadi bagian penting dari kehidupan digital sehari-hari bagi jutaan orang di seluruh dunia yang memungkinkan penggunanya terhubung dan berinteraksi dengan orang lain di komunitas seperti kota, tempat kerja, kampus, dan lingkungan sekitar (M. Ali Nurhasan Islamy & Ika Laksmiwati, 2020). Dengan jumlah pengguna aktif bulanan yang mencapai lebih dari 2,9 miliar pada tahun 2022, Facebook telah menjadi platform yang sangat berpengaruh dalam membentuk opini dan tren publik di berbagai bidang, termasuk dalam hal rating dan review produk serta layanan digital.

Salah satu fitur utama Facebook adalah kemampuannya menyediakan forum diskusi bagi pengguna, termasuk dengan memberikan review dan rating untuk aplikasi seluler yang tersedia di toko aplikasi seperti Google Play Store. Ulasan aplikasi di Facebook adalah sumber informasi berharga bagi calon pengguna untuk menilai kualitas dan kinerja aplikasi sebelum mengunduhnya (Darmawan, A. C., 2023). Namun, seperti platform lainnya, ulasan aplikasi Facebook di Google Play Store juga rentan terhadap aktivitas ulasan palsu dan ulasan yang dibuat oleh bot. Motivasi atas perilaku ini merupakan upaya dari pengembang aplikasi untuk meningkatkan peringkat aplikasi secara tidak etis hingga kampanye perusahaan pesaing untuk menghentikan aplikasi pesaing dengan ulasan negatif yang tidak valid. Ulasan palsu yang dibuat oleh bot dapat memberikan informasi yang menyesatkan, memengaruhi persepsi pengguna terhadap aplikasi, dan mengganggu proses pengambilan keputusan yang seharusnya didasarkan pada informasi yang obyektif dan dapat diandalkan. Selain itu, perilaku ini dapat mengganggu persaingan yang sehat dan mengurangi kepercayaan pengguna terhadap platform.

Cara efektif untuk mengatasi masalah ini adalah diperlukan alat untuk mendeteksi ulasan yang dihasilkan oleh bot secara akurat. Salah satu pendekatan yang menjanjikan adalah dengan menggunakan algoritma support vector machine (SVM), algoritma ini dikenal karena kemampuannya dalam menghasilkan performa tinggi pada berbagai tugas klasifikasi teks, seperti analisis sentimen dan klasifikasi ujaran kebencian, dengan akurasi yang seringkali melebihi 90% (Habibi, M., & Cahyo, P. W., 2020). Dengan menggunakan fitur linguistik dan metadata yang tersedia dalam ulasan, algoritma SVM dapat digunakan untuk membedakan antara ulasan manusia dan bot.

Aplikasi Facebook di Google Play Store sering menjadi sasaran bot karena popularitasnya yang tinggi. Kombinasi fitur seperti panjang teks, frekuensi kata unik, dan analisis sentimen dapat meningkatkan akurasi pendeteksian ulasan palsu menggunakan algoritma Support Vector Machine (SVM) (Elshrif Ibrahim Elmurngi & Abdelouahed Gherbi, 2018). Dalam konteks ini, penelitian menerapkan SVM yang bertujuan untuk mendeteksi ulasan yang dihasilkan bot dalam ulasan aplikasi Facebook di Google Play Store. Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin yang cukup populer untuk masalah klasifikasi. Algoritma SVM berupaya menemukan hyperplane (garis pemisah) optimal yang memisahkan dua kelas data dengan memaksimalkan margin (jarak) antara titik data terdekat dari kedua kelas, titik data terdekat ini disebut vektor pendukung. Saat mengklasifikasikan tinjauan nyata dan bot, diasumsikan bahwa dua kelas dalam ruang fitur tertentu dapat dipisahkan secara linear oleh hyperplane.

e-ISSN3032-601X & p-ISSN3032-7105

SVM menemukan hyperplane optimal yang memaksimalkan jarak antara dua kelas. Hal ini memberikan batasan isolasi yang kuat dan kemampuan generalisasi yang baik untuk data baru. Menurut Nurachim, R. I. (2019), walaupun waktu pelatihan SVM umumnya lambat, metode ini lebih akurat karena kemampuannya menangani model-model nonlinier yang kompleks. SVM juga kurang rentan terhadap overfitting dibandingkan metode lainnya. SVM dapat digunakan untuk tujuan prediksi dan klasifikasi. Dengan mengukur hasil tingkat akurasi, tingkat presisi, recall, dan F1-score dengan harapan dapat memperoleh hasil terbaik dibandingkan dengan metode klasifikasi lainnya (Parapat, I. M., et al, 2018).

METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah SVM. Dengan metode ini, dapat mendeteksi review buatan bot pada ulasan aplikasi facebook di google play store. Berikut pada Gambar 1 adalah alur metode SVM.

Gambar 1. Alur Metode Support Vector Machine

A. *Pengumpulan Data*

Pada penelitian ini, data yang dipakai adalah dataset yang dikumpulkan atau didapatkan melalui https://www.kaggle.com/datasets/bwandowando/facebook-app-google-store-reviews-31-countries/data dan https://huggingface.co/datasets/NicolaiSivesind/human-vs-machine/tree/main. Pada dataset pertama berisi tentang ulasan aplikasi Facebook di Google Play Store dan pada dataset kedua berisi data teks buatan manusia dan teks buatan bot.

B. *Pre-processing*

Pre-processing data dilakukan untuk menghilangkan berbagai masalah yang mungkin mempengaruhi pemrosesan data. Hal ini karena banyak data yang tidak diformat secara konsisten. Preprocessing data merupakan teknik pertama sebelum data mining. Namun, ada beberapa proses, yaitu:

1) Lowercase

Lowercase merupakan proses awal dalam pre-processing, pada tahap ini semua karakter teks dalam data diubah menjadi huruf kecil (Fani et al., 2021). Langkah ini dilakukan untuk memastikan konsistensi dalam analisis teks. Pada langkah ini, data review dimuat dari file '.csv'. Pada proses ini tiap review Facebook di Google Playstore yang terdapat karakter atau huruf kapital akan diubah menjadi huruf kecil.

2) Stopwords Removal

Stopwords Removal merupakan proses menghilangkan kata - kata umum yang sering muncul dalam dataset tetapi tidak memiliki nilai informasi yang signifikan (Ulfah, A. N., & Anam, M. K., 2020). Kata - kata tersebut biasanya tidak memberikan banyak informasi tentang isi teks dan dapat menghambat analisis atau model prediksi. Daftar stopwords yang digunakan dalam

program penelitian adalah bahasa Inggris. Data yang telah diproses (tanpa stopwords) dapat disimpan kembali untuk analisis lebih lanjut.

3) Lemmatization
Lemmatization merupakan langkah terakhir pada pre-processing, mengubah kata-kata pada teks review ke bentuk dasar mereka yang disebut lema (Bahtiar, S. A. H., 2023). Lema adalah bentuk kata yang ada dalam kamus dan mewakili makna utama atau dasar dari kata tersebut. Tujuannya adalah untuk mengurangi variasi kata yang disebabkan oleh infleksi (perubahan bentuk kata) dan mengembalikan setiap kata ke bentuk dasar yang sederhana. Proses ini lebih kompleks dibandingkan dengan stemming karena mempertimbangkan konteks dan arti kata dalam kalimat. Data yang telah diproses (dengan lemmatization) dapat disimpan kembali untuk analisis lebih lanjut.

C. *Data Training & Data Testing*
Pemisahan data menjadi data training dan data testing adalah untuk memungkinkan model pembelajaran machine learning dari data yang tersedia dan menjalankan pengujian pada data yang belum pernah ada sebelumnya. Hal ini penting untuk mengukur kemampuan generalisasi model (Djuardi, R., & Rochadiani, T., 2024). Pada penelitian ini, data pelatihan dan data testing dibagi dengan rasio 80:20. Dari 20.000 data dalam dataset, 16.000 dialokasikan sebagai data training. 4.000 data dari data training digunakan untuk validasi internal, dan 4.000 data lainnya digunakan untuk validasi eksternal untuk mendeteksi overfitting model. Pembagian data ini dilakukan menggunakan metode random sampling untuk memastikan representasi yang adil dari semua kelas dalam dataset. Rasio 80:20 dipilih karena dianggap optimal untuk menyeimbangkan antara akurasi model dan efisiensi pelatihan. Data validasi internal digunakan untuk memantau performa model selama pelatihan dan membantu mencegah overfitting. Data validasi eksternal digunakan untuk mengevaluasi performa model pada data yang belum pernah dilihat sebelumnya, sehingga memberikan gambaran yang lebih realistis tentang kemampuan generalisasi model.

D. *Term Frequency - Inverse Document Frequency (TF-IDF)*
Menurut Melita, R. (2018), Metode Term Frequency-Inverse Document Frequency (TF-IDF) adalah metode statistik yang umum digunakan dalam sistem ini untuk mengukur pentingnya sebuah term dalam sebuah dokumen relatif terhadap koleksi dokumen yang lebih besar. Metode ini membantu dalam pemeringkatan dokumen, pencarian informasi relevan, pemfilteran dokumen, dan ekstraksi ringkasan. TF (Term Frequency) merupakan ukuran yang menunjukkan seberapa sering sebuah kata muncul dalam sebuah dokumen. Sedangkan, IDF (Inverse Document Frequency) merupakan ukuran yang menilai seberapa penting sebuah kata berdasarkan frekuensi kemunculannya di seluruh dokumen (Fajrina, D. P. et al, 2023).

Pada tahap ini, mengukur seberapa penting sebuah kata relatif terhadap semua dokumen dalam dataset. Memberikan bobot pada setiap kata dalam dokumen berdasarkan term frequency (TF) dan mengurangi bobot pada suatu *term* jika kemunculannya banyak tersebar di seluruh dokumen berdasarkan Inverse Document Frequency (IDF). Berikut persamaan dari TF, IDF dan TF-IDF.

Term Frequency (TF) : $TF\ (t,d) = \frac{f(t,d)}{\Sigma k\ f(k,d)}$……..……..(1)

e-ISSN3032-601X & p-ISSN3032-7105

Inverse Document Frequency (IDF) : *IDF (t)* = $\log\left(\frac{N}{1+n(t)}\right)$……..…..…(2)

TF-IDF : *TF-IDF(t,d)=TF(t,d)×IDF(t)*.....(3)

E. *Support Vector Machine (SVM)*
Model SVM dilatih pada data pelatihan dengan tujuan menemukan hyperplane yang memisahkan kedua kelas secara optimal (Widodo, P. P. et al, 2013). SVM menggunakan fungsi kernel untuk memetakan data ke dalam ruang berdimensi tinggi yang memfasilitasi pemisahan kelas, penelitian ini menggunakan kernel linear. Parameter seperti parameter C (regularization parameter) ditetapkan untuk mengontrol trade-off antara margin dan penalti kesalahan klasifikasi.

F. *Validation*
Setelah melatih model SVM, evaluasi dilakukan menggunakan data uji untuk mengukur kinerja model. Proses ini untuk memastikan bahwa model dapat digeneralisasi dengan baik pada data baru. Metode evaluasi yang umum digunakan meliputi akurasi, presisi, recall, dan F1-score. Akurasi pada persamaan (4) adalah alat pengukuran akurasi yang baik.

Accuracy= *(TP+TN) / (TP+FP+FN+TN)* ..... (4)

F1-score (pada persamaan (5)) merupakan bobot rata-rata dari Precision dan Recall. F1-score biasanya lebih berguna daripada akurasi, terutama ketika distribusi kelas yang dihasilkan tidak merata.

F1 score = *2*(Recall * Precision) / (Recall + Precision)* .....(5)

Presisi (Persamaan (6)) adalah proporsi kumpulan data yang diprediksi dengan benar dibandingkan dengan seluruh kumpulan data yang diprediksi dengan benar.

Presisi= *TP/TP+FP* .....(6)

Recall adalah ukuran yang melengkapi presisi. Secara matematis, recall didefinisikan dengan menggunakan persamaan yang berbeda, yaitu pada persamaan 7.

Recall = *TP/TP+FN*.....(7)

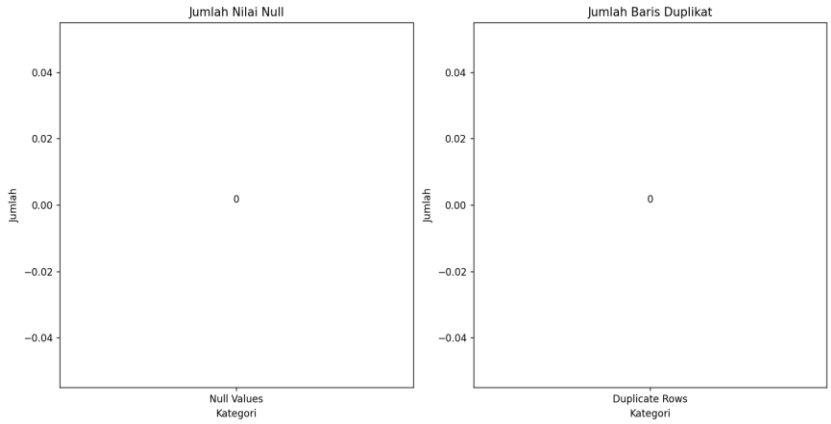
HASIL DAN PEMBAHASAN

1) Analisis Data Eksploratif

Analisis visualisasi telah dilakukan pada tahap ini dengan hasil seperti berikut:

A. Bar Chart

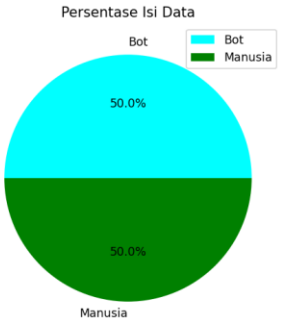
e-ISSN3032-601X & p-ISSN3032-7105



Gambar 2. Bar Chart

Visualisasi bar chart pada Gambar 2 menunjukkan bahwa dataset ini bebas dari data null dan data duplikat. Ketidakhadiran data-data bermasalah ini merupakan keuntungan besar dalam penelitian ini. Data yang bersih dan bebas dari duplikasi akan menghasilkan analisis yang lebih akurat dan terpercaya, sehingga kesimpulan yang ditarik dari penelitian ini akan lebih valid dan dapat diandalkan.

B. Pie Chart



Gambar 3. Pie Chart

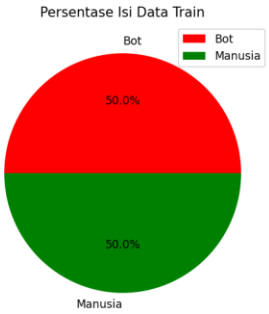
Berdasarkan visualisasi pie chart yang dipaparkan pada gambar 3, terlihat bahwa dataset penelitian ini memiliki jumlah data bot dan data manusia sama banyak. Keseimbangan data bot dan data manusia ini memungkinkan pembuatan model yang seimbang dan terhindar dari bias terhadap salah satu jenis data.

2) Prapemprosesan

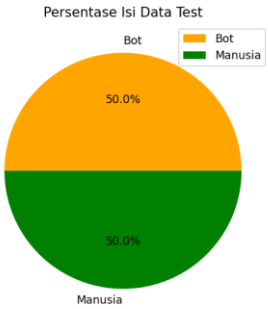
Analisis dataset penelitian mengungkapkan bahwa dataset tersebut bebas dari data outlier dan nilai null. Hal ini memungkinkan pemanfaatan seluruh kumpulan data tanpa perlu melakukan proses pembersihan data terlebih dahulu.

A. Data Splitting

Pada penelitian ini, data training dan data testing dibagi dengan rasio 80:20. Dari 20.000 data dalam dataset, 16.000 dialokasikan sebagai data testing. 4.000 data dari data training digunakan untuk validasi internal, dan 4.000 data lainnya digunakan untuk validasi eksternal untuk mendeteksi overfitting model. Dataset training dan testing keduanya memuat 50% data buatan manusia dan 50% data buatan bot. Hal ini dilakukan untuk meminimalkan bias dalam model, seperti yang ditunjukkan menggunakan pie chart pada Gambar 4 dan Gambar 5.



Gambar 4. Pie Chart Persentase Bot dan Manusia di Dataset Training



Gambar 5. Pie Chart Persentase Bot dan Manusia di Dataset Testing

Penerapan pembagian data 50:50 ini bertujuan untuk meminimalkan bias dalam model, sehingga model tidak condong terhadap salah satu jenis data, baik data buatan manusia maupun data buatan bot. Dengan meminimalkan bias, model yang dihasilkan akan lebih akurat dan adil dalam memproses dan menghasilkan prediksi, karena model tidak dipengaruhi oleh karakteristik salah satu jenis data secara berlebihan.

B. Feature Selection

Model ini menggunakan dua kolom dari file CSV: kolom *label* dan kolom *text*. Kolom label berisi nilai biner 0 atau 1, di mana 1 menunjukkan data buatan bot dan 0 menunjukkan data buatan manusia. Kolom text berisi kalimat yang akan diklasifikasikan oleh model sebagai buatan bot atau buatan manusia. Pemilihan kolom-kolom ini didasarkan pada asumsi bahwa kolom label berisi informasi yang secara langsung menunjukkan apakah data tersebut dibuat oleh bot atau manusia, sedangkan kolom text berisi data yang akan dianalisis oleh model untuk membuat klasifikasinya.

3) Pemodelan

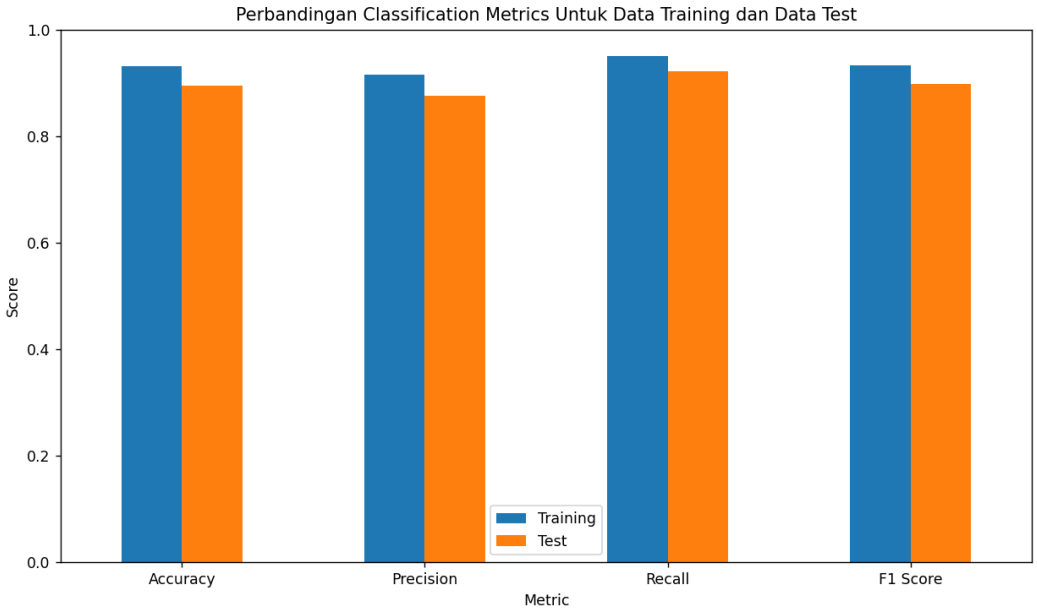
Pemodelan dalam tahap ini menggunakan algoritma Support Vector Machine (SVM), khususnya SVM Linear, model SVM dilatih menggunakan GridSearchCV untuk menemukan kombinasi hyperparameter optimal dari parameter C, gamma, dan kernel. Sebelum pemodelan, data training dibersihkan terlebih dahulu untuk mengatasi inkonsistensi format data, proses pembersihan data meliputi konversi semua kalimat ke huruf kecil, lemmatization, dan penghapusan stopword. Selanjutnya, teks diubah menjadi vektor TF-IDF agar dapat dipahami oleh model. Setelah konversi ke vektor TF-IDF, model SVM dilatih dengan GridSearchCV untuk menemukan kombinasi hyperparameter optimal dari C, gamma, dan kernel. Setelah modeling selesai, model dan vektorizer disimpan menggunakan joblib untuk penggunaan di masa depan tanpa perlu melatih ulang model.

Tahap selanjutnya adalah evaluasi model menggunakan *Classification Metrics* dimana berbagai metrik klasifikasi seperti akurasi, presisi, recall, dan f1-score akan dihitung untuk mengevaluasi performa model. Kemudian hasil evaluasi akan dibandingkan dengan menggunakan dua set data test yang berbeda: data test yang berasal dari data training dan data test yang berasal dari luar data training untuk mengidentifikasi potensi overfitting.

4) Evaluasi

Tahap terakhir dari penelitian adalah evaluasi dan menampilkan hasil. Dan didapatkan hasilnya sebagai berikut:

A. Classification Report



Gambar 6. Bar Chart Perbandingan Classification Metrics Data Training dan Data Test

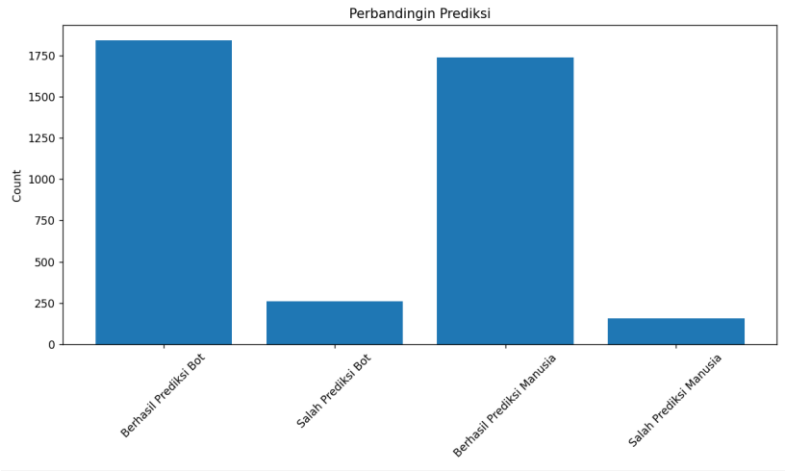
Berdasarkan hasil evaluasi menggunakan Classification Report diperoleh nilai presisi, recall, akurasi, dan f1-score yang serupa pada kedua dataset, baik dataset training maupun dataset testing. Hal ini menunjukkan bahwa model tidak mengalami overfitting karena selisih performa antara kedua dataset hanya sekitar 3-4%, yang tergolong rentang yang wajar.

B. Hasil

Tabel 1. *Classification Metrics* Data Test

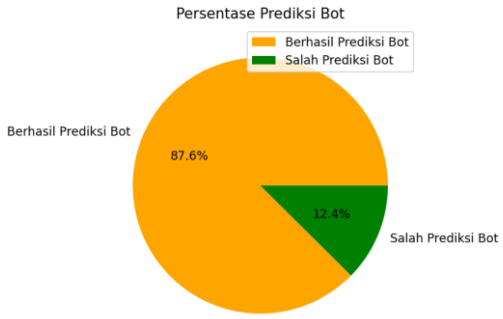
Accuracy	Precision	Recall	F1-score
0,8955	0,8760	0,9215	0.8981

Dengan tingkat F1-score yang relatif tinggi sebesar 89,81%, hasil ini menunjukkan bahwa algoritma SVM dapat menjadi alat yang efektif untuk mendeteksi review yang dihasilkan bot pada ulasan aplikasi Facebook di Google Play Store.

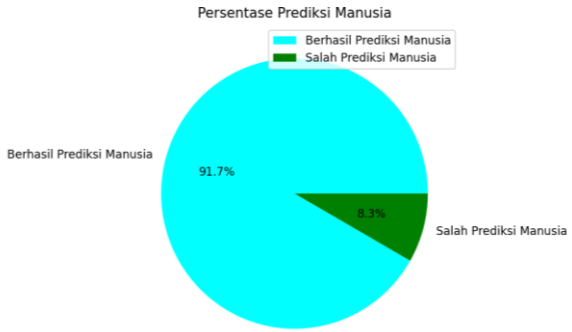


Gambar 7. Bar Chart Distribusi Prediksi Model

Gambar 7 menunjukkan distribusi prediksi model terhadap data test. Berdasarkan Gambar 7, model berhasil mengklasifikasikan 1.843 data test sebagai buatan bot dengan benar, di mana data-data tersebut memang dikategorikan sebagai buatan bot dan 1.739 data test sebagai buatan manusia dengan benar, di mana data-data tersebut memang dikategorikan sebagai buatan manusia. Namun, terdapat 261 data yang seharusnya dikategorikan sebagai buatan manusia yang diklasifikasikan sebagai bot, dan 157 data yang seharusnya dikategorikan sebagai bot diklasifikasikan sebagai buatan manusia.



Gambar 8. Pie Chart Distribusi Prediksi Bot



Gambar 9. Pie Chart Distribusi Prediksi Manusia

Gambar 8 dan 9 di atas menyajikan distribusi tingkat keberhasilan dan kegagalan model dalam bentuk pie chart. Berdasarkan diagram, model mencapai tingkat keberhasilan 87,6% dalam memprediksi data buatan bot dan 91,7% dalam memprediksi data buatan manusia.

KESIMPULAN

Penelitian ini menghasilkan pengembangan model Support Vector Machine (SVM) yang memuaskan untuk klasifikasi ulasan buatan bot dan buatan manusia. Dataset yang digunakan terdiri dari 20.000 data, terbagi secara merata (50:50) antara data buatan bot dan data buatan manusia untuk meminimalisir bias. Pembagian data dilakukan dengan alokasi 80% untuk pelatihan dan 20% untuk pengujian. Model SVM yang dibangun menggunakan kernel linear dan dioptimalkan dengan GridSearchCV untuk menemukan kombinasi hyperparameter optimal, yaitu parameter C, gamma, dan kernel. Hasil pelatihan menunjukkan akurasi 0,8955, presisi 0,8760,recall 0,9215, dan F1-score 0,8985. Model kemudian disimpan menggunakan joblib. Hasil prediksi menunjukkan bahwa dari 2.000 data buatan bot, 1.843 berhasil diidentifikasi sebagai buatan bot, dan dari 2.000 data buatan manusia, 1.843 berhasil diidentifikasi sebagai buatan manusia. Hal ini menunjukkan bahwa model yang dibangun sesuai dengan hasil yang diharapkan dan mampu mengklasifikasikan ulasan buatan bot dengan baik. Namun, masih terdapat tantangan dalam menggeneralisasi model deteksi ulasan palsu di seluruh platform dan bahasa. Oleh karena itu, penelitian lebih lanjut diperlukan untuk mengevaluasi kinerja Support Vector Machine (SVM) dalam konteks berbeda dan mengembangkan model yang lebih kuat.

DAFTAR PUSTAKA

Bahtiar, S. A. H. (2023). *Perbandingan Naïve Bayes dan Logistic Regression dalam Sentiment Analysis pada Review Marketplace menggunakan Rating-Based Labeling* (Doctoral dissertation). https://dspace.uii.ac.id/handle/123456789/dspace.uii.ac.id/123456789/46497

Darmawan, A. C. (2023). Pengembangan Aplikasi Berbasis Web dengan Python Flask untuk Klasifikasi Data Menggunakan Metode Decision Tree C4. https://dspace.uii.ac.id/handle/123456789/42602

Djuardi, R., & Rochadiani, T. (2024). Klasifikasi Suara Anjing Menggunakan Pretrained Model Yet Another Mobile Network Berbasis Convolutional Neural Network. *Building of Informatics,*

Technology and Science (BITS), *6*(1), 30−42. https://doi.org/10.47065/bits.v6i1.5165

Elmurngi, E. I., & Gherbi, A. (2018). Unfair reviews detection on amazon reviews using sentiment analysis with supervised learning techniques. *J. Comput. Sci.*, *14*(5), 714-726.

Fahrimal, Y. (2018). Netiquette: Etika jejaring sosial generasi milenial dalam media sosial.

Fajrina, D. P., Syafriandi, S., Amalita, N., & Salma, A. (2023). Sentiment Analysis of TikTok Application on Twitter using The Naïve Bayes Classifier Algorithm. *UNP Journal of Statistics and Data Science*, *1*(5), 392-398. https://doi.org/10.24036/ujsds/vol1-iss5/103

Fani, S. M., Santoso, R., & Suparti, S. (2021). Penerapan Text Mining Untuk Melakukan Clustering Data Tweet Akun Blibli Pada Media Sosial Twitter Menggunakan K-Means Clustering. *Jurnal Gaussian*, *10*(4), 583-593. Retrieved from https://ejournal3.undip.ac.id/index.php/gaussian/

Habibi, M., & Cahyo, P. W. (2020). Journal Classification Based on Abstract Using Cosine Similarity and Support Vector Machine. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *4*(3), 185-192. https://doi.org/10.14421/jiska.2020.43-06

Islamy, M. A. N., & Laksmiwati, I. (2020). Pemanfaatan Media Sosial Sebagai Sarana Promosi Layanan Perpustakaan Institut Seni Indonesia Surakarta. *Nusantara Journal of Information and Library Studies (N-JILS)*, *3*(1), 75–87. https://doi.org/10.30999/n-jils.v3i1.804

Melita, R. (2018). *Penerapan metode term frequency inverse document frequency (tf-IDF) dan cosine similarity pada sistem temu kembali INFORMASI untuk mengetahui Syarah Hadits Berbasis web (Studi Kasus: Hadits Shahih Bukhari-Muslim)* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta). https://repository.uinjkt.ac.id/dspace/handle/123456789/54945

Nurachim, R. I. (2019). Pemilihan model prediksi indeks harga saham yang dikembangkan berdasarkan algoritma Support Vector Machine (SVM) atau Multilayer Perceptron (MLP) studi kasus: saham PT Telekomunikasi Indonesia Tbk. *Jurnal Teknologi Informatika dan Komputer*, *5*(1), 29-35. https://doi.org/10.37012/jtik.v5i1.243

Parapat, I. M., Furqon, M. T., & Sutrisno, S. (2018). Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *2*(10), 3163-3169. https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2577

Sari, A., Prasetya, D. A., Al Haromainy, M. M., Aditiawan, F. P., Sihananto, A. N., & SJ Saputra, W. (2022, November 29). Analisis Faktor Kesuksesan Penggunaan eBelajar Menggunakan Metode Hot-Fit di STIKI Malang. PROSIDING SEMINAR NASIONAL SAINS DATA, 2(1), 92-102. Retrieved from https://prosiding-senada.upnjatim.ac.id/index.php/senada/article/view/54

Ulfah, A. N., & Anam, M. K. (2020). Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM). JATISI (Jurnal Teknik Informatika dan Sistem Informasi), 7(1), 1-10. https://doi.org/10.35957/jatisi.v7i1.196

Widodo, P. P., Handayanto, R. T., & Herlawati, H. (2013). Penerapan data mining dengan Matlab. *Bandung: Rekayasa Sains*.