PENERAPAN ALGORITMA K-NEAREST NEIGHBORS (KNN) DALAM ANALISIS SENTIMEN FILM 'GADIS KRETEK' DI TWITTER

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Informatika

disusun oleh

GHALUH BHELBY ENDENA

20.11.3524

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

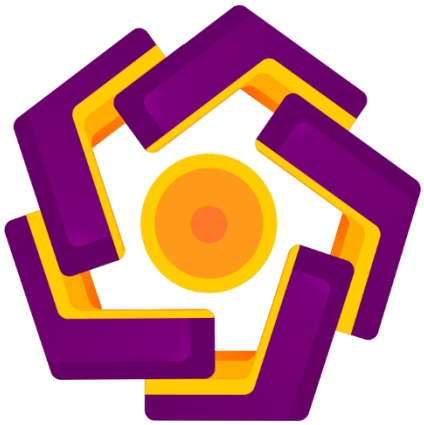
2024

PENERAPAN ALGORITMA K-NEAREST NEIGHBORS (KNN) DALAM ANALISIS SENTIMEN FILM 'GADIS KRETEK' DI TWITTER

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Informatika



disusun oleh

GHALUH BHELBY ENDENA

20.11.3524

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2024

HALAMAN PERSETUJUAN

SKRIPSI

Penerapan Algoritma K-Nearest Neighbors (KNN) dalam Analisis Sentimen Film 'Gadis Kretek' di Twitter

yang disusun dan diajukan oleh

GHALUH BHELBY ENDENA

20.11.3524

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Agustus 2024

Dosen Pembimbing,

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

HALAMAN PENGESAHAN

SKRIPSI

Penerapan Algoritma K-Nearest Neighbors (KNN) dalam Analisis Sentimen Film 'Gadis Kretek' di Twitter

yang disusun dan diajukan oleh

GHALUH BHELBY ENDENA

20.11.3524

Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Agustus 2024

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Rizqi Sukma Kharisma, M.Kom **NIK. 190302215**	
Ike Verawati, M. Kom **NIK. 190302237**	
Hanif Al Fatta, S.Kom., M.Kom., Ph.D. **NIK. 190302096**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 12 Agustus 2024

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ghaluh Bhelby Endena
NIM : 20.11.3524

Menyatakan bahwa Skripsi dengan judul berikut:

Penerapan Algoritma K-Nearest Neighbors (KNN) dalam Analisis Sentimen Film 'Gadis Kretek' di Twitter

Dosen Pembimbing : Hanif Al Fatta,S.Kom., M.Kom., Ph.D.

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, 12 Agustus 2024

Yang Menyatakan,

Ghaluh Bhelby Endena

HALAMAN PERSEMBAHAN

Pertama saya ucapkan puji syukur kepada Allah SWT atas segala nikmat dan kasih sayangnya. Skripsi ini penulis persembahkan untuk:

1. Bapak dan Mama yang sangat saya cintai, di mana telah memberikan kasih sayang yang berlimpah terhadap penulis sehingga bisa membuat penulis bertahan hidup sampai sekarang. Terimakasih telah menjadi orangtua yang sangat hebat.
2. Masba selaku Kakak penulis yang sudah menopang dan menyupport hidup saya selama ini, terimakasih atas jasamu yang sangat besar.
3. Bapak Hanif Al Fatta, M.Kom., Ph.D. selaku dosen pembimbing saya yang sangat baik dan mau membimbing saya selama ini. Terima kasih atas setiap waktu, tenaga, dan perhatian yang Bapak berikan. Semoga segala kebaikan Bapak mendapatkan balasan yang setimpal dari Tuhan Yang Maha Esa.
4. Untuk Seli, Ranya, Ian, Hapis, Rangga alias sahabat di bangku kuliah yang telah menemani *ups and downs* dari hidup penulis, terimakasih telah mendengarkan keluh-kesah serta mengerti dan memahami penulis setiap saat.
5. Untuk Rachel, Istianingsih, dan Eyin alias sahabat penulis dari tempat asal yang masih tetap *reach out* dan membantu penulis selama ini.
6. Untuk Tim Pascasarjana Amikom yang turut memotivasi penulis.
7. Untuk pemilik NIM 20330046 yang telah menemani dan membantu kehidupan penulis hingga jadi berwarna, terimakasih.
8. *Last but not least, I wanna thank me, I wanna thank me for believing in me, I wanna thank me for doing all this hard work, I wanna thank me for having no days off, I wanna thank me for, for never quitting. I wanna thank me all the time.*

KATA PENGANTAR

Segala puji bagi Allah SWT atas limpahan rahmat, petunjuk dan karunia-Nya yang telah membantu peneliti menyelesaikan penyusunan skripsi ini. Skripsi ini disusun sebagai tahapan penting untuk memenuhi persyaratan kelulusan program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Penyusunan skripsi ini tidak lepas dari doa, bantuan, dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada:

1. Keluarga tersayang, yang senantiasa memberikan doa terbaik, dukungan moral, materi, serta semangat dalam perjalanan penyusunan skripsi ini.
2. Hanif Al Fatta, M.Kom., Ph.D., Selaku dosen pembimbing, atas dukungan, arahan, serta masukan berharga yang telah membimbing langkah-langkah penyusunan skripsi ini.
3. Sahabat dan teman, yang telah berbagi ilmu, pengalaman, dan inspirasi sepanjang perjalanan akademik di Universitas Amikom Yogyakarta.
4. Semua individu yang turut menyumbangkan gagasan, pandangan, dan sokongan dalam berbagai bentuk, yang telah membantu kelancaran perjalanan penyelesaian skripsi ini.

Yogyakarta, 12 Agustus 2024

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMBANG DAN SINGKATAN

KNN	K-Nearest Neighbors
API	Application Programming Interface
CSV	Comma-Separated Values
K-fold CV	K-fold Cross Validation
TfidfTransformer	Term Frequency-Inverse Document Frequency Transformer
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative

DAFTAR ISTILAH

Vektor	besaran yang mempunyai arah

INTISARI

Penelitian ini bertujuan untuk menganalisis sentimen terhadap film 'Gadis Kretek' di Twitter dengan menggunakan algoritma K-Nearest Neighbors (KNN). Data dikumpulkan dari Twitter yang mencakup 13.180 tweet, dengan proses pre-processing untuk memastikan data bersih dan konsisten. Komentar dalam bahasa Indonesia dipisahkan menjadi 2.227 komentar negatif dan 2.105 komentar positif. Model KNN diterapkan pada data yang telah diproses menggunakan TF-IDF untuk transformasi teks menjadi fitur numerik. Evaluasi model dilakukan dengan cross-validation untuk menentukan nilai K yang optimal, yang ditemukan sebesar 48 dengan akurasi 69,60%. Pencarian parameter terbaik menggunakan RandomizedSearchCV menunjukkan bahwa parameter optimal adalah {'knn__weights': 'uniform', 'knn__metric': 'euclidean'}, yang menghasilkan akurasi 70,30% pada data uji.

Eksperimen dengan berbagai ukuran data uji (30%, 20%, dan 10%) menunjukkan bahwa ukuran data uji 30% memberikan akurasi 71% dengan recall tinggi untuk komentar negatif (84%) dan recall lebih rendah untuk komentar positif (56%). Precision untuk komentar positif adalah 76% dan untuk komentar negatif 68%. Ukuran data uji 20% menghasilkan akurasi 70% dengan recall 55% untuk komentar positif dan 85% untuk komentar negatif, serta precision 77% untuk komentar positif dan 67% untuk komentar negatif. Ukuran data uji 10% menunjukkan akurasi meningkat menjadi 71,43%. Hasil penelitian ini menekankan pentingnya menemukan keseimbangan antara ukuran data pelatihan dan data uji untuk memperoleh model KNN yang stabil dan akurat.

Kata Kunci: Analisis Sentimen, K-Nearest Neighbors, TF-IDF, Cross-Validation, RandomizedSearchCV

ABSTRACT

This study aims to analyze sentiments towards the film 'Gadis Kretek' on Twitter using the K-Nearest Neighbors (KNN) algorithm. Data was collected from Twitter, consisting of 13,180 tweets, and underwent pre-processing to ensure data cleanliness and consistency. Comments in Indonesian were categorized into 2,227 negative and 2,105 positive comments. The KNN model was applied to the processed data using TF-IDF for text transformation into numerical features. Model evaluation was performed with cross-validation to determine the optimal K value, found to be 48 with an accuracy of 69.60%. Parameter tuning using RandomizedSearchCV indicated that the optimal parameters were {'knn__weights': 'uniform', 'knn__metric': 'euclidean'}, resulting in an accuracy of 70.30% on the test data.

Experiments with various test sizes (30%, 20%, and 10%) revealed that a 30% test size yielded an accuracy of 71% with high recall for negative comments (84%) and lower recall for positive comments (56%). Precision for positive comments was 76%, while for negative comments it was 68%. The 20% test size resulted in an accuracy of 70% with a recall of 55% for positive comments and 85% for negative comments, and precision of 77% for positive comments and 67% for negative comments. The 10% test size showed an accuracy increase to 71.43. The findings underscore the importance of balancing training and test data sizes to achieve a stable and accurate KNN model.

***Keywords**: Sentiment Analysis, K-Nearest Neighbors, TF-IDF, Cross-Validation, RandomizedSearchCV*