PERBANDINGAN METODE *RANDOM FOREST* DAN *K-NEAREST NEIGHBORS* PADA ANALISIS SENTIMEN PENGGUNA TWITTER MENGENAI PROMO GOJEK

SKRIPSI

ASHIL HAFIDH DHIYA

1810511121

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2023

PERBANDINGAN METODE *RANDOM FOREST* DAN *K-NEAREST NEIGHBORS* PADA ANALISIS SENTIMEN PENGGUNA TWITTER MENGENAI PROMO GOJEK

SKRIPSI

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Komputer

ASHIL HAFIDH DHIYA

1810511121

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2023

PERNYATAAN BEBAS PLAGIAT

PERNYATAAN BEBAS PLAGIAT

Skripsi ini adalah hasil karya sendiri dan sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Ashil Hafidh Dhiya

NIM : 1810511121

Tanggal : 30 November 2022

Judul Skripsi : Perbandingan Metode Random Forest dan K-Nearest Neighbors pada Analisis Sentimen Pengguna Twitter mengenai Promo Gojek

Bilamana dikemudian hari ditemukan ketidaksamaan dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 30 November 2022

Ashil Hafidh Dhiya

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan dibawah ini :

Nama : Ashil Hafidh Dhiya

NIM : 1810511121

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

Perbandingan Metode *Random Forest* dan *K-Nearest Neighbors* pada Analisis Sentimen Pengguna Twitter mengenai Promo Gojek

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih-media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas akhir saya selama tetap mencamtumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 30 November 2022

Yang menyatakan,

(Ashil Hafidh Dhiya)

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama : Ashil Hafidh Dhiya

NIM : 1810511121

Program Studi : S1 Informatika

Judul Tugas Akhir : Perbandingan Metode Random Forest dan K-Nearest Neighbors pada Analisis Sentimen Pengguna Twitter mengenai Promo Gojek

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Dr. Ermatita, M. Kom.
Penguji I

Helena Nurramdhani Irmanda, S.Pd., M.Kom.
Penguji II

Ati Zaidiah, S.Kom, MTI.
Pembimbing

Dr. Ermatita, M.Kom
Dekan

Dr. Widya Cholil, S.Kom., M.I.T.
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 11 Januari 2023

PERBANDINGAN METODE *RANDOM FOREST* DAN *K-NEAREST NEIGHBORS* TERHADAP ANALISIS SENTIMEN PADA TWITTER MENGENAI PROMO GOJEK

Ashil Hafidh Dhiya

ABSTRAK

Layanan transportasi *online* merupakan salah satu topik yang sedang hangat dibicarakan. Banyak orang mengandalkan lalu lintas *online*, atau pengemudi dan pelanggan. Transportasi *online* menjadi topik yang ramai diperbincangkan karena kebutuhan transportasi umum di beberapa daerah sulit terpenuhi. Respon masyarakat terhadap pelayanan yang diberikan oleh penyedia jasa transportasi *online* bervariasi, ada yang memberikan hasil positif dan negatif. Analisis sentimen merupakan salah satu cara untuk memahami pendapat seseorang atau sekelompok orang. Data berupa tweet akan dikumpulkan melalui Twitter sebanyak 453 tweet dan dibagi dua menjadi 80% data latih dan 20% data uji. Setelah itu data akan di *preprocessing* menggunakan *case folding*, *cleansing data*, normalisasi, *stemming*, *stopword removal* dan tokenisasi. Setelah data di *preprocess*, maka data akan diberi *term weighting* menggunakan TF-IDF. Setelah data diberi bobot, data akan diolah menggunakan *K-Nearest Neighbors* dan *Random Forest*. Penelitian ini diharapkan bisa mendapatkan informasi akan sentimen opini publik terhadap promo Gojek serta mengetahui performa metode *K-Nearest Neighbors* dan *Random Forest.* Pada pengujian pertama yaitu menggunakan metode *K-Nearest Neighbors* diperoleh nilai akurasi sebesar 75%, nilai presisi sebesar 80%, nilai recall sebesar 64%, dan nilai *f-1 score* sebesar 71%. Pengujian kedua yaitu dengan menerapkan metode Random Forest dan diperoleh nilai akurasi sebesar 77%, nilai presisi sebesar 75%, nilai recall sebesar 77%, dan nilai *f-1 score* sebesar 75%. Berdasarkan perolehan hasil evaluasi tersebut, metode *Random Forest* lebih baik dibandingkan dengan metode *K-Nearest Neighbors* dengan nilai akurasi sebesar 77%.

Kata Kunci: Promo Gojek, Analisis Sentimen, *K-Nearest Neighbors*, *Random Forest*

COMPARISON OF RANDOM FOREST AND K-NEAREST NEIGHBORS METHODS TOWARDS ANALYSIS OF SENTIMENT ON TWITTER ABOUT GOJEK PROMOTION

Ashil Hafidh Dhiya

ABSTRACT

Online transportation services are one of the hotly discussed topics. Many people rely on online traffic, or drivers and customers. Online transportation is a topic that is often discussed because public transportation needs in some areas are difficult to meet. Public responses to the services provided by online transportation service providers vary, some giving positive and negative results. Sentiment analysis is one way to understand the opinion of a person or group of people. Data in the form of tweets will be collected via Twitter as many as 453 tweets and divided into 80% training data and 20% test data. After that, the data will be preprocessed using case folding, data cleansing, normalization, stemming, stopword removal and tokenization. After the data is preprocessed, the data will be given term weighting using TF-IDF. After the data is weighted, the data will be processed using K-Nearest Neighbors and Random Forest. This research is expected to be able to obtain information on the sentiment of public opinion towards the Gojek promo and determine the performance of the K-Nearest Neighbors and Random Forest methods. In the first test, using the K-Nearest Neighbors method, the accuracy value is 75%, the precision value is 80%, the recall value is 64%, and the f-1 score is 71%. The second test is by applying the Random Forest method and obtained an accuracy value of 77%, a precision value of 75%, a recall value of 77%, and an f-1 score of 75%. Based on the evaluation results, the Random Forest method is better than the K-Nearest Neighbors method with an accuracy value of 77%.

Keyword: *Gojek Promotion, Sentiment Analysis, K-Nearest Neighbors, Random Forest*.

KATA PENGANTAR

Puji dan syukur penulis panjatkan atas berkat, rahmat dan karunia-Nya kepada Allah SWT, penulis dapat menyelesaikan Tugas Akhir/Skripsi yang berjudul "Perbadingan Metode Random Forest dan K-Nearest Neighbors pada Analisis Sentimen Pengguna Twitter mengenai Promo Gojek" dengan baik.

Dalam penyelesaian tugas akhir ini, penulis tidak lupa mengucapkan terima kasih kepada pihak-pihak yang telah memberikan dukungan dan masukan kepada penulis, serta membantu penulis dalam merealisasikan tugas akhir ini, pihak-pihak tersebut adalah :

1. Kedua orang tua penulis beserta keluarga yang selalu memberikan doa dan dukungan penuh sehingga dapat menyelesaikan tugas akhir ini.
2. Ibu Ati Zaidiah S.Kom., MTI. Selaku dosen pembimbing yang telah memberikan saran dan masukan yang sangat bermanfaat.
3. Bapak/Ibu dosen Informatika Universitas Pembangunan Nasional Veteran Jakarta, terima kasih atas ilmu-ilmu yang selama ini sudah diajarkan.
4. Teman-teman seperti Nadhifa Zhafira, Vanesa, Hanif Razka, Kavindra Razik Afif, Nabila Adhari, Daniel Manurung, Krisna Jonathan, serta teman-teman terdekat yang tiada henti membantu dan memberikan dukungan dalam setiap proses penulisan tugas akhir.
5. Seluruh pihak yang terlibat dalam kelancaran pembuatan tugas akhir ini dan yang belum disebutkan di atas, penulis ucapkan terima kasih.

Akhir kata, semoga tugas akhir/skripsi ini dapat bermanfaat bagi para pembacanya.

Jakarta, 30 November 2022

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN