PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR, RANDOM FOREST, DAN LOGISTIC REGRESSION TERHADAP KATEGORI REVIEW PADA MARKETPLACE XYZ

SKRIPSI

ARIA NANDA HERDIAWAN

1910511083

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

2023

PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR, RANDOM FOREST, DAN LOGISTIC REGRESSION TERHADAP KATEGORI REVIEW PADA MARKETPLACE XYZ

SKRIPSI

Diajukan Sebagai Salah Satu Syarat Memperoleh Gelar

Sarjana Komputer

ARIA NANDA HERDIAWAN

1910511083

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

2023

PERNYATAAN ORISINALITAS

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

Nama	:	Aria Nanda Herdiawan
NIM	:	1910511083
Tanggal	:	19 Januari 2024

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

Jakarta, 19 Januari 2024

(Aria Nanda Herdiawan)

LEMBAR PERSETUJUAN PUBLIKASI

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertandatangan di bawah ini:

Nama	:	Aria Nanda Herdiawan
NIM	:	1910511083
Fakultas	:	Ilmu Komputer
Program Studi	:	S1 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 ALGORITMA K-NEAREST NEIGHBOR, RANDOM FOREST, DAN LOGISTIC REGRESSION TERHADAP KATEGORI REVIEW PADA MARKETPLACE XYZ

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 kata (basis data), merawat dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 19 Januari 2024

Yang menyatakan,

(Aria Nanda Herdiawan)

LEMBAR PENGESAHAN

Tugas akhir ini diajukan oleh:

Nama	:	Aria Nanda Herdiawan
NIM	:	1910511083
Program Studi	:	S1 Informatika
Judul Tugas Akhir	:	Perbandingan Algoritma K-Nearest Neighbor, Random Forest, dan Logistic Regression Terhadap Kategori Review pada Marketplace XYZ

Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

(Theresia Wati, S.Kom., MTI.)
Penguji I

(Anita Muliawati, S.Kom., MTI.)
Penguji II

(Jayanta, S. Kom., M. Si)
Pembimbing

(Dr. Ir. Krisnanik S.Kom., MM.)
Plt Dekan

(Dr. Widya Cholil, M.I.T.)
Kepala Program Studi

Ditetapkan di : Jakarta
Tanggal Ujian : 20 Desember 2023

Perbandingan Algoritma K-Nearest Neighbor, Random Forest, dan Logistic Regression Terhadap Kategori Review pada Marketplace XYZ

Aria Nanda Herdiawan

ABSTRAK

Dengan perkembangan teknologi di dunia yang sangat cepat, dapat mempermudah dalam melakukan segala hal. Salah satunya adalah berbelanja secara *online*. Salah satu platform yang dapat digunakan untuk berbelanja *online* adalah *marketplace*. Banyak *review* yang diberikan pembeli ketika selesai melakukan transaksi pada *marketplace*. Sehingga dibutuhkan klasifikasi untuk membedakan kategori *review* pembeli, selain itu dapat diketahui perbedaan kinerja dari algoritma klasifikasi *K-nearest neighbor, random forest,* dan *logistic regression*. Dengan menggunakan data *review* sebanyak 1000 data yang dibagi menjadi kelas produk 618 data, pengiriman 214 data, dan pelayanan 168 data. Data dilakukan *preprocessing* dengan tahap *case folding, cleansing,* normalisasi, tokenisasi, *filtering,* dan *stemming.* Kemudian dihitung bobot kata pada data dengan *tf-idf*. Data akan dibagi menjadi *data train* dan *data test*. Setelah itu dilakukan klasifikasi dengan *K-nearest neighbor, random forest,* dan *logistic regression*. Yang menghasilkan akurasi tertinggi dari *K-nearest neighbor* 76%, *random forest* 81%, dan *logistic regression* 80%.

Kata Kunci: *Review*, *Marketplace*, Klasifikasi Teks, *K-Nearest Neighbor, Random Forest, Logistic Regression*

ABSTRACT

With the rapid development of technology in the world, it can make doing everything easier. One of them is shopping online. One platform that can be used for online shopping is a marketplace. Many reviews are given by buyers when completing transactions in marketplace. So classification is needed to distinguish categories of buyer reviews, besides that we can know the difference performance of the K-nearest neighbor, random forest, and logistic regression classification algorithms. By using review data of 1000 data which is divided into product classes 618 data, delivery 214 data, and service 168 data. The data was preprocessed using case folding, cleansing, normalization, tokenization, filtering, and stemming stages. Then the word weights in the data are calculated using tf-idf. The data will be divided into train data and test data. After that, classification was carried out with K-nearest neighbor, random forest, and logistic regression. Which produces the highest accuracy of K-nearest neighbor 76%, random forest 81%, and logistic regression 80%.

Key Word: Review, Marketplace, Text Classification, K-Nearest Neighbor, Random Forest, Logistic Regression

KATA PENGHANTAR

Puji dan syukur selalu tercurahkan kepada Allah SWT atas nikmat dan karunia yang diberikan-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul "Perbandingan Algoritma K-Nearest Neighbor, Random Forest, dan Logistic Regression Terhadap Kategori Review pada Marketplace XYZ".

Dalam penulisan skripsi ini tidak mungkin selesai tanpa adanya dukungan, bantuan, saran, serta bimbingan dari berbagai pihak. Oleh karena itu penulis ingin memberikan terimakasih pada berbagai pihak yang telah membantu selama proses penulisan skripsi ini, yaitu kepada:

1. Bapak Andy Sosiawan dan Ibu Heti Herliati selaku orang tua penulis yang selalu memberikan semangat serta doa selama pengerjaan skripsi ini.
2. Bapak Jayanta, S. Kom., M. Si. selaku dosen pembimbing skripsi, yang telah membantu dan memberikan saran selama proses penulisan skripsi ini.
3. Bapak Hamonangan Kinantan Prabu, M.T. selaku dosen pembimbing akademik penulis.
4. Ibu Dr. Widya Cholil, M.I.T selaku Kaprodi Informatika Universitas Pembangunan Nasional Veteran Jakarta.
5. Luthfiyah Putri Ayuningtyas dan Ayuni Fitria selaku teman penulis yang telah membantu dalam proses pelabelan data.
6. Serta keluarga, kerabat, dan rekan penulis yang telah membantu dan memberikan dukungan selama penulisan skripsi ini.

Tangerang Selatan, 19 Januari 2023

Penulis,

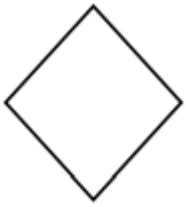
Aria Nanda Herdiawan

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR SIMBOL

GAMBAR	NAMA	KETERANGAN
	*Terminator*	Melambangkan awal dimulai dan akhir dari proses
	*Flow*	Melambangkan penghubung antar simbol untuk alur proses
	*Data*	Melambangkan data *input* atau *output* dalam proses
	*Process*	Melambangkan proses yang dijalankan
	*Decision*	Melambangkan percabangan proses yang memiliki putusan yang berbeda