IMPLEMENTASI ALGORITMA *RANDOM FORESTS* UNTUK KLASIFIKASI *SPAM* PADA CITRA DAN *TEXT* INSTAGRAM®

TUGAS AKHIR

RIZKY NOVRIYEDI PUTRA

1132001001

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK DAN ILMU KOMPUTER

UNIVERSITAS BAKRIE

JAKARTA

2017

IMPLEMENTASI ALGORITMA *RANDOM FORESTS* UNTUK KLASIFIKASI *SPAM* PADA CITRA DAN *TEXT* INSTAGRAM®

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer

RIZKY NOVRIYEDI PUTRA

1132001001

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK DAN ILMU KOMPUTER

UNIVERSITAS BAKRIE

JAKARTA

2017

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.

Nama	**: Rizky Novriyedi Putra**
NIM	**: 1132001001**
Tanda Tangan	**:** 
Tanggal	**: 12 September 2017**

...

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama	: Rizky Novriyedi Putra
NIM	: 1132001001
Program Studi	: Informatika
Fakultas	: Teknik dan Ilmu Komputer
Judul Skripsi	: Implementasi Algoritma *Random Forests* Untuk Klasifikasi *Spam* Pada Citra dan *Text* Instagram

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik dan Ilmu Komputer Universitas Bakrie

DEWAN PENGUJI

Pembimbing : Guson P. Kuntarto, ST, Msc. ( )

Penguji I : Berkah I. Santoso, S.T, M.T.I. ( )

Penguji II : Prof. Dr. Hoga Saragih, ST, MT. ( )

Ditetapkan di : Jakarta

Tanggal : 12 September 2017

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan di bawah ini:

Nama	: Rizky Novriyedi Putra
NIM	: 1132001001
Program Studi	: Informatika
Fakultas	: Teknik dan Ilmu Komputer

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

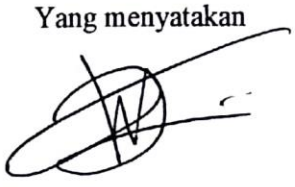
Implementasi Algoritma *Random Forests* Untuk Klasifikasi *Spam* Citra dan *Text* Instagram

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 12 September 2017

Yang menyatakan



(Rizky Novriyedi Putra)

IMPLEMENTASI ALGORITMA *RANDOM FORESTS* UNTUK KLASIFIKASI *SPAM* PADA CITRA DAN *TEXT* INSTAGRAM

Rizky Novriyedi Putra

ABSTRAK

Pada saat sekarang ini media sosial memegang peranan penting dalam berkomunikasi dan berbagi pengalaman, Instagram merupakan sebuah *platform* media sosial yang menggunakan gambar, dan *text* dalam komunikasi sesama pengguna. Informasi yang tersebar di Instagram dapat berupa promosi barang, dan konten yang tidak berhubungan dengan penerima dan lain sebagainya, hal tesebut merupakan *spam* yang dapat menggangu kenyamanan. Permasalahan tersebut dapat dicegah dengan mengklasifikasi data gambar dan *text* pada Instagram kedalam kategori *spam* dan *non-spam*. Untuk melakukan hal tersebut diperlukannya sebuah *dataset* yang dapat diproses menggunakan metode *gray level coocurance matrix* (GLCM) untuk data gambar dan *term frequency invers document frequency* (TF/IDF) untuk data *text*. Proses klasifikasi pada penelitian ini menggunakan 250 data *spam* dan 250 data *non-spam*, dimana masing-masing *dataset images* dan *dataset text* berjumlah 500 data. Algortima *random forests* digunakan sebagai metode klasifikasi dalam *machine learning* yang memiliki tahapan *random feature selection*, dan *bootstrap aggrigation* dalam pembentukan model, untuk proses klasifikasi *random forests* menggunakan *majority vote*, hasil penelitian ini memiliki *f-measure* rata-rata 70% untuk *dataset* gambar dan rata-rata 60% untuk *dataset text* dengan melakukan perubahan parameter *bootstrap aggregation* menjadi 3 diantaranya 1/2, 1/6, 2/3 data dari *dataset*.

Kata Kuci: Instagram, *Gray Level Coocurance Matrix* (GLCM), *Term Frequency/ Invers Document Frequency* (TF/IDF), *Machine Learning*, *Random Forests*, *Random Feature Selection*, *Bagging*, *Majority Vote*.

IMPLEMENTATION OF RANDOM FORESTS ALGORITHM FOR SPAM CLASSIFICATION ON IMAGE AND TEXT INSTAGRAM

Rizky Novriyedi Putra

ABSTRACT

Today, social media is a important part of communication and sharing experience to the people in network. One of that is Instagram, a social media platform that using picture and text in communication to each other user. Information spread on the Instagram have a many variation like a promotion, and content that doesn't related to other user and etc, that will disturb the comfort to some user. The problems can be solved with classification the data of text and picture to spam and non-spam category. To do that, we'll need some dataset that can be processed with method of Gray Level Concurrence Matrix (GLCM) to data picture and Term Frequency invers Document Frequency (TF/IDF) to data text. The process classification on this study using 250 spam and 250 non-spam data where in each other data set total is 500 data. Algorithm Random Forest used in classification method for machine learning that have a random feature selection and bootstrap aggregation in forming the model, for process classification random forest used majority vote. The result of this study have Average F-Measure 70% on picture dataset and 60% on text dataset with 3 (three) parameter bootstrap aggregation among others, 1/2, 1/6, 2,3 data from dataset.

Key Words: Instagram, *Gray Level Coocurance Matrix* (GLCM), *Term Frequency/ Invers Document Frequency* (TF/IDF), *Machine Learning*, *Random Forests*, *Random Feature Selection*, *Bagging*, *Majority Vote*.

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR RUMUS

DAFTAR LAMPIRAN