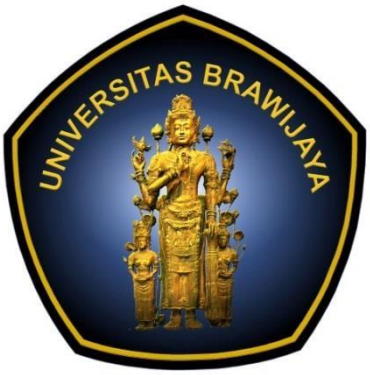
PENERAPAN KLASIFIKASI *TWEETS* PADA BERITA TWITTER MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* DAN *QUERY EXPANSION* BERBASIS *DISTRIBUTIONAL SEMANTIC*

SKRIPSI

Untuk memenuhi sebagian persyaratan memperoleh
gelar Sarjana Komputer

Disusun oleh:
Galih Nuring Bagaskoro
NIM: 135150218113027



PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2018

PENGESAHAN

Penerapan Klasifikasi *Tweets* Pada Berita Twitter Menggunakan metode *K-Nearest Neighbor* dan *Query Expansion* berbasis *Distributional Semantic*

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana Komputer

Disusun Oleh :
Galih Nuring Bagaskoro
NIM: 135150218113027

Skripsi ini telah diuji dan dinyatakan lulus pada
12 Januari 2018 Telah diperiksa dan disetujui
oleh:

Dosen Pembimbing I

M. Ali Fauzi S.Kom , M.Kom
NIK. 201502 890101 1 001

Dosen Pembimbing II

Putra Pandu Adikara, S.Kom, M.Kom
NIP. 19850725 200812 1 002

Mengetahui
Ketua Jurusan Teknik Informatika

Tri Astoto Kurniawan, S.T, M.T, Ph.D
NIP. 19710518 200312 1 001

IDENTITAS TIM PENGUJI

PEMBIMBING 1

Mochammad Ali Fauzi, S.Kom, M.Kom

NIK. 201502 890101 1 001

Email: moch.ali.fauzi@ub.ac.id

PEMBIMBING 2

Putra Pandu Adikara, S.Kom, M.Kom

NIP. 19850725 200812 1 002

Email: adikara.putra@ub.ac.id

PENGUJI 1

Rizal Setya Perdana, S.Kom, M.Kom

NIP. 2016039101181001

Email: rizalespe@ub.ac.id

PENGUJI 2

Lailil Muflikhah, S.kom, M.Sc

NIP. 197411132005012001

Email: lailil@ub.ac.id

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsurunsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 12 Januari 2018

Galih Nuring Bagaskoro

NIM: 135150218113027

RIWAYAT HIDUP

Penulis dilahirkan di Tulungagung pada tanggal 11 Juni 1995 dari Ibu yang bernama Sri Wulan Ros Indah dan Ayah bernama Mudjiono Emje. Penulis menyelesaikan pendidikan sekolah dasar di SDN KAMPUNG DALEM VI pada tahun 2007 dengan mengikuti organisasi pramuka dan berbagai ekstrakulikuler yang diikuti. Kemudian melanjutkan ke sekolah menengah pertama di SMP Islam Al-huda Kediri dengan tahun kelulusan 2010 dengan mengikuti organisasi OSIS. Dan melanjutkan studi ke jenjang sekolah menengah atas di SMAN 8 Kota Kediri dengan menyelesaikan masa studi pada tahun 2013

Pada tahun 2018 penulis telah berhasil menyelesaikan pendidikannya di Universitas Brawijaya pada Program Studi Teknik Informatika Jurusan Teknik Informatika Fakultas Ilmu Komputer. Selain itu penulis aktif menjadi Anggota Pengusaha Muda Kediri, serta menjadi aktivis sosial yang bergerak dengan Seni.

"Impossible Is Nothing"

Assalamualaikum wr.wb.

Alhamdulillah, Terimakasih Ya Rabb atas Kehadirat Allah SWT, serta junjungan besar Nabi Muhammad SAW.

Tugas Akhir ini saya persembahkan untuk Orang Tua dan Keluarga Besar saya yang selalu mendukung saya serta kepada seluruh teman-

temanku dan orang terdekat yang tak berhenti memberikan semangat dan motivasi.

ABSTRAK

Galih Nuring Bagaskoro, 2018. Penerapan Klasifikasi *Tweets* pada Berita Twitter Menggunakan Metode *K-Nearest Neighbor* dan *Query Expansion* Berbasis *Distributional Semantic.*

Fakultas Ilmu Komputer Universitas Brawijaya. Pembimbing : Ali Fauzi S.Kom, M.Sc dan Putra Pandu Adikara S.Kom, M.Kom

Penggunaan teks pendek berbasis digital sampai saat ini masih berkembang dan meluas hingga diberbagai media sosial. Media sosial Twitter memiliki fitur kategori jenis informasi melalui *tweets* yang di unggah. Setiap pengelompokan jenis informasi dilakukan agar mempermudah pengguna untuk memanfaatkannya. Tujuan dari penggunaan kategori dalam hal ini klasifikasi, untuk meningkatkan kualitas media sosial dalam pengelompokan kategori isi dari konten yang disediakan. Klasifikasi tradisional sampai saat ini masih digunakan, namun hasil yang diperoleh terkadang tidak maksimal, perlu dilakukan ekspansi kata untuk menambahkan kata kedalam teks agar dapat meningkatkan akurasi. Ekspansi kata digunakan dengan berbasis *distributional semantic* dengan teknik *Euclidean distance* untuk menemukan kata terdekat dari sumber eksternal agar menjadi kueri yang akan ditambahkan ke teks data uji. Dengan menggunakan data uji 105 dan data latih 400, klasifikasi yang menggunakan *K-Nearest Neighbor* dapat memperoleh hasil 90% dengan tetangga terdekat *K*=5. Hasil tersebut sama halnya dengan hasil pengujian yang dilakukan dengan tanpa menggunakan teknik ekspansi kata. Sedangkan pengujian yang dilakukan dengan menambahkan ekspansi kata dengan *threshold* 0,5 dan nilai tertangga terdekat *K-Nearest Neighbor* K=5 memperoleh hasil akurasi 92%.

Kata kunci: *twitter, tweet, ekspansi kata, distributional semantic, euclidean distance, klasifikasi, k-nearest neighbor*

ABSTRACT

Galih Nuring Bagaskoro, 2018. Application of Tweets Classification on Twitter News Using K-Nearest Neighbor Method and Query Expansion Based on Distributional Semantic.

Faculty of Computer Science University of Brawijaya. Advisor : Ali Fauzi, S.Kom. M.Sc and Putra Pandu Adikara, S.Kom. M.Kom

The use of short text based on digital to date is still growing and extending to various social media. Twitter has news features in tweets to represent information representing each type. Each categorization of this type is done to make it easier for users to use it. The purpose of the use of categories in this classification, to evaluate and improve the quality of social media in grouping categories of content of the content provided. Traditional classification is still used today, but the results are sometimes not maximal, it is necessary to expand the word to add words to the text in order to improve the accuracy. Word expansion is used with a semantic-based distributional euclidean distance technique to find the closest word from an external source to be a query to be added to the test data text. Using test data 105 and training data 400, the classification using K-Nearest Neighbor can obtain 90% results with nearest neighbor K=5. These results are similar to the results of tests conducted without using word expansion techniques. While the test is done by adding the expansion of words with threshold 0.5 and the nearest immediate value K-Nearest Neighbor K=5 obtained an accuracy of 92%.

***Keywords**: twitter, tweet, word expansion, distributional semantic, euclidean distance, classification, k-nearest neighbor*

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Penerapan Klasifikasi Tweets pada Berita Twitter Menggunakan metode K-Nearest Neighbor dan Query Expansion Berbasis Distributional Semantic". Skripsi ini diajukan untuk memenuhi prasyarat kelulusan dalam menyelesaikan pendidikan dan memperoleh gelar Sarjana Komputer konsentrasi Komputasi Cerdas dan Visualisasi di program studi Informatika Fakultas Ilmu Komputer Universitas Brawijaya.

Dalam penyelesaian skripsi ini, penulis telah mendapatkan banyak dukungan dan bantuan dari berbagai banyak kalangan pihak. Mungkin hanya ucapan terima kasih dari penulis yang bias membalas bantuan dan dukungan dari beberapa pihak. Atas bantuan yang telah diberikan, penulis ingin menyampaikan sebuah ucapan terima kasih yang sebesar-besarnya kepada :

1. Bapak Ali Fauzi, S.Kom., M.Kom., selaku dosen pembimbing utama yang telah meluangkan waktu untuk memberikan pengarahan dan masukan kepada penulis
2. Bapak Putra Pandu Adikara, S.Kom, M.Kom selaku dosen pendamping kedua yang juga telah meluangkan waktu untuk memeberikan pengarahan dan masukan pada penulis
3. Wayan Firdaus Mahmudy, S.Si., M.T., Ph.D., Ir. Heru Nurwasito, M.Kom., Drs. Mardji M.T, dan Edy Santoso, S.Si., M.Kom., selaku Dekan, wakil Dekan 1, Wakil Dekan 2, Wakil Dekan 3 Fakultas Ilmu Komputer Universitas Brawijaya
4. Tri Astoto Kurniawan, S.T, M.T, Ph.D dan M. Tanzil Furqon, S.Kom, M.CompSc., selaku Ketua dan Sekretaris Program Studi Teknik Informatika Universitas Brawijaya
5. Seluruh Dosen Fakultas Ilmu Komputer Universitas Brawijaya atas kesediaannya membagi ilmu kepada penulis
6. Bapak dan Ibu kandung serta Adik-Adik yang senantiasa selalu memeberikan dukungan dan doa kepada penulis demi kelancaran pengerjaan skripsi.
7. Seorang bernama Alfi Rizky Anita Fajarina yang tidak luput dari bantuan masukan dan inspirasi, serta dukungan terbaiknya yang membuat penulis terus bersemangat.
8. Teman-teman dan para sahabat penulis Wildan Afif, Nasrul Ashar, Cristian Herlando, Dhimas Tungga, Afif Ridwan, Erda Endika, dan Ilham

Zunaidy yang memberikan bantuan, semangat dan doa demi terselesaikannya skripsi ini.

9. Semua pihak yang tidak dapat disebutkan satu persatu oleh penulis yang memberikan dukungan dan semangat secara langsung maupun tidak langsung demi terselesainya skripsi ini.

Meskipun penulis berharap isi dari skripsi ini bebas dari kekurangan dan kesalahan, namun karena ini adalah buatan dari manusia yang dengan kata lain pasti tidaklah sempurna. Oleh karena itu, penulis mengharapkan sebuah kritik dan saran yang dapat membangun agar skripsi ini dapat lebih baik. Akhir kata penulis berharap skripsi ini dapat bermanfaat pagi segala kalangan pembaca.

Malang, 8 Januari 2018

penulis

galihnuring2@gmail.com

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE PROGRAM

DAFTAR PERSAMAAN

DAFTAR LAMPIRAN