KLASIFIKASI KOMENTAR PADA DATASET PEMILU PRESIDEN INDONESIA 2014 DENGAN METODE IMPROVED K-NEAREST NEIGHBOR

Skripsi

oleh

ANDRO ARDIYANTO

71130066

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS KRISTEN DUTA WACANA

2017

KLASIFIKASI KOMENTAR PADA DATASET PEMILU PRESIDEN INDONESIA 2014 DENGAN METODE IMPROVED K-NEAREST NEIGHBOR

Skripsi

Diajukan kepada Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Duta Wacana

Sebagai Salah Satu Syarat dalam Memperoleh Gelar

Sarjana Komputer

Disusun oleh :

ANDRO ARDIYANTO

71130066

PROGRAM STUDI TEKNIK INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS KRISTEN DUTA WACANA

2017

PERNYATAAN KEASLIAN SKRIPSI

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

KLASIFIKASI KOMENTAR PADA DATASET PEMILU PRESIDEN INDONESIA 2014 DENGAN METODE IMPROVED K-NEAREST NEIGHBOR

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Teknik Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 5 Juni 2017

ANDRO ARDIYANTO
71130066

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

Judul Skripsi	:	KLASIFIKASI KOMENTAR PADA DATASET PEMILU PRESIDEN INDONESIA 2014 DENGAN METODE IMPROVED K-NEAREST NEIGHBOR
Nama Mahasiswa	:	ANDRO ARDIYANTO
N I M	:	71130066
Matakuliah	:	Skripsi (Tugas Akhir)
Kode	:	TIW276
Semester	:	Genap
Tahun Akademik	:	2016/2017

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 5 Juni 2017

Dosen Pembimbing I	Dosen Pembimbing II
Yuan Lukito, S.Kom., M.Cs.	Antonius Rachmat C., S.Kom.,M.Cs.

HALAMAN PENGESAHAN

HALAMAN PENGESAHAN

KLASIFIKASI KOMENTAR PADA DATASET PEMILU PRESIDEN INDONESIA 2014 DENGAN METODE IMPROVED K-NEAREST NEIGHBOR

Oleh: ANDRO ARDIYANTO / 71130066

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Teknik Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 29 Mei 2017

Yogyakarta, 5 Juni 2017
Mengesahkan,

Dewan Penguji:

1. Yuan Lukito, S.Kom., M.Cs.
2. Antonius Rachmat C., S.Kom.,M.Cs.
3. Lucia Dwi Krisnawati, Dr.
4. Laurentius Kuncoro Probo Saputra, S.T., M.Eng.

Dekan

(Budi Susanto, S.Kom., M.T.)

Ketua Program Studi

(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Dalam menyelesaikan penelitian tugas akhir ini, penulis telah banyak mendapatkan bimbingan, saran serta dukungan dari berbagai pihak. Oleh karena itu penulis mengucapkan terima kasih kepada :

1. Bapak Budi Susanto, S.Kom., M.T. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
2. Ibu Gloria Virginia, S.Kom., MAI. selaku Ketua Program Studi Teknik Informatika Universitas Kristen Duta Wacana.
3. Bapak Yuan Lukito, S.Kom., M.Cs. selaku dosen pembimbing I yang telah membantu memberikan saran dan arahan dalam penyusunan tugas akhir ini.
4. Bapak Antonius Rachmat, S.Kom., M.Cs. selaku dosen pembimbing II yang telah membantu memberikan saran dan arahan dalam penyusunan tugas akhir ini.
5. Orangtua yang memberikan dukungan baik melalui doa maupun motivasi untuk menyelesaikan tugas akhir ini.
6. Teman-teman terdekat Teknik Informatika 2013 yang memberikan motivasi dan semangat untuk segera menyelesaikan tugas akhir ini.
7. Semua pihak yang tidak dapat disebutkan satu per satu yang ikut memberikan arahan, saran maupun nasihat dalam pembuatan/penelitian tugas akhir ini.

Penulis menyadari dalam penelitian ini terdapat banyak kekurangan, baik dalam penelitian maupun penulisan laporan akhir ini. Akhir kata penulis ingin berterima kasih pada semua pihak yang telah mendukung penulis, dan harapannya tugas akhir ini dapat bermanfaat bagi semua pihak terutama Universitas Kristen Duta Wacana.

Penulis

KATA PENGANTAR

Puji syukur dan terima kasih kepada Tuhan Yang Maha Esa karena anugerah dan bimbingan-Nya, penulis dapat menyelesaikan penelitian tugas akhir dengan judul "KLASIFIKASI KOMENTAR PADA DATASET PEMILU PRESIDEN INDONESIA 2014 DENGAN METODE IMPROVED K-NEAREST NEIGHBOR".

Dalam penulisan laporan tugas akhir ini diajukan guna melengkapi sebagai syarat dalam mencapai gelar sarjana strata satu (S1) di Fakultas Teknologi Informasi Program Studi Teknik Informatika Universitas Kristen Duta Wacana Yogyakarta. Penulis menyadari banyak kekurangan dalam penulisan maupun penelitian ini, namun penulis sudah berusaha semaksimal mungkin dalam pembuatan sistem maupun analisis sistem sesuai dengan tingkat pengetahuan dan kemampuan penulis. Penulis mengharapkan kritik maupun saran yang membangun untuk menyempurnakan tugas akhir ini.

Dalam pengerjaan tugas akhir ini banyak sekali kendala, namun karena berkat Tuhan Yang Maha Esa, laporan dan penelitian ini dapat selesai dengan baik. Penulis juga memberikan banyak terima kasih kepada Bapak Yuan Lukito dan Bapak Antonius Rachmat karena telah membimbing penulis serta mengarahkan penulis untuk memberikan yang terbaik dalam pembuatan program maupun laporan tugas akhir, terima kasih untuk waktu, tenaga dan sumbangan idenya, karena secara langsung maupun tidak langsung, hal tersebut telah membantu penulis dalam menyelesaikan tugas akhir dengan baik.

Penulis

INTISARI

SENTIMEN ANALISIS KOMENTAR PADA DATASET PEMILU PRESIDEN INDONESIA 2014 DENGAN METODE IMPROVED K-NEAREST NEIGHBOR

Perbedaan porsi data latih dari setiap kategori dapat mempengaruhi hasil klasifikasi untuk lebih condong ke arah porsi data latih yang paling besar. Pada algoritma k-*Nearest Neighbor*, nilai k berpengaruh dalam menentukan proses klasifikasi dari suatu data uji. Proses klasifikasinya juga tergantung porsi data terbanyak dari tetangga yang diambil, kemunculan paling banyak pada jumlah tetangga terdekatnya. *Improved KNN* muncul untuk mengatasi hal tersebut. Dalam hal ini *dataset* yang dipakai memiliki jumlah total 2796 data (2406 data positif dan 390 data negatif).

Pertama-tama data uji akan melalui proses *preprocessing* yang terdiri dari (*convert emoticon, cleansing, casefolding, tokenizing, filtering, stemming*). Data tersebut akan diberi bobot sesuai dengan TF-IDF lalu akan dilanjutkan pada proses *cos-similiarity*. Pada proses tersebut akan terjadi pemilihan jumlah tetangga terbesar sesuai nilai k, lalu proses *improvement KNN* dijalankan dan data uji tersebut diklasifikasikan.

Penelitian ini menghasilkan bahwa *Improved KNN* sukses dalam menaikkan akurasi pengklasifikasian. Penggunaan *feature selection* meningkatkan akurasi pada data latih dengan perbedaan 1800 data sebanyak 1,01%, dari 76,52% menjadi 77,53%. Peningkatan akurasi terbesar sebesar 1,48% terjadi pada skenario dengan ketimpangan data latih 900 buah, sedangkan pada data seimbang dan perbedaan data latih 300 buah, tidak terjadi perubahan akurasi jika dibandingkan dengan *Default KNN*.

Kata Kunci : [*text mining, improved k-nn, knn, sentimen analyst*]

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

INTISARI

SENTIMEN ANALISIS KOMENTAR PADA DATASET PEMILU PRESIDEN INDONESIA 2014 DENGAN METODE IMPROVED K-NEAREST NEIGHBOR

Perbedaan porsi data latih dari setiap kategori dapat mempengaruhi hasil klasifikasi untuk lebih condong ke arah porsi data latih yang paling besar. Pada algoritma k-*Nearest Neighbor*, nilai k berpengaruh dalam menentukan proses klasifikasi dari suatu data uji. Proses klasifikasinya juga tergantung porsi data terbanyak dari tetangga yang diambil, kemunculan paling banyak pada jumlah tetangga terdekatnya. *Improved KNN* muncul untuk mengatasi hal tersebut. Dalam hal ini *dataset* yang dipakai memiliki jumlah total 2796 data (2406 data positif dan 390 data negatif).

Pertama-tama data uji akan melalui proses *preprocessing* yang terdiri dari (*convert emoticon, cleansing, casefolding, tokenizing, filtering, stemming*). Data tersebut akan diberi bobot sesuai dengan TF-IDF lalu akan dilanjutkan pada proses *cos-similiarity*. Pada proses tersebut akan terjadi pemilihan jumlah tetangga terbesar sesuai nilai k, lalu proses *improvement KNN* dijalankan dan data uji tersebut diklasifikasikan.

Penelitian ini menghasilkan bahwa *Improved KNN* sukses dalam menaikkan akurasi pengklasifikasian. Penggunaan *feature selection* meningkatkan akurasi pada data latih dengan perbedaan 1800 data sebanyak 1,01%, dari 76,52% menjadi 77,53%. Peningkatan akurasi terbesar sebesar 1,48% terjadi pada skenario dengan ketimpangan data latih 900 buah, sedangkan pada data seimbang dan perbedaan data latih 300 buah, tidak terjadi perubahan akurasi jika dibandingkan dengan *Default KNN*.

Kata Kunci : [*text mining, improved k-nn, knn, sentimen analyst*]

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi merupakan salah satu sarana yang mengalami perkembangan pesat dari waktu ke waktu. Contoh dari perkembangan tersebut adalah semakin meningkatnya aktifitas komunikasi yang dilakukan oleh seseorang melalui dunia maya. Jejaring sosial seperti *facebook, twitter* merupakan salah satu alternatif yang dapat digunakan oleh pengguna untuk mengekspresikan fakta, pandangan dan pendapat mereka secara bebas (Renata, 2012). Dengan fasilitas tersebut, maka pengguna bisa secara aktif dan bebas untuk memposting/menyatakan sesuatu yang mereka suka maupun tidak suka di akun jejaring sosial mereka tanpa terbatas, dan setiap orang yang membaca postingan tersebut juga memiliki hak untuk berkomentar dan ikut andil dalam diskusi/bahasan tersebut. *Posting* maupun komentar yang dilontarkan oleh beberapa orang di jejaring sosial merupakan sebuah informasi teks yang mewakili pendapat seseorang mengenai sebuah topik.

Informasi teks dapat dikategorikan menjadi dua jenis utama yaitu fakta dan opini. Opini atau pendapat biasanya memiliki sifat subyektif antara satu dan lainnya. Kasus yang menjadi perhatian pada tugas akhir ini adalah mengenai komentar/opini yang dilontarkan seseorang terkait pada pemilihan umum Presiden Indonesia 2014 kemarin. Terlebih lagi pemilihan umum Presiden Indonesia 2014 kemarin, menjadi bahan perbincangan rakyat Indonesia, karena kandidat presiden dan wakil presidennya memiliki banyak perbedaan yang mengakibatkan pro dan kontra hingga Mahkamah Konstitusi ikut andil untuk menyelesaikan masalah tersebut (Harli, 2014). Banyak sekali komentar-komentar yang dilontarkan oleh berbagai macam pihak mulai dari pelajar hingga politikus dalam pemilihan umum tersebut, tidak banyak juga yang memberikan komentar-komentar pedas untuk menjatuhkan maupun menghasut

lawannya. Komentar tersebut dapat memiliki sifat yang bermacam-macam yaitu positif dan negatif, sifat komentar tersebut yang menjadi fokus utama dari tugas akhir ini.

Metode *Improved K-Nearest Neighbor* adalah suatu metode yang digunakan untuk membantu mengklasifikasikan jenis-jenis komentar tersebut. Improvisasi yang dilakukan pada algoritma *k-Nearest Neighbor* berupa modifikasi nilai *k* pada *k-Nearest Neighbor,* dimana nilai *k* tersebut akan ditinjau kembali nilainya setelah selesai melakukan perangkingan kesamaan (*similiarity*), dengan menggunakan proporsi kategori pada banyaknya data di data latih. Setelah menhitung nilai *k* baru maka akan diuji kembali sesuai dengan probabilitas data latih, probabilitas kategori terbesar pada data latih menentukan nilai *k* efektif untuk kasus/data uji tersebut. Maka dari itu, dalam tugas akhir ini digunakan metode *Improved k-Nearest Neighbor* untuk mengklasifikasikan opini-opini tersebut tergolong jenis opini/komentar yang berbau positif atau negatif. Dengan menggunakan metode ini diharapkan pengklasifikasian komentar-komentar tersebut memiliki tingkat akurasi yang tinggi.

1.2 Rumusan Masalah

Rumusan masalah dalam tugas akhir ini adalah :

1. Bagaimana mengklasifikasikan suatu komentar tergolong komentar dengan tipe positif atau negatif dengan metode *Improved K-Nearest Neighbor* ?
2. Apakah hasil klasifikasi komentar/opini dengan menggunakan metode *Improved K-Nearest Neighbor* memberikan hasil (*confusion matrix*) yang lebih baik daripada metode *K-Nearest Neighbor (kNN)* ?

1.3 Batasan Masalah

Berdasarkan rumusan masalah diatas maka pembuatan sistem ataupun penelitian akan dibatas sesuai dengan parameter-parameter berikut ini :

1. Dalam melakukan tugas akhir ini, komentar-komentar yang akan diteliti hanya digolongkan menjadi 2 bagian yaitu komentar positif dan komentar negatif.

2. Dari sisi kebahasaan, komentar yang akan diteliti adalah komentar yang mengandung bahasa Indonesia saja.
3. Data sentipol (sentimen politik) tahun 2016 yang telah dilakukan oleh Yuan Lukito dan Antonius Rachmat (http://ti.ukdw.ac.id/~crowd/dataset.php) akan dijadikan acuan dan digunakan dalam pengklasifikasian komentar dalam format **.csv**.

1.4 Tujuan Penelitian

Tujuan penelitian dalam tugas akhir ini adalah mengklasifikasikan komentar dengan metode *Improved K-Nearest Neighbor* dan membandingkan hasil klasifikasi metode *Improved K-Nearest Neighbor* dengan metode *K-Nearest Neighbor* dengan menghitung *confusion matrix*-nya.

1.5 Manfaat Penelitian

Manfaat penelitian dalam tugas akhir ini :

1. Memudahkan pengklasifikasian jenis komentar (positif atau negatif) yang terdapat pada jejaring sosial.
2. Mengetahui tingkat pro dan kontra yang terjadi melalui klasifikasi komentar pada data sentipol (sentimen politik).
3. Mengetahui tingkat kecocokan metode *Improved K-Nearest Neighbor* dalam mengklasifikasikan jenis-jenis komentar pada data sentipol dibandingkan metode *K-Nearest Neighbor (kNN)*.

1.6 Metode Penelitian

1.6.1 Tahap Pengumpulan dan Persiapan Data

Data sentipol (sentimen politik) tahun 2016 yang telah dilakukan oleh Yuan Lukito dan Antonius Rachmat (http://ti.ukdw.ac.id/~crowd/dataset.php) akan dijadikan

acuan dan digunakan dalam pengklasifikasian komentar. *Data* sentipol tersebut memiliki format dokumen *csv,* dan *data* tersebut berjumlah 3400 data (2940 data positif, 337 data negatif, 123 data netral) namun data yang digunakan oleh peneliti berjumlah 2796 data (setelah mengalami proses pemfilteran) yang terdiri dari *data* berlabel positif sebanyak 2406 buah dan *data* berlabel negatif sebanyak 291 buah. *Data* berlabel netral sebanyak 99 buah dijadikan data negatif.

Tahap selanjutnya adalah tahap persiapan data, dokumen tersebut akan dipecah menjadi kata-kata yang memiliki makna sesuai kamus bahasa Indonesia. Pemecahan dokumen menjadi kata-kata tersebut akan melalui *pre-processing* terlebih dahulu. Hasil dari *pre-processing* ini adalah kata yang akan dijadikan acuan dalam melakukan sebuah klasifikasi komentar data uji yang akan diteliti. Kata tersebut nantinya akan diberi bobot sesuai dengan frekuensi kemunculannya dalam suatu dokumen data latih.

1.6.2 Tahap Implementasi Sistem

Pada tahap implementasi ini, penelitian akan diterapkan pada sebuah *web*. Pengguna akan memasukkan dokumen yang akan diuji, dan *web* akan menganalisa setiap kata dari dokumen tersebut. Nilai k paling optimal diimplementasikan pada sistem berdasarkan hasil pengujian. Hasil akhirnya *web* akan menampilkan hasil klasifikasi secara sistem apakah dokumen tersebut digolongkan menjadi data positif maupun data negatif. Jika data yang diinputkan lebih dari 1 pada 1 proses penginputan, maka data tersebut akan diuji dengan menggunakan *confusion matrix*, namun jika tidak maka akan diberikan *detail* klasifikasi, mulai dari *df*, nilai *idf*, *term*, nilai k dan *cos-similiarity.*

1.6.3 Tahap Pengujian Sistem

Dalam melakukan pengujian sistem, akan dilakukan 7 skenario. Ketujuh skenario tersebut memiliki tujuan apakah porsi data latih untuk masing-masing kategori memberikan pengaruh terhadap pengklasifikasian/pengujian. Skenario I merupakan data apa adanya, sedangkan skenario lainnya diberikan jarak sekitar 300 data pada data latih positif, sedangkan data negatifnya tetap 300. Pengujian dilakukan dari sejumlah data apa adanya (besar) hingga seimbang. Untuk data uji, data negatifnya diberikan data sejumlah 90 data, sedangkan data positifnya bervariasi dari besar hingga seimbang pada skenario ketujuh.

1.6.4 Tahap Analisis Sistem dan Pengambilan Kesimpulan

Pada akhir penelitian akan diadakan analisis *data* berdasarkan *precision*, *recall*, *F-measure* dan *Accuracy* dari berbagai skenario yang akan ditawarkan. Setiap skenario tersebut memiliki jumlah *data* latih dan nilai *k* yang berbeda, dari setiap skenario tersebut, akan disajikan nilai *precision*, *recall*, *F-measure* dan *Accuracy* dari masing-masing skenario.

1.7 Sistematika Penulisan

Bab 1, Pendahuluan, bab ini berisi gambaran umum mengenai penelitian penulis yang terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistemika penelitian.

Bab 2, Tinjauan pustaka dan landasan teori, tinjauan pustaka akan membahas mengenai jurnal/*paper* yang berkaitan dengan penelitian tersebut, dalam tinjauan pustaka juga berisi mengenai hasil akhir/kesimpulan dari masing-masing jurnal/*paper*

tersebut. Landasan teori berisi mengenai konsep, teori maupun rumus-rumus yang mendukung proses penelitian.

Bab 3, Perancangan sistem, bab ini membahas rancangan sistem yang dibangun mulai dari spesifikasi sistem, rancangan diagram sistem, rancangan antar-muka sistem dan tahap-tahapan yang berkaitan dengan proses dan pembuatan sistem tersebut.

Bab 4, Implementasi dan analisis sistem, bab ini akan menguraikan hasil implementasi dari metode-metode yang digunakan pada penelitian penulis dan analisis sistem secara teoritis berdasarkan *confusion matrix*.

Bab 5, Kesimpulan dan saran, bab ini akan membahas mengenai hasil analisis dari penelitian yang sudah dilakukan oleh penulis. Penulis juga akan memberikan saran yang mendukung supaya penelitian tersebut menjadi lebih baik. Dan diharapkan dari saran tersebut dapat memperbaiki kinerja sistem tersebut.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

1. *Improved KNN* menaikkan rata-rata *accuracy* 0,49% dibandingkan dengan *Default KNN* untuk seluruh skenario. Peningkatan terbesar sebesar 1.48% terjadi pada ketimpangan data latih 900 data. Tidak terjadi peningkatan akurasi pada ketimpangan data latih 300 data maupun data latih yang seimbang.
2. Pencarian k optimal dari range 1 - 25 menghasilkan k optimal sebesar 22. Pencarian *feature selection* optimal dari *range* 10% - 100% menghasilkan *feature selection* optimal sebesar 20%. Penggunaan *feature selection* 20% dapat meningkatkan *accuracy* klasifikasi sebanyak 1,01% dari 76,52% menjadi 77,53%.
3. Rumus pengubah nilai k pada *Improved KNN* sukses mengecilkan nilai k untuk kategori tertentu dengan nilai k minimal 2 pada skenario pertama. Nilai k tersebut digunakan pada kategori yang memiliki porsi yang lebih kecil, namun pengklasifikasian juga dipengaruhi/ditentukan dari nilai *cos-similiarity* terbesar dan jumlah tetangga yang akan dijadikan patokan.

5.2 Saran

Pada data latih sentipol di *social media* terdapat kata-kata singkatan maupun kata khas perpolitikan yang tidak terdapat pada KBBI maka penulis menyarankan untuk pada penelitian selanjutnya, data latih tersebut memiliki kamus singkatan maupun istilah perpolitikan sehingga kata-kata pada data latih dapat dipakai semua, karena bisa saja kata-kata tersebut berpengaruh dalam proses klasifikasi.

DAFTAR PUSTAKA

Baoli, L., Shiwen, Y., & Quin, L. (2003). An Improved k-Nearest Neighbor Algorithm for Text Categorization. *Proceedings Of The 20Th International Conference On Computer Processing Of Oriental Languages*.

Bose, S., Badawy, A., Baili, J., Changalasetty, S., Ghribi, W., & Bangali, H. . (2016). A Hybrid GA/kNN/SVM Algorithm for Classification of data. *Biohouse Journal Of Computer Science, 2(2)*.

C, R., Antonius. & Lukito, Y. . (n.d.). Implementasi Sistem Crowdsourced Labelling Berbasis Web dengan Metode Weighted Majority Voting. *Jurnal Sistem Informasi, 6(2)*, pp. 76-139.

Chopra, D., Joshi N., Mathur I. (2016). *Mastering Natural Language Processing with Python.* Livery Place: Packt Publishing Ltd.

Feldman, R. & Sanger, J. (n.d.). *The text mining handbook.* Cambridge: Cambridge University Press.

Herdiawan. (n.d.). Analisis Sentimen terhadap Telkom Indihome berdasarkan Opini Publik menggunakan Metode Improved K-Nearest Neighbor. *urnal Ilmiah Komputer Dan Informatika (KOMPUTA)*, 2-8.

Indriati, & Ridok, A. (2016). Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted k-Nearest Neighbor (NWKNN). *Journal of Enviromental Engineering & Sustainable Technology, 03*, 23-32.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining.* Morgan & Claypoll Publishers.

Maatwebsite/Laravel-Excel. (n.d.). Retrieved from www.github.com: https://github.com/Maatwebsite/Laravel-Excel

Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to Information Retrieval.* Cambridge University Press.

Manning, C.D., Schutze, H. (2002). *Foundation Of Statistical Natural Language Processing.* United States of America: Fifth printing.

Memon, A. (2017). *Advances in Computers, Volume 105.* Academic Press.

Muin, H. (2014). *Selesaikan Pro dan Kontra, MK Cari Selamat.* Retrieved from www.kompasiana.com: http://www.kompasiana.com/harli/selesaikan-pro-dan-kontra-mk-cari-selamat_552bc9096ea834da148b458f

Mustafa, Atika, Ali Akbar, Ahmer Sultan. (April). Knowledge Discovery using Teks Mining : A Programmable Implementation on Information Extraction and Categoritazion. *International Journal of Multimedia Ubiquitous Engineering Vol 4.No.2*.

Nugues, P. M. (2006). *An Introduction to Language Processing with Perl and Prolog.* Sweden: Springer.

Nurgoho, E. (2011). Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karpi, Skripsi, Program Studi Ilmu Komputer Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya Malang.

Putri, A. P., Ridok, A., & Indriati. (2013). Implementasi Metode Improved K-Nearest Neighbor pada Analisis Sentimen Twitter Berbahasa Indonesia (1st ed., pp. 4-5).

Renata, A., Maharani, W., & Kurniati, P. A. (2012). Analisis Klasifikasi Opini Pada Jejaring Sosial Twitter Menggunakan Algoritma K-Nearest Neighbor (KNN).

Ridok, A. (2016). Sentiment Analysis For Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN). *Journal Of Environmental Engineering & Sustainable Technology, 03(01)*, 23-32.

Rijsbergen, C. V. (1979). *Information Retrieval, 2nd ed.* London: Butterworth-Heinemann.

Savani, H. (2016, September 18). *Laravel 5.3 - import export csv and excel file into database*. Retrieved from www.itsolutionstuff.com: http://itsolutionstuff.com/post/laravel-53-import-export-csv-and-excel-file-into-databaseexample.html

Similar Data Finder for Excel. (n.d.). Retrieved from www.mapilab.com: https://www.mapilab.com/excel/similar_data_finder/

Suguna, N., Thanushkodi, K. . (2010). An Improved k-Nearest Neighbor Classification Using Genetic Algorithm. *JCSI International Journal Of Computer Science Issues, 7(4)*.

Tf-idf :: A Single-Page Tutorial - Information Retrieval and Text Mining. (2016, October 11). Retrieved from www.tfidf.com: http://www.tfidf.com/

Triawati, C. (2013, Maret 14). *Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia.* Retrieved from www.digilib.ittelkom.ac.id: http://digilib.ittelkom.ac.id/index.php?option=com_content&view=article&id=590:text-mining&catid=20:informatika&Itemid=14

Zanasi, A., Ebecken, N.F.F., and Almorza Gomar, D. (eds.). (2009). *Data Mining IX. Data Mining, Protection, Detection and other Security Technologies.* Southampton: UK:WIT Press.