

PROSIDING SEMINAR NASIONAL

Teknologi Informasi Kesehatan (SNATIK)
Hasil Penelitian 2017

TEKNOLOGI INFORMASI MENUJU

Smart Health Care

Prosiding Seminar Nasional Teknologi Informasi Kesehatan (SNATIK)

Hasi Penelitian Bidang Kesehatan dan Teknologi Informasi
Teknologi Informasi Menuju Smart Health Care

Copyright ©

Lembaga Penelitian dan Pengabdian Masyarakat STMIK El Rahma Yogyakarta
Lembaga Penelitian dan Pengabdian Masyarakat STIKES Surya Global Yogyakarta

Cetakan Pertama, Desember 2017

21 x 29.7; viii+505 hlm

ISBN : 978-602-278-045-8

Penyunting

Hariza Adnani, S.KM., M.Pd., M.Kes
Suparyanto, S.T., M.Cs

Pengantar

Eko Riswanto, S.T., M.Cs

Kerabat Kerja

Nor Wijayanti, M.Kes
Rustiana Setyowati, SE., M.P.H., MM
Drs. H. Nurhidayat Pamungkas, M.Pd

Pra Cetak : Gatot Satriyo, S.Kom

Lay Out : Jamhari, S.Kom

Cover : M. Amir Mutharom, S.Kom

Diterbitkan Oleh:

Kurnia Kalam Semesta

Bekerjasama dengan

Lembaga Penelitian dan Pengabdian Masyarakat STMIK El Rahma Yogyakarta

Alamat:

Jl. Sisingamangaraja No. 76 Yogyakarta Telp/Fax. (0274) 377982

Lembaga Penelitian dan Pengabdian Masyarakat STIKES Surya Global Yogyakarta

Alamat

Jl Ringroad Selatan Km. 6,7 Blado Potorono Banguntapan Bantul Yogyakarta Telp. (0274) 4469098 /
4469099 , Fax. (0274) 4469101

All right reserved. Semua hak cipta © dilindungi undang-undang. Tidak diperkenankan memproduksi ulang, atau mengubah dalam bentuk apapun melalui cara elektronik, fotocopy, atau rekaman sebagian atau seluruh buku ini tanpa ijin tertulis dari pemilih hak cipta.

SUSUNAN PANITIA

Ketua	:	Eko Riswanto, ST., M.Cs
Wakil Ketua	:	Drs. H. Nurhidayat Pamungkas, M.Pd
Sekretaris	:	Suparyanto, S.T., M.Eng
Wakil Sekretaris	:	Nor Wijayanti, M.Kes
Bendahara	:	Rustiana Setyowati, SE., M.P.H., MM
Reviewer		
Bidang Teknologi Informasi	:	Dr. Awang Hendrianto Pratomo, S.T.,M.T.
Bidang Kesehatan	:	1. Dr. Heru Subaris Kasjono, S.KM, M.Kes 2. Dr. Tri Pitara Mahanggoro, S.Si, M.Kes
Seksi-Seksi		
1. Seminar	:	1. Hariza Adnani, S.KM., M.Pd., M.Kes 2. Sri Sularsih Indartiwi, S.KM., M.P.H 3. Andri Syafrianto, S.Kom, M.Cs 4. Amir Mutharom, S.Kom
2. Paparan Call Paper	:	1. Woro Ispandiah, S.KM., M.P.H 2. Tutik wahyuningsih, M.P.H 3. Yuli Nurlaili 4. Ani Mashunatul Mahmudah 5. Herdiesel Santoso, ST., S.Kom., M.Cs 6. Momon Muzakkar, ST., M.Eng
3. Sekretariat	:	1. Dwi Septiana 2. Dwi kurniawati 3. Fitri Dian Kurniati, S.Kep 4. Jamhari, S.Kom
4. Konsumsi, dan Pubdekdok	:	1. Rr. Viantika Kusumasari, S.Kep., M.Kep., Ners 2. Anna Nur, S.Kep 3. Dedy Fathurrahim 4. Sholeh 5. Suib, S.Kep., M.Kep 6. Muskhah Eko Riyadi, S.Kep., M.Kep., Ners

DAFTAR ISI



No	JUDUL	Hal
BIDANG KESEHATAN		
1	Alternatif Penyempurnaan Pelaksanaan Pemantauan Status Gizi (PSG) Balita (Noviati Fuada dan Irlina Raswanti)	1-14
2	Dukungan keluarga terhadap pemberian ASI pada Bayi di Posyandu Desa Tileng Kecamatan Girisubo Gunungkidul DIY (Dwi Hariyati dan Charlina Yekti Yuli Ningsih)	15-27
3	Evaluasi Penggunaan Sistem Informasi Manajemen Obat Pada Instalasi Farmasi Rumah Sakit Grhasia Yogyakarta (Rendita Dwibarto)	28-32
4	Analisis Pelaksanaan Ketepatan Identifikasi Pasien Sesuai Standar Akreditasi Rumah Sakit Versi 2012 Di RSKIA PKU Muhammadiyah Kota Gede Yogyakarta (Totok Sundoro)	33-52
5	Hubungan Status Gizi, Status Sosial Ekonomi Dan Konsumsi Makanan Cepat Saji Dengan Usia <i>Menarche</i> Pada Siswi kelas V-VI di SD Muhammadiyah Pakel Program PLUS Yogyakarta Tahun 2017 (Sri Sularsih Endartiwi dan Indri Napolisti)	53-71
6	<i>Prevalence Rate</i> dan Faktor Resiko Hipertensi Primer di wilayah Puskesmas Gondokusuman 1 Yogyakarta Tahun 2016 (Nurul Fatimah dan Hariza Adnani)	72-86
7	Pengaruh Terapi Wedang Uwuh Terhadap Perubahan Kolesterol Pada Lansia Hiperkolesterolemia Di Wilayah Pleret Bantul Yogyakarta (Bety Agustina Rahayu, Sarni Anggoro, Rahmania Hidayati)	87-97
8	Analisis Persepsi Pekerja Tentang Keselamatan Dan Kesehatan Kerja (K3) Dengan Perilaku <i>Safety</i> Di PT. Arthur Alfathuh Yuda Brebes 2016 (Akhmad Faizal Himawan dan Sri Yuni Tursilowati)	98-105
9	Pengaruh Lingkungan Terhadap Kejadian Diare Di Puskesmas Banguntapan di Bantul Daerah Istimewa Yogyakarta (Amyati)	106-115
10	Faktor Yang Mempengaruhi Tingginya Angka Kuman Di Ruang Operasi Rumah Sakit Umum PKU Muhammadiyah Bantul (Ubaidillah dan Trea Aprilia Patiah)	116-134
11	Pemahaman Orangtua Tentang Praktik Penggunaan Jamu Dengan Imunitas Dan Tumbuh Kembang Balita Di Padukuhan Pungkuran Wetan Pleret Bantul Yogyakarta (Sarni Anggoro, Tanti Idayati, Bety Agustina Rahayu)	135-142
12	Pengujian Kualitas Sistem Informasi DeLone dan Mc Lean pada Sistem Informasi Manajemen Pendaftaran Rawat Jalan Di Rumah Sakit Umum PKU Muhammadiyah Bantul (Tutik Wahyuningsih)	143-152

13	Gambaran Kejadian Anemia Pada Santri Pondok Pesantren Fadlun Minalloh Yogyakarta (Mudita Sri Hidayah)	153-157
14	Teknologi Informasi Berbasis Komputerisasi Dan Sumber Daya Manusia Terhadap Efektivitas Penerapan Sistem Informasi Manajemen (Anggi dan Ivan Tinarbudi Gavinov)	158-172
15	Hubungan Spiritualitas dengan Tingkat Penyalahgunaan Napza pada Awak Bus di Terminal Giwangan Yogyakarta (Sri Setyowati dan Dian Mar'atu Zakyyah)	173-186
16	Pelatihan Bermain Pada Pengasuh Dapat Meningkatkan Sensitivitas Pengasuhan Anak Prasekolah (Eka Oktavianto dan Ketut Anggun Paramita)	187-201
17	Gambaran Karakteristik Kejadian Kehamilan Tidak Diinginkan Pada Remaja Di Wilayah Kerja Puskesmas Banguntapan Bantul Yogyakarta (Ismarwati, Dwi Ernawati dan Herlin Fitriani Kurniawati)	202-206
18	Hubungan Pengetahuan, Sikap, Dan Perilaku Ibu Dalam Pemenuhan Nutrisi Dengan Status Gizi Balita Di Posyandu Harapan Ibu Desa Sumberharjo Kecamatan Pacitan Kabupaten Pacitan Provinsi Jawa Timur Tahun 2016 (Titik Mariati)	207-214
19	Pengaruh Pemberian Latihan <i>Range Of Motion</i> (ROM) Terhadap Kekuatan Otot Kaki Pada Pasien Diabetes Mellitus Di RSUP Dr. Soeradji Tirtonegoro Klaten (Jawa Tengah) (Fatma Hapida dan Riza Yulina Amri)	215-238
20	Pengaruh Iklan Susu Formula Dan Pekerjaan Ibu Terhadap Motivasi Memberikan Asi Eksklusif Di Wilayah Kerja Puskesmas Temon II (Nafisaturochmah dan Norra Hendarni)	239-254
21	Hubungan Pengetahuan Dengan Perilaku Masyarakat Mengonsumsi Obat Paten Dibanding Obat Generik di Desa Babakan Ciwaringin Cirebon Tahun 2017 (Woro Ispandiyah)	255-262
22	Hubungan Pengetahuan dan Sikap Dalam Pencegahan Demam Berdarah Dengue (DBD) Di Wilayah Kerja Puskesmas Kasihan II Bantul DIY (Nor Wijayanti)	263-269
23	Faktor-Faktor yang Mempengaruhi Lamanya Waktu Penyediaan Dokumen Rekam Medis Pasien Lama di Instalasi Rawat Jalan (Teguh Kurniadi dan Tri Widayanti)	270-277
24	Hubungan Tingkat Pengetahuan dan Perilaku Ibu Tentang Stimulasi Perkembangan Dengan Perkembangan Motorik Kasar Pada Anak Usia 3-5 Tahun Di TK Baik Krpyak Kulon Yogyakarta (Sinta Sundari)	278-286

BIDANG TEKNOLOGI INFORMASI

- | | | |
|----|---|---------|
| 25 | Aplikasi Sistem Pendukung Keputusan Seleksi Penerimaan Siswa Baru Pada Smp Negeri 8 Samarinda Menggunakan Metode Simple Additive Weighting (Anton Topadang, Abdul Najib) | 287-297 |
| 26 | Klasifikasi Komentar Spam Pada Instagram Berbahasa Indonesia Menggunakan K-NN (Antonius Rachmat Chrismanto, Yuan Lukito) | 298-306 |
| 27 | Implementasi Decision Support System Untuk Analisis Kelayakan Pendirian Minimarket Dengan Metode Fuzzy Analytical Hierarchy Process (Wahyu Widodo) | 307-321 |
| 28 | Sistem Pakar Untuk Menentukan Jenis Terapi Nyeri Punggung Bawah Myogenic (Alfian Gema Negara, Izzati Muhimmah) | 322-333 |
| 29 | Sistem Pakar Untuk Mendeteksi Gangguan <i>Pervasive Developmental Disorder</i> Dengan Implementasi Inferensi Forward Chaining Menggunakan Metode <i>Certainty Factor</i> (Avrillaila Akbar Harahap, Arfi Joyendri²) | 334-343 |
| 30 | Klasifikasi Sentimen Terhadap Layanan Penerbangan Pada Microblogging Menggunakan Metode Support Vector Machine (Tien Rahayu Tulili, Muhammad Farman Andrijasa) | 344-352 |
| 31 | Intranet Berbasis Opensource Sebagai Solusi Pusat Data (Data Center) Untuk Sekolah Di Pedesaan (Hari Aspriyono, Yode Arliando) | 353-362 |
| 32 | Perancangan Sistem Informasi Pada Bengkel Fauzan Motor Menggunakan Metode Waterfall (Moehamad Irfan, S. Thya Safitri, Didi Supriyadi) | 363-372 |
| 33 | Sistem Informasi Pelaporan Dan Monitoring Kinerja Dosen Dengan Metode Rapid Application Development (RAD) (Jayanti Tri Herawati, Didi Supriyadi, Dwi Januarita Ardianing Kusuma) | 373-383 |
| 34 | Rekomendasi Rute Perjalanan Multi Destinasi Dengan Algoritma Genetika (Edi Iskandar, Sudarmanto, Herdiesel Santoso) | 384-394 |
| 35 | Sistem Informasi Alumni Stikes Surya Global Yogyakarta (Herdiesel Santoso, Ahmad Arif Nurrahman, Edi Iskandar) | 395-407 |
| 36 | Efektivitas Pengawasan Pemilihan Umum Berbasis Sistem Informasi (Roni Marzuki, Fadlul Amdi Yul) | 408-417 |
| 37 | Implementasi Metode Forward Chaining Untuk Mendeteksi Kriteria User Dalam Menggunakan Smartphone (Putri Aylin Marsha, Indra Hidayatulloh, Muhammad Zidny Naf'an) | 418-428 |
| 38 | Analisis Dan Desain Perancangan Database Pada Sistem Informasi Camp Bebas Riba (Amalia Indriani Dwian Pramudita, Didi Supriyadi) | 429-436 |
| 39 | Algoritma Kriptografi Playfair Chiper Dengan Java (Yuli Praptomo PHS., S.Kom., M.Cs) | 437-446 |
| 40 | Efek Aperture, Dept Of Field Citra Dengan Segmentasi Citra | 447-452 |

(Sugiyatno)

- | | | |
|----|--|---------|
| 41 | Sistem Rekomendasi Rute Tempat Wisata Menggunakan Topsis Dan Algoritma Genetika (Gunawan Abdillah, Esmeralda C. Djamal) | 453-460 |
| 42 | Aplikasi Sistem Pakar Sekolah Siaga Bencana Menggunakan Metode <i>Certainty Factor</i> (CF) (Joko Siswanto) | 461-473 |
| 43 | Penerapan Teknologi Augmented Reality Pada Brosur Promosi Stmik El Rahma Yogyakarta (Ahmad Aspuri, Asih Winantu) | 474-488 |
| 44 | Simulator Pengenal String Yang Diterima Sebuah Deterministic Finite Automata (DFA) Menggunakan Linked List (Suparyanto) | 489-496 |
| 45 | Survei Kerangka Kerja <i>Business Process Reengineering</i> (Brata Widiyantoro) | 497-505 |

KLASIFIKASI KOMENTAR SPAM PADA INSTAGRAM BERBAHASA INDONESIA MENGGUNAKAN K-NN

Antonius Rachmat Chrismanto¹, Yuan Lukito²

¹Fakultas Teknologi Informasi/Prodi Informatika, Universitas Kristen Duta Wacana
Yogyakarta,

²Fakultas Teknologi Informasi/Prodi Informatika, Universitas Kristen Duta Wacana
Yogyakarta

e-mail: ¹anton@ti.ukdw.ac.id, ²yuanlukito@ti.ukdw.ac.id

Abstrak

Para publik figur, terutama aktor dan artis menggunakan Instagram (IG) sebagai salah satu media sosial berbasis foto untuk mempromosikan kegiatan dan menjalin relasi dengan para penggemarnya. Para penggemar dapat memfollow, melihat foto, dan berkomentar pada setiap status IG artis idolanya. Permasalahannya, banyak sekali komentar spam yang ditulis pada IG yang sampai saat ini belum ada penyelesaiannya, terutama untuk spam berbahasa Indonesia. Penelitian ini melakukan pengumpulan dataset komentar IG dari 10 artis Indonesia berfollower terbanyak sejumlah 16461 data, melakukan pelabelan pada dataset training, pre-processing, pembobotan TF-IDF, dan klasifikasi komentar spam berbahasa Indonesia menggunakan k-NN. Hasil dari penelitian ini menggunakan k terbaik 3 adalah: (1) untuk dataset tanpa stemming, rata-rata akurasi mencapai 87.84% pada unbalanced dataset dan 86.17% untuk balanced dataset, (2) untuk dataset dengan stemming, rata-rata akurasi mencapai 88.44 % pada unbalanced dataset dan 85.86% untuk balanced dataset, (3) diperlukan penanganan lebih serius mengenai karakter unicode dan simbol.

Kata Kunci: Instagram, Klasifikasi Komentar Spam, k-NN

PENDAHULUAN

Media sosial banyak digunakan oleh pengguna Internet untuk tetap eksis sekaligus bersosialisasi di dunia maya. Para publik figur, seperti politikus dan artis / aktor, Indonesia banyak menggunakan media sosial seperti Facebook, Twitter, Instagram (IG) dan Path. Jika Facebook dan Twitter lebih banyak menggunakan teks sebagai statusnya, IG dan Path menggunakan foto dan *caption* foto sebagai statusnya. IG dengan jumlah pengguna mencapai 500 juta, per hari terdapat 95 juta foto yang diunggah [1]. Hal ini menjadikannya media sosial berbasis foto yang sangat terkenal dan lebih disukai para penggunanya, sebab dengan berbasis foto/gambar lebih bersifat visual dan nyata dibandingkan dengan sekedar tulisan. Permasalahannya, IG ternyata memiliki banyak sekali komentar spam pada status-status para publik figur, terutama artis/aktor yang memiliki *follower* sangat banyak. Komentar spam tersebut bisa berupa tulisan yang tidak berhubungan sama sekali dengan statusnya, komentar jualan, endorse, bahkan link ke suatu website berbahaya tertentu, dan lain lain. Semakin terkenal suatu artis/aktor, semakin banyak follower-nya, maka semakin banyak pula mendapat komentar spam. Komentar spam mengganggu karena: (1) dapat menyebabkan misleading informasi jumlah komentar. Jumlah komentar yang banyak merupakan salah satu indikator popularitas seorang publik figur, namun ternyata jumlah komentar yang banyak adalah komentar spam, (2) mengganggu alur diskusi yang terdapat pada suatu status sehingga pengguna akan kesulitan dalam mencari informasi dengan cepat dan tepat.

Sampai saat ini belum ada sistem penanggulangan komentar spam di IG yang handal. Untuk IG berbahasa Inggris, yang dapat dilakukan oleh pengguna IG adalah dengan

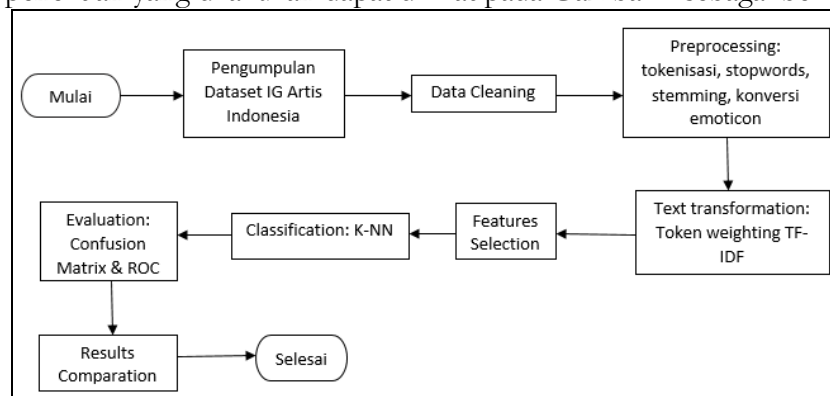
melaporkan bahwa suatu komentar adalah spam dan kemudian dapat dihapus secara manual jika dikehendaki [2], atau melakukan *filtering* otomatis berbasis *keywords* tertentu untuk spam berbahasa Inggris yang juga harus diaktifkan sendiri oleh pengguna tersebut [3]. Untuk bahasa lain selain Inggris, belum ada mekanisme lain yang bagus dan otomatis. Pengguna IG bisa saja membuat akunnya menjadi privat dan menonaktifkan komentar pada setiap status, namun hal itu tentu saja tidak akan dilakukan oleh para aktor / artis karena akun mereka tetap harus bersifat publik agar para penggemar dapat mem-*follow*-nya sebagai sarana promosi kegiatan artis/aktor tersebut.

Arti penting penelitian ini adalah berusaha mengambil peran dalam menangani komentar spam pada IG berbahasa Indonesia, dengan memulainya dari penelitian sebelumnya, yaitu deteksi komentar spam menggunakan metode Naive Bayes [4]. Deteksi spam memang sudah banyak dilakukan, misalnya untuk deteksi email spam menggunakan ELM & SVM [5], spam pada *closed group* Facebook menggunakan Random Forest dengan tingkat akurasi mencapai 98% [6] dan menggunakan Maximum Entropy [7], deteksi spam pada media sosial Twitter menggunakan konten hibrid dan algoritma graf [8], deteksi spam review pada produk-produk yang terdapat pada *marketplace*, seperti Amazon.com [9], dan lain sebagainya. Perbedaan yang penting dari komentar spam pada IG adalah bahwa komentar spam pada IG memiliki ciri: (1) mempunyai panjang karakter yang beraneka ragam, pendek atau panjang (lebih dari 160 karakter, tidak seperti *microblogging* Twitter), (2) lebih banyak mengandung karakter Unicode dan simbol-simbol (*emoticon* dan gambar-gambar lainnya seperti bunga, jempol, bibir, dan lain-lain yang mengandung arti suka/tidak suka), dan (3) lebih terlihat jelas komentar spam yang spesifik terhadap profil pengguna tertentu saja.

Penelitian ini dilakukan dengan menggunakan dataset dari 10 artis Indonesia yang memiliki *follower* di atas 10 juta, berdasarkan referensi pada [10], di mana dari masing-masing artis hanya diambil 50 status, dan dari masing-masing status hanya diambil 50 komentar. Penelitian ini menggunakan k-NN sebagai algoritma untuk melakukan klasifikasi komentar spam pada IG. K-NN dipilih karena: (1) k-NN termasuk algoritma yang mudah diimplementasikan, cepat dari sisi waktu pemrosesan, dan akurasi tinggi [11] (2) k-NN memiliki banyak dukungan *library* pemrograman, (3) k-NN akan dibandingkan akurasinya dengan algoritma Naive Bayes yang sudah diuji pada penelitian sebelumnya [4]. Pada tahap penelitian berikutnya, algoritma yang terbaik akan diimplementasikan dalam bentuk sistem berbasis layanan berbasis awan (*cloud services system*) yang kemudian dikembangkan pada implementasi nyata dengan membangun *browser's extension* untuk mengidentifikasi dan menandai komentar spam secara otomatis..

METODE PENELITIAN

Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1 sebagai berikut:



Gambar 1. Tahap-tahap Penelitian

A. Tahap *Data Collection*

Tahap pengumpulan data (*data collection*) merupakan tahap untuk pengambilan data yang akan digunakan sebagai data latih dan data uji. Dataset yang dihasilkan dari tahap ini berupa dataset status artis / aktor Indonesia yang memiliki *follower* lebih dari 10 juta beserta komentar-komentar pada status-statusnya. Pada tahap ini data diambil dengan menggunakan InstagramGrabber yang berbasis PHP dan dimodifikasi dari aplikasi buatan Vassard dan Borlander [12]. Daftar 10 akun artis/aktor IG Indonesia berfollower lebih dari 10 juta yang digunakan sesuai urutan jumlah follower dari besar ke kecil adalah @ayutingting92, @princessyahrini, @raffinagita1717, @laudyacynthiabella, @prillylatuconsina96, @juliaperrezz: @chelseaoliviana, @raisa6690, @lunamaya, @agnezmo [10].

B. Tahap *Data Cleaning*

Dari dataset sejumlah 17.007 data, yang terdiri dari data spam sejumlah 10.719, dan data non spam sejumlah 6.288, *data cleaning* dilakukan dengan cara menghapus karakter-karakter khusus yang tidak bermakna, menghapus angka, menghapus link / URL, dan data-data kosong (null) dan yang tidak lengkap. Dari hasil pada tahap *data cleaning*, terjadi penyusutan dataset akhir menjadi 16.641.

C. Tahap *Pre-Processing*

Tahap *pre-processing* yang dilakukan adalah: (1) tokenisasi, yaitu mengubah kalimat-kalimat menjadi token-token yang bermakna, (2) penghapusan token-token yang masuk dalam *stopwords*, yaitu data token yang tidak memiliki makna dan dapat dibuang menggunakan kamus khusus stopwords berjumlah 1050 token, (3) *stemming*, yaitu tahap untuk mengubah token-token yang ada menjadi token (kata) dasar dengan tujuan untuk mengurangi jumlah token yang akan digunakan untuk proses berikutnya, dan terakhir (4) *symbol dan emoticon handler*, yaitu tahap untuk menangani konversi *emoticon* dan simbol-simbol khas Instagram yang diperoleh dari tahap sebelumnya.

D. Tahap *Text Transformation*

Tahap ini adalah tahap untuk melakukan perubahan dari data token teks menjadi data vector yang memiliki bobot tertentu sehingga dapat dilakukan perhitungan untuk klasifikasi. Pembobotan vector token menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) sehingga akhirnya semua token akan memiliki bobot tertentu yang kemudian dilakukan normalisasi. Pemilihan TF-IDF digunakan pada penelitian ini karena TF-IDF termasuk baik dalam merepresentasikan bobot token untuk pembobotan [13].

E. Tahap *Features Selection*

Dari semua vektor token yang dibentuk pada tahap sebelumnya, tentunya masih berjumlah sangat banyak sehingga diperlukan suatu tahapan lagi untuk memilih token-token tertentu saja yang bisa dianggap sebagai fitur-fitur representatif yang mewakili komentar yang akan diklasifikasi / dideteksi berikutnya. Tujuan dari tahap ini adalah mengurangi jumlah token yang akan digunakan untuk tahap klasifikasi namun dapat mewakili tiap-tiap data komentar [14].

F. Tahap Klasifikasi

Tahap klasifikasi adalah proses inti yang akan dilakukan dengan menggunakan metode k-NN. Metode k-NN dipilih nilai k tertentu dari 1 s/d 10 dan kemudian dicari nilai k terbaik. Parameter algoritma klasifikasi k-NN dapat dilihat pada Tabel 1.

Tabel 1. Parameter Klasifikasi pada k-NN

Parameter	Nilai
Jumlah k	1 - 5
<i>Weighted vote</i>	<i>Yes</i>
<i>Measure Type</i>	<i>Mixed measure</i>
<i>Mix measure</i>	<i>Mixed euclidian distance</i>

G. Tahap Evaluasi

Tahap evaluasi adalah tahap akhir untuk menentukan baik buruknya metode klasifikasi yang digunakan pada dataset. Tahap evaluasi diukur menggunakan confusion matrix [15]. Dari confusion matrix yang digunakan, diambil nilai akurasi, *precision*, dan *recall*-nya. Metode pengujian menggunakan k-fold validation berjumlah 10 kali dan kemudian diambil rata-rata akurasinya. Parameter lengkap metode evaluasi dapat dilihat pada Tabel 2.

Tabel 2. Parameter Evaluasi k-NN

Parameter	Nilai
Metode validasi	<i>k-Fold Validation</i>
Matrik pengujian	<i>Confusion Matrix</i>
Jumlah data	16641 data
Tool	RapidMiner 7.6
Kriteria output yang dihasilkan	<i>Accuracy</i>
Sampling tape	<i>Shuffled sampling</i>

HASIL DAN PEMBAHASAN

Pengujian pertama dilakukan dengan melakukan pencarian nilai k terbaik dengan konfigurasi dataset *unbalanced* dan *non stemming* untuk digunakan sebagai k *default* pada sistem. Tabel percobaan nilai k dapat dilihat pada Tabel 3 berikut.

Tabel 3. Percobaan nilai k terbaik

Jumlah K	Akurasi
k=1	87.4 %
k=2	86.84 %
k=3	87.84 %
k=4	87.4 %
k=5	86.32 %

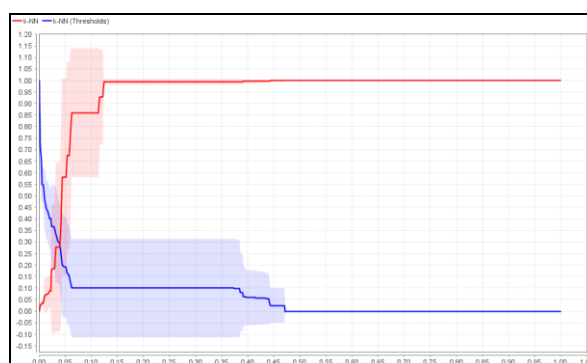
Dari penelitian yang dilakukan penulis menggunakan skenario berikut sebagai sarana untuk melakukan pengujian dan analisis:

1. Skenario I: *Unbalanced Dataset, Non-Stemming*

Skenario I menggunakan *unbalanced* dataset yang berisi *dataset* apa adanya, yaitu berjumlah 10.399 untuk data spam dan 6062 untuk data *not spam* tanpa dilakukan *stemming*. Hasil klasifikasi menggunakan k-NN dengan k=3 dapat dilihat pada Tabel 4 dan kurva ROC dapat dilihat pada Gambar 2. Dari hasil tersebut diperoleh akurasi rata-rata 87.84 %, dan algoritma k-NN sangat baik dalam melakukan klasifikasi.

Tabel 4. *Confusion Matrix* Skenario I

	True Spam	True Not Spam	Class Precision
Predicted Spam	4879	291	94.37 %
Predicted Not Spam	1183	5771	82.99 %
Class Recall	80.48 %	95.2 %	



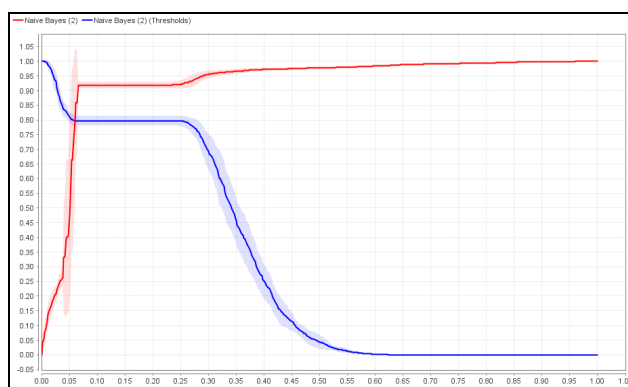
Gambar 2. ROC Skenario I

2. Skenario II : *Unbalanced Dataset, Stemming*

Skenario II menggunakan *unbalanced* dataset yang berisi dataset apa adanya, yaitu berjumlah 10.399 untuk data spam dan 6062 untuk data *not spam* dengan terlebih dahulu dilakukan *stemming*. Hasil klasifikasi menggunakan k-NN dengan k=3 dapat dilihat pada Tabel 5 dan kurva ROC dapat dilihat pada Gambar 3. Dari hasil tersebut diperoleh rata-rata akurasi 88.4 % yang lebih dari dari skenario I dan kurva ROC nya lebih halus serta memiliki luasan yang lebih luas.

Tabel 5. *Confusion Matrix* Skenario II

	True Spam	True Not Spam	Class Precision
Predicted Spam	8972	476	94.96 %
Predicted Not Spam	1427	5586	79.65 %
Class Recall	86.28 %	92.15 %	



Gambar 4. ROC Skenario II

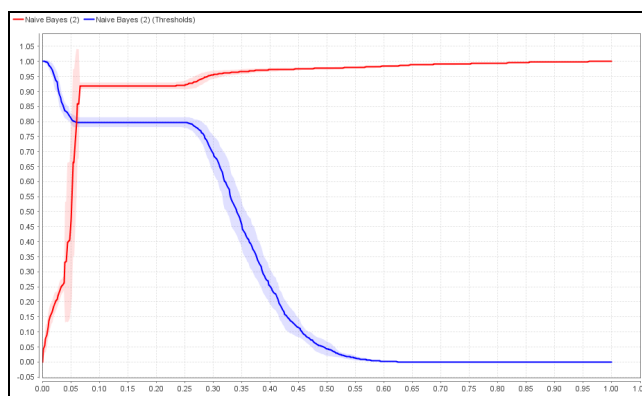
3. Skenario III : *Balanced Dataset, Non-Stemming*

Skenario III menggunakan *balanced* dataset yang berisi dataset yang diseimbangkan, yaitu berjumlah 6062 untuk data spam dan 6062 untuk data *not spam* tanpa dilakukan *stemming*. Hasil klasifikasi menggunakan k-NN dengan k=3 dapat dilihat pada Tabel 6 dan kurva ROC dapat dilihat pada Gambar 5. Dari hasil penelitian diperoleh rata-rata akurasi 86.17 % dan

mengalami sedikit penurunan dibandingkan skenario I dan II, namun kurva ROC mendekati skenario II.

Tabel 6. *Confusion Matrix* Skenario III

	True Spam	True Not Spam	Class Precision
Predicted Spam	4717	332	93.42 %
Predicted Not Spam	1345	5730	80.99 %
Class Recall	77.81 %	94.52 %	



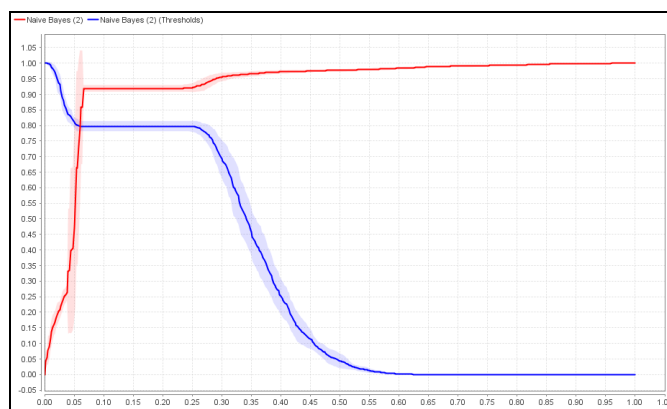
Gambar 5. ROC Skenario III

4. Skenario IV : *Balanced Dataset, Stemming*

Skenario IV menggunakan *balanced* dataset yang berisi dataset yang diseimbangkan, yaitu berjumlah 6062 untuk data spam dan 6062 untuk data *not spam* tanpa dilakukan *stemming*. Hasil klasifikasi menggunakan k-NN dengan k=3 dapat dilihat pada Tabel 7 dan kurva ROC dapat dilihat pada Gambar 6. Hasil dari penelitian menunjukkan rata-rata akurasi adalah 85,86 %, dan paling rendah dari skenario-skenario lainnya.

Tabel 7. *Confusion Matrix* Skenario IV

	True Spam	True Not Spam	Class Precision
Predicted Spam	4698	350	93.07 %
Predicted Not Spam	1364	5712	80.72 %
Class Recall	77.5 %	94.23 %	

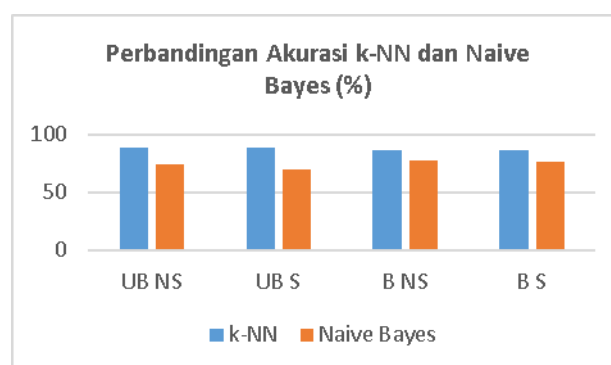


Gambar 6. ROC Curve Skenario IV

Dari hasil tersebut kemudian dilakukan perbandingan akurasi dari penelitian sebelumnya [4] yang menggunakan metode Naive Bayes dan yang menggunakan k-NN berdasarkan berbagai skenario seperti yang dilihat pada Tabel 8. Jika dilihat dalam bentuk grafik dapat dilihat pada Gambar 7. Keterangan gambar 7 adalah: UB NS=*Unbalanced Non Stem*, UB S=*Unbalanced Stem*, B NS=*Balanced Non Stem*, dan B S=*Balanced Stem*.

Tabel 8. Perbandingan Akurasi Akhir

Metode	Parameter Dataset	Akurasi
k-NN	<i>Unbalanced, Non Stem (UB NS)</i>	87.84 %
	<i>Unbalanced, Stem (UB S)</i>	88.4 %
	<i>Balanced, Non Stem (B NS)</i>	86.17 %
	<i>Balanced, Stem (B S)</i>	85.86 %
	RATA-RATA	87.07 %
Naïve Bayes	<i>Unbalanced, Non Stem (UB NS)</i>	74.31 %
	<i>Unbalanced Stem (UB S)</i>	69.96 %
	<i>Balanced Non Stem (B NS)</i>	77.25 %
	<i>Balanced Stem (B S)</i>	76.1 %
	RATA-RATA	74.40 %



Gambar 7. Grafik Perbandingan Akurasi k-NN dan Naive Bayes

KESIMPULAN

Kesimpulan yang diperoleh dari hasil penelitian yang telah dilakukan adalah:

1. Algoritma k-NN dapat diimplementasikan dengan sangat baik pada kasus klasifikasi komentar spam pada Instagram berbahasa Indonesia dengan tingkat rata-rata akurasi dari 4 skenario 87.07 %. Hal ini masuk dalam kategori hasil yang sangat baik.

2. Penelitian ini telah menguji sistem deteksi komentar spam pada IG menggunakan algoritma klasifikasi k-NN dan diperoleh rata-rata akurasi sebesar 87.84 % untuk skenario I (UB NS) dan sebesar 88,4% untuk skenario II (UB S). Terjadi peningkatan sebesar 0.56 % akibat dari penggunaan *stemming*.
3. Penelitian ini telah menguji sistem deteksi komentar spam pada IG menggunakan algoritma klasifikasi k-NN dan diperoleh rata-rata akurasi sebesar 86.17 % untuk skenario III (B NS) dan sebesar 85.86 % untuk skenario IV (B S). Terjadi penurunan 0.31 % pada dataset seimbang (B).
4. Jika dibandingkan dengan penelitian sebelumnya, k-NN memiliki tingkat akurasi yang lebih tinggi yaitu selisih 12.67 %.
5. Tahapan *pre-processing* pada Instagram bahasa Indonesia yang sangat perlu diperhatikan adalah: penanganan khusus karakter Unicode (UTF-8) dan konversi simbol serta *emoticon*.

SARAN

Saran-saran yang sebaiknya dilakukan pada tahap penelitian berikutnya adalah:

1. Proses *pre-processing* yang berkaitan dengan karakter *Unicode* perlu dilakukan dengan lebih baik lagi agar sistem lebih siap diproses lebih lanjut dalam implementasi nyata dan digunakan secara publik.
2. Penelitian perlu dilanjutkan ke tahap selanjutnya yaitu pengembangan sistem berbasis *cloud service* agar dapat digunakan dalam ekstensi browser.

DAFTAR PUSTAKA

- [1] R. Grant, "Why ALL ACTORS Need To Get On Instagram NOW!," Act on This - The TV Actor Network, March 2017. [Online]. Available: <https://www.actonthis.tv/2017/03/actors-need-get-instagram-now/>. [Diakses 9 August 2017].
- [2] Instagram, "Bagaimana cara melaporkan komentar yang melanggar atau spam?," Instagram, 2017. [Online]. Available: <https://help.instagram.com/198034803689028>. [Diakses 24 October 2017].
- [3] J. Vanian, "Instagram Turns to Artificial Intelligence to Fight Spam and Cruel Comments," Fortune, 29 June 2017. [Online]. Available: <http://fortune.com/2017/06/29/instagram-artificial-intelligence-offensive-comments/>. [Diakses 24 October 2017].
- [4] A. Rachmat dan Y. Lukito, "Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes," *Ultimatics - Jurnal Informatika*, vol. 9, no. 1, pp. 50-58, 1 June 2017.
- [5] S. O. Olatunji, "Extreme Learning machines and Support Vector Machines models for email spam detection," dalam *IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, Windsor, ON, Canada, 2017.
- [6] N. Watcharenwong dan K. Saikaew, "Spam detection for closed Facebook groups," dalam *14th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Nakhon Si Thammarat, Thailand, 2017.
- [7] T.-H. Vuong, V.-H. Tran, M.-D. Nguyen, C.-V. T. Nguyen, T.-H. Pham dan M.-V. Tran, "Social-spam profile detection based on content classification and user behavior," dalam *Eighth International Conference on Knowledge and Systems Engineering (KSE)*, Hanoi, 2016.
- [8] M. Mateen, M. A. Iqbal, M. Aleem dan M. A. Islam, "A hybrid approach for spam detection for Twitter," dalam *14th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, Islamabad, 2017.

- [9] S. K. Chauhan, A. Goel, P. Goel, A. Chauhan dan M. K. Gurve, "Research on product review analysis and spam review detection," dalam *4th International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida, India, 2017.
- [10] M. Deoranje, "10+ Akun Instagram Dengan Followers Terbanyak di Indonesia," Musdeoranje.net, August 2016. [Online]. Available: <http://www.musdeoranje.net/2016/08/akun-instagram-dengan-followers-terbanyak-di-indonesia.html>. [Diakses 9 August 2017].
- [11] I. Witten, E. Frank dan M. Hall, *Data mining: Practical Machine Learning Tools and Techniques*, 3rd penyunt., MA: Morgan Kaufmann, 2011.
- [12] K. Vassard dan T. Borlander, "Bolandish/PHP-Instagram-Grabber," Github, 13 February 2017. [Online]. Available: <https://github.com/Bolandish/PHP-Instagram-Grabber>. [Diakses 15 August 2017].
- [13] W. Zhang, T. Yoshida dan X. Tang, "A comparative study of TF-IDF, LSI, and multi-words for text classification," *Expert Systems with Application*, vol. 38, no. 2011, pp. 2758-2765, 2010.
- [14] G. Forman, "An extensive empirical study of feature selection metrics for text classification," *Journal of machine learning research*, vol. 3, no. March, pp. 1289-1305, 2003.
- [15] K. Markham, "Simple guide to confusion matrix terminology," Data School, 25 March 2014. [Online]. Available: <http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/>. [Diakses 11 August 2017].