

J-INTTECH

Journal of Information and Technology

Volume 06 Nomor 01, Bulan Juni Tahun 2018



STIKI

SEKOLAH TINGGI INFORMATIKA & KOMPUTER INDONESIA

Jl. Raya Tidar 100 Malang, 65146

Telp. (0341)560823, Fax (0341)562525

ISSN: 2303-1425 E-ISSN: 2580-720X

J-INTTECH

Journal of Information and Technology
Volume 06 Nomor 01, Bulan Juni 2018



LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT

STIKI

SEKOLAH TINGGI INFORMATIKA & KOMPUTER INDONESIA
Jl. Raya Tidar 100, Malang; Phone: 0341-560823; Fax: 0341-562525; <http://www.stiki.ac.id>; mail@stiki.ac.id

PENGANTAR REDAKSI

J-INTECH merupakan jurnal yang diterbitkan oleh Sekolah Tinggi Informatika dan Komputer Indonesia Malang guna mengakomodasi kebutuhan akan perkembangan Teknologi Informasi serta guna mensukseskan salah satu program DIKTI yang mewajibkan seluruh Perguruan Tinggi untuk menerbitkan dan mengunggah karya ilmiah mahasiswanya dalam bentuk terbitan maupun jurnal online.

Pada edisi ini, redaksi menampilkan beberapa karya ilmiah mahasiswa yang mewakili beberapa mahasiswa yang lain, yang dianggap cukup baik sebagai media pembelajaran bagi para lulusan selanjutnya.

Tentu saja diharapkan pada setiap penerbitan memiliki nilai lebih dari karya ilmiah yang dihasilkan sebelumnya sehingga merupakan nilai tambah bagi para adik kelas maupun pihak-pihak yang ingin studi atau memanfaatkan karya tersebut selanjutnya.

Pada kesempatan ini kami juga mengundang pihak-pihak dari PTN/PTS lain sebagai kontributor karya ilmiah terhadap jurnal J-INTECH, sehingga Perkembangan IPTEK dapat dikuasai secara bersama-sama dan membawa manfaat bagi institusi masing-masing.

Akhir redaksi berharap semoga dengan terbitnya jurnal ini membawa manfaat bagi para mahasiswa, dosen pembimbing, pihak yang bekerja pada bidang Teknologi Informasi serta untuk perkembangan IPTEK di masa depan.

REDAKSI

J-INTECH

Journal of Information and Technology

Volume 06 Nomor 01, Bulan Juni 2018

DAFTAR ISI

Sistem Informasi Pelayanan Terpadu di Restoran Berbasis Android <i>Hery Kuswandi</i>	01-08
Pemanfaatan <i>Raspberry Pi</i> Dan Webcam Sebagai Kamera Pemantau Dan <i>Cloud Drive</i> Sebagai Media Penyimpanan <i>Ady Noegroho</i>	09-17
Sistem Penunjang Keputusan Berbasis <i>Webgis</i> Dengan Metode AHP Untuk Pemilihan Lokasi Usaha..... <i>Sya'roni</i>	18-22
Aplikasi Manajemen <i>Inventory</i> Berbasis <i>Mobile</i> <i>Angga Eka Syaputra</i>	23-32
Aplikasi Perencanaan Kebutuhan Produksi Menggunakan <i>Demand</i> <i>Forecasting</i> Dengan Pendekatan Proyektif..... <i>Samuel Pusirumang Makahanap</i>	33-42
Membangun Aplikasi <i>E-Commerce</i> Dengan Sistem Penunjang Keputusan Metode Apriori Untuk Memberikan Rekomendasi Kepada Calon Pembeli Di Toko Islam Malang <i>Alamsyah Ady Nugroho</i>	43-47
Sistem Informasi <i>Inventory</i> pada UD. MM GoDAM “NENENG” Berbasis Web Guna Memudahkan Pengolahan Data Barang..... <i>Widia Normalasari</i>	48-52
Aplikasi Pembelajaran Menulis Permulaan Berbasis Android Menggunakan <i>Unity 2D</i> <i>Andi Fiqqih Adiqro</i>	53-62
Sistem Pakar Identifikasi Penyakit Burung Puyuh Menggunakan Metode <i>Inferensi Forward Chaining</i> Berbasis Android <i>Mahartin Hendra Sukmawan</i>	63-77

Sistem Keamanan <i>Database</i> Berbasis <i>Restfull</i> Pada <i>Content Management System Wordpress</i> (Studi Kasus : STIKI Malang).....	78-89
<i>Ridho Valentin</i>	
Sistem Informasi Pengolahan Data Surat Masuk dan Keluar di Kantor BARENLITBANG Kota Malang.....	90-93
<i>Antonius Lorensius</i>	
Sistem Penunjang Keputusan Pemilihan Jurusan Perguruan Tinggi Menggunakan Teori Psikologi <i>Rothwell Miller Interest Blank</i> (RMIB)	94-104
<i>Muhammad Hanifudin</i>	
Permainan Ular Tangga Berbasis Android Menggunakan <i>Unity</i>	105-118
<i>Novanda Bayhakky</i>	
Sistem Informasi Manajemen Pakan Guna Meningkatkan Indikator Keberhasilan Panen Ternak pada PT Berkah Benua Farm	119-140
<i>Burhannudin</i>	
Klasifikasi Artikel Berbahasa Indonesia untuk Mendeteksi <i>Clickbait</i> Menggunakan Metode Naïve Bayes	141-147
<i>Ali Fahnnur Yavi</i>	
Sistem Informasi Akademik SMK Bhakti Luhur Malang Berbasis Web	148-152
<i>Fransiskus Sina Witi</i>	
Pencarian Resep Masakan Menggunakan Metode <i>Vector Space Model</i> (VSM) Berbasis Android	153-160
<i>Bulan Dewi Gulita</i>	
Pemanfaatan Sensor Gyroscope pada Game Casual Berbasis Android.....	161-165
<i>Dionisius Aditya Remy Susanto</i>	
Penerapan Teknologi Augmented Reality pada <i>Game</i> Pengenalan Hewan Berdasarkan Jenis Makanannya Berbasis <i>Mobile</i>	166-172
<i>Herjuno Daud Pramono</i>	
Sistem Penunjang Keputusan Pemilihan Rumah Kontrakan untuk Keluarga di Kota Malang Menggunakan Metode Fuzzy Sugeno.....	173-176
<i>Slamet Nur Huda</i>	

ISSN: 2303-1425 E-ISSN: 2580-720X

J-INTTECH

Journal of Information and Technology
Volume 06 Nomor 01, Bulan Juni 2018

- Pelindung** : Ketua STIKI
- Penasehat** : Puket I, II, III
- Pembina** : Ka. LPPM
-
- Editor** : Subari, S.Kom, M.Kom
- Section Editor** : Daniel Rudiaman S.,ST, M.Kom
-
- Reviewer** : Dr. Eva Handriyantini, S.Kom, M.MT.
Evi Poerbaningtyas, S.Si, M.T.
Laila Isyriyah, S.Kom, M.Kom
Anita, S.Kom, M.T.
-
- Layout Editor** : Siti Aminah, S.Si, M.Pd
Nira Radita, S.Pd., M.Pd
Muh. Bima Indra Kusuma

Klasifikasi Artikel Berbahasa Indonesia untuk Mendeteksi *Clickbait* menggunakan Metode Naïve Bayes

Ali Fahnnur Yavi

Program Studi Teknik Informatika, Sekolah Tinggi Informatika dan Komputer Indonesia (STIKI) Malang
alvin.tkjunity@gmail.com

ABSTRAK

Clickbait menjadi salah satu cara untuk mencari pendapatan dengan meningkatkan *traffic* pembaca dan pengunjung. Semakin banyak pengunjung situs, semakin banyak pula kemungkinan untuk mendapatkan pendapatan dari situs tersebut. Banyak pengguna *clickbait* yang tidak mengutamakan kepuasan pembaca berita dengan cara menampilkan judul yang terkesan lebih – lebihkan, atau mengada – ada, tetapi isi konten tidak sesuai dengan yang tertera pada judul berita. Eksperimen yang dilakukan untuk mengklasifikasi artikel *clickbait* berbahasa Indonesia menggunakan metode Naïve Bayes classifier dengan menggunakan sampel dari judul berita online yang di ambil dari beberapa situs website. Selain itu juga melihat jumlah *like* dan *share* berita berdasarkan kategori yang sudah ditentukan yaitu *clickbait* dan bukan *clickbait*.

Kata Kunci: *naïve bayes*, klasifikasi, artikel berbahasa Indonesia, *clickbait*.

1. PENDAHULUAN

Internet adalah salah satu teknologi informasi yang sangat diminati oleh banyak kalangan, karena internet telah menyentuh berbagai aspek kebudayaan manusia mulai dari gaya hidup, pendidikan, penelitian hingga ke dunia bisnis. Salah satu manfaat dari internet adalah sebagai penyedia informasi yang cepat dan fleksibel, namun terkadang informasi yang diberikan tidak sesuai dengan judul artikel di media elektronik. Judul berita yang menjebak (*Clickbait*) merupakan modus media online untuk meningkatkan *traffic* pengunjung (Wiputra, 2015).

Clickbait menjadi salah satu cara untuk mencari pendapatan dengan meningkatkan *traffic* pembaca dan pengunjung. Semakin banyak pengunjung situs, semakin banyak pula kemungkinan untuk mendapatkan pendapatan dari situs tersebut. Faktor pendorong maraknya *clickbait* di media online adalah persaingan yang semakin ketat antar media untuk mendapatkan pembaca atau pengunjung dengan mengesampingkan kepuasan pengunjung situs menggunakan *clickbait* (Tea, 2016). Ciri-ciri *clickbait* adalah penggunaan kalimat yang hiperbola sehingga membuat pembaca tertarik untuk mengunjungi situs tersebut. *Clickbait* dapat menjadi alat yang sangat berguna bagi penyedia situs untuk menarik lebih banyak pembaca dengan tetap memperhatikan dan memberikan konten yang berkualitas (Akbar, 2016). Namun tidak sedikit yang menyalahgunakan *clickbait* dengan memberikan konten berita yang tidak sesuai dengan judulnya bahkan yang bersifat palsu atau dibuat-buat. Dapat dikatakan bahwa dengan adanya *clickbait* pembaca dapat dimanipulasi bahkan ditipu, yang kemudian pembaca tidak akan membagikan (*share*) situs karena merasa dirugikan.

Berdasarkan masalah penyalahgunaan *clickbait* diatas, diperlukan pengklasifikasian untuk mengidentifikasi artikel yang termasuk *clickbait* dan bukan. Penelitian sebelumnya tentang pengklasifikasian ini adalah identifikasi konten *e-government* untuk mengetahui perkembangan konten politik dan ekonomi berdasarkan frekuensi kata yang sering muncul pada sebuah artikel dengan menggunakan metode Naïve Bayes (Wijaya & Santoso, 2016). Sedangkan penelitian ini menggunakan metode Naïve Bayes untuk mengklasifikasikan judul artikel *clickbait*.

Algoritma *Naïve Bayes Classifier* (NBC) merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat. Kelebihan NBC adalah kemampuannya mengklasifikasikan dokumen dengan sederhana dan kecepatan komputasinya yang tinggi. Metode NBC juga memiliki kinerja yang baik untuk pengklasifikasian data dokumen yang mengandung angka maupun teks. Sebelum tahap klasifikasi, dokumen harus dipersentasikan menjadi vektor (Wijaya & Santoso, 2016).

Penerapan algoritma NBC dalam penelitian ini digunakan untuk mengklasifikasikan artikel berita Berbahasa Indonesia sehingga dapat dikelompokkan artikel *clickbait* dan bukan *clickbait*. Kemudian dilakukan pengumpulan data statistik berdasarkan jumlah *like* dan *share* pada artikel berita untuk mengetahui pengaruh dari *clickbait* yaitu perbandingan *like* dan *share* pengunjung situs antara artikel *clickbait* dan bukan *clickbait*.

2. ANALISA DAN PERANCANGAN

Analisa Masalah

Maraknya penyalahgunaan *clickbait* pada artikel media elektronik menyebabkan banyak pengunjung atau pembaca berita *online* merasa tertipu karena judul yang tertera pada berita atau artikel *online* menggunakan kata yang dilebih-lebihkan sehingga membuat pembaca tertarik untuk mengunjungi situs tersebut terlepas dari kesesuaian konten berita dengan judul yang dicantumkan. Hal ini dikarenakan persaingan yang semakin ketat antara media untuk mendapatkan *traffic* pembaca atau pengunjung dengan mengesampingkan kepuasan pengunjung situs. Semakin banyak pengunjung situs semakin banyak pula pendapatan yang diperoleh dari situs tersebut yang menjadikan *clickbait* sebagai sumber pendapatan,

Pemecahan Masalah

Untuk mengklasifikasi artikel *clickbait* dan bukan *clickbait* menggunakan metode Naive Bayes Classifier pada artikel berita berbahasa Indonesia. Naive Bayes Classification merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (atau aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat.

Analisa pengaruh *like* dan *share* pada artikel *clickbait* dan bukan *clickbait* dilakukan dalam pemecahan masalah, agar diketahui apakah penggunaan *clickbait* pada artikel berita dapat mempengaruhi jumlah *like* dan *share* pada artikel tersebut. Berdasarkan solusi yang telah disebutkan diatas, selain agar pembaca berita *online* dapat mengelompokkan berita *clickbait* maupun bukan *clickbait*, pengaruh penggunaan *clickbait* juga diharapkan dapat membantu penyedia berita *online* dalam melakukan inovasi baru terkait dengan memaksimalkan jumlah *like* dan *share* pada artikel berita.

Pengumpulan Data

Pengumpulan data dilakukan dengan mengumpulkan berita – berita pada situs berita *online*. Data yang dikumpulkan sebanyak 1000 data, yaitu 800 sebagai data latih, dan 200 sebagai data uji. Pengumpulan berita *online* tersebut dilakukan dengan cara *copy - paste* dari artikel *online* ke dalam Microsoft Excel. Data yang telah terkumpul, dilabeli sesuai dengan kategorinya, yaitu *clickbait* dan bukan *clickbait*. Pelabelan dilakukan secara manual, yaitu melabeli satu per satu artikel yang sudah terkumpul.

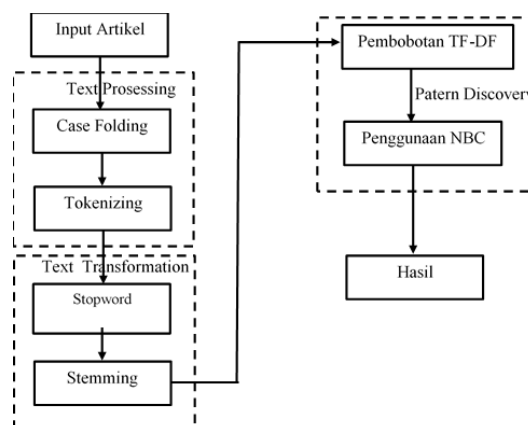
Pengelompokan artikel dilakukan melalui sensus data terhadap 5 orang responden yang memberikan nilai terhadap setiap artikel yang telah dikumpulkan yaitu artikel dengan kategori *clickbait* (0), ragu – ragu (1), dan bukan *clickbait* (2). Setelah semua nilai artikel telah diisi oleh responden,

kemudian dilakukan pemberian skor pada setiap data, pemberian skor diambil dari nilai rata – rata responden per data, sehingga apabila rata – rata yang didapatkan adalah lebih dari atau sama dengan satu maka hasil yang didapatkan adalah *clickbait*, sebaliknya apabila kurang dari satu maka hasilnya bukan *clickbait*.

Perancangan Aplikasi

Pengumpulan data dilakukan dengan mengumpulkan berita – berita pada situs berita *online*. Data yang dikumpulkan sebanyak 1000 data, yaitu 800 sebagai data latih, dan 200 sebagai data uji. Pengumpulan berita *online* tersebut dilakukan dengan cara *copy - paste* dari artikel *online* ke dalam Microsoft Excel. Data yang telah terkumpul, dilabeli sesuai dengan kategorinya, yaitu *clickbait* dan bukan *clickbait*. Pelabelan dilakukan secara manual, yaitu melabeli satu per satu artikel yang sudah terkumpul.

Pengelompokan artikel dilakukan melalui sensus data terhadap 5 orang responden yang memberikan nilai terhadap setiap artikel yang telah dikumpulkan yaitu artikel dengan kategori *clickbait* (0), ragu – ragu (1), dan bukan *clickbait* (2). Setelah semua nilai artikel telah diisi oleh responden, kemudian dilakukan pemberian skor pada setiap data, pemberian skor diambil dari nilai rata – rata responden per data, sehingga apabila rata – rata yang didapatkan adalah lebih dari atau sama dengan satu maka hasil yang didapatkan adalah *clickbait*, sebaliknya apabila kurang dari satu maka hasilnya bukan *clickbait*.



Gambar 1. Alur Tahapan Klasifikasi

Untuk mempermudah pemahaman, berikut adalah cara pengklasifikasian artikel yang sudah dilabeli dengan jumlah dokumen berjumlah 8:

Tabel 1. Data judul sebagai contoh kasus

Dokumen	Isi	Kategori
D1	Wow, Ternyata hanya Jokowi Presiden yang Terjun ke TKP Bom.	Clickbait
D2	Heboh Pria di Jagakarsa Gantung Diri Sambil Live Facebook.	Clickbait
D3	Astaga! Ternyata ada Guru Predator, Cabuli 35 Siswi Sendiri	Clickbait
D4	Persiapan Mudik Lebaran, Polisi Survei Beberapa Lokasi	Bukan Clickbait
D5	Jamaah Umroh asal Gresik Meninggal saat Tiba di Bandara Juanda	Bukan Clickbait
D6	Wow, Ternyata Ular King Kobra Minum dari Botol Air Mineral	Clickbait (Data Uji)
D7	Heboh! Polisi Gendong Biduan di Atas Panggung Jadi Viral	Clickbait (Data Uji)
D8	Ahok Bakal Putar Video Gus Dur di Sidang Ke-17	Bukan Clickbait (Data Uji)

Text Preprocessing

Dalam tahap ini kata menjadi elemen penting untuk proses pembangunan pengetahuan dan proses klasifikasi. Kata dapat didefinisikan sebagai kumpulan huruf, alfanumerik maupun tanda baca, yang diapit oleh spasi, kumpulan huruf alfabetik dengan beberapa tanda baca yang didefinisikan, atau sebagai kumpulan huruf alfabetik saja.

Pada tahap *text preprocessing*, langkah - langkah yang perlu dilakukan antara lain:

1. *Case Folding* yaitu mengubah semua karakter pada artikel menjadi huruf kecil, sehingga dari contoh sebelumnya maka hasil yang didapat adalah:

Dokumen	Hasil Case Folding
D1	wow, ternyata hanya jokowi presiden yang terjun ke tkp bom.
D2	heboh pria di jagakarsa gantung diri sambil live facebook.
D3	astaga! ternyata ada guru predator, cabuli 35 siswi sendiri
D4	persiapan mudik lebaran, polisi survei beberapa lokasi
D5	jamaah umroh asal Gresik meninggal saat tiba di bandara juanda
D6	wow, ternyata ular king kobra minum dari botol air mineral
D7	heboh! polisi gendong biduan di atas panggung jadi viral
D8	ahok bakal putar video gus dur di sidang ke-17

2. *Tokenizing* yaitu memecah kalimat menjadi satuan kata atau menghilangkan tanda baca.

Text Transformation

Sedangkan tahap *text transformation* akan dilakukan penghilangan *stopword*. Untuk setiap kata yang didapatkan, aplikasi akan melakukan pengecekan kedalam daftar *stopword*. Jika kata merupakan *stopword* maka kata itu akan dibuang. Jika bukan, maka kata tersebut akan diteruskan pada tahap *stemming*.

1. *stopword* yaitu menghilangkan kata hubung, atau kata – kata yang tidak deskriptif, sehingga dari proses *text preprocessing*.
2. *Stemming* yaitu merubah kata berimbuhan menjadi kata dasar.

Pattern Discovery

Setelah dilakukan tahap *Teks Preprocessing* dan *Text Transformation* selanjutnya dapat dilakukan tahap penentuan pola (*Pattern discovery*). *Pattern discovery* merupakan tahapan dari keseluruhan proses pengklasifikasian otomatis untuk menentukan pola atau pengetahuan dari keseluruhan teks. Pada tahapan ini algoritma Naive Bayes dapat diterapkan, namun sebelumnya terlebih dahulu

dilakukan perhitungan frekuensi kemunculan kata pada dokumen. Proses penghitungan TF-DF dibagi menjadi dua penghitungan, yaitu penghitungan TF dan penghitungan DF. TF (*Term Frequency*) merupakan frekuensi kemunculan suatu kata (t) pada dokumen (d), sedangkan DF (*Document Frequency*) adalah banyaknya dokumen yang mengandung suatu kata (t). Langkah awal perhitungan adalah menghitung tf berdasarkan kategori, kemudian menghitung df berdasarkan kategori, hasil dari perhitungana df masing masing kategori akan dimasukan ke dalam perhitungan algoritma Naive Bayes *classifier*.

Tabel 2. Contoh Penghitungan TF-DF berdasarkan kategori

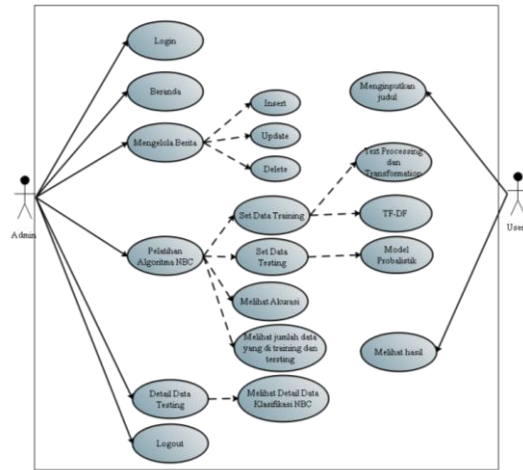
NO	TERM	TF			DF (C)			TF			DF (BC)		
		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	
1	wow	1	0	0	1	0	0	0	0	0	0	0	
2	nyata	1	0	1	2	0	0	0	0	0	0	0	
3	jokowi	1	0	0	1	0	0	0	0	0	0	0	
4	presiden	1	0	0	1	0	0	0	0	0	0	0	
5	terjun	1	0	0	1	0	0	0	0	0	0	0	
6	tkp	1	0	0	1	0	0	0	0	0	0	0	
7	bom	1	0	0	1	0	0	0	0	0	0	0	
8	heboh	0	1	0	1	0	0	0	0	0	0	0	
9	pria	0	1	0	1	0	0	0	0	0	0	0	
10	jagakarsa	0	1	0	1	0	0	0	0	0	0	0	
11	gantung	0	1	0	1	0	0	0	0	0	0	0	
12	diri	0	1	0	1	0	0	0	0	0	0	0	
13	sambil	0	1	0	1	0	0	0	0	0	0	0	
14	live	0	1	0	1	0	0	0	0	0	0	0	
15	facebook	0	1	0	1	0	0	0	0	0	0	0	
16	astaga	0	0	1	1	0	0	0	0	0	0	0	
17	ada	0	0	1	1	0	0	0	0	0	0	0	
18	guru	0	0	1	1	0	0	0	0	0	0	0	
19	predator	0	0	1	1	0	0	0	0	0	0	0	
20	cabul	0	0	1	1	0	0	0	0	0	0	0	
21	35	0	0	1	1	0	0	0	0	0	0	0	
22	siswi	0	0	1	1	0	0	0	0	0	0	0	
23	siapa	0	0	0	0	1	0	1	0	1	0	1	
24	mudik	0	0	0	0	1	0	1	0	1	0	1	
25	lebaran	0	0	0	0	1	0	1	0	1	0	1	
26	polisi	0	0	0	0	1	0	1	0	1	0	1	
27	survei	0	0	0	0	1	0	1	0	1	0	1	
28	lokasi	0	0	0	0	1	0	1	0	1	0	1	

Pelatihan Algoritma Naive Bayes Classifier

Secara umum proses pelatihan Naive Bayes dibagi menjadi beberapa tahap. Tahap-tahap tersebut dapat dilihat pada gambar 2 berikut:



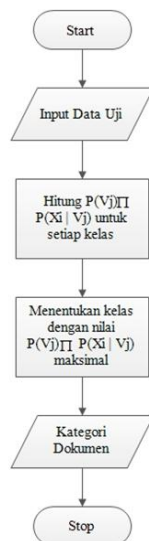
Gambar 2. Flowchart Proses Pelatihan Naïve Bayes Classifier



Gambar 4. Use Case Sistem Aplikasi

Pengujian Algoritma Naïve Bayes Classifier

Pada proses ini, data uji akan melewati proses klasifikasi berdasarkan data latih. Flowchart untuk tahap klasifikasi dapat dilihat pada Gambar 3.



Gambar 3. Flowchart Proses Klasifikasi Naïve Bayes Classifier

Use Case Sistem Aplikasi

Diagram use case dari sistem aplikasi pengklasifikasian artikel yang akan dikembangkan dapat dilihat pada gambar 4.

3. IMPLEMENTASI DAN PEMBAHASAN Text Preprocessing

Text preprocessing pada penelitian ini meliputi *casefolding* dan *tokenizing*, salah satu unsur suatu artikel merupakan artikel *clickbait* adalah terdapat tanda baca di dalamnya, dalam proses *tokenizing* terdapat banyak tanda baca yang hilang yang menandakan bahwa proses *tokenizing* merupakan kelemahan dalam proses penentuan kategori artikel.

Tabel 3. Ciri Tanda Baca Mengandung *Clickbait*

Artikel Clickbait	Proses Tokenizing
hakim cantik ini bikin cowok-cowok rela ? disidang?	hakim cantik ini bikin cowok-cowok rela disidang
jangan biarkan anak bermain slime jika tak ingin bernasib seperti bocah ini !	jangan biarkan anak bermain slime jika tak ingin bernasib seperti bocah ini
boy william kerap kerja bareng cewek cantik, apa reaksi sang pacar?	boy william kerap kerja bareng cewek cantik, apa reaksi sang pacar
"terungkap", ini alasan cewek jepang lebih suka cowok bertubuh pendek	terungkap ini alasan cewek jepang lebih suka cowok bertubuh pendek
pembelian server untuk unbk dananya ternyata...	pembelian server untuk unbk dananya ternyata

Dari 800 kata yang telah latih ditemukan beberapa contoh kata yang menjadi ciri – ciri *clickbait* hilang karena proses *tokenizing* sehingga artikel yang dianggap sebagai *clickbait* terkadang oleh sistem dianggap bukan *clickbait*.

Text Transformation

Text Transformation merupakan lanjutan dari proses *preprocessing* di mana terdapat dua proses dalam penentuan artikel *clickbait* atau bukan *clickbait*. Proses tersebut adalah *stopword* dan *stemming*. Setelah proses penelitian ada beberapa kata yang menjadi ciri kata unik dari artikel *clickbait* yang hilangkan saat proses *stopword* sebagai berikut

Tabel 4. Ciri Kata *Clickbait*

Kata	Frekuensi dan kategori
Wow	11 pada <i>Clickbait</i>
Heboh	16 pada <i>Clickbait</i>
Alasan	25 pada <i>Clickbait</i>
Astaga	4 pada <i>Clickbait</i>
Begini	32 pada <i>Clickbait</i>
kenapa	9 pada <i>Clickbait</i>
mengapa	11 pada <i>Clickbait</i>
bagaimana	4 pada <i>Clickbait</i>
siapa	8 pada <i>Clickbait</i>
Apa	11 pada <i>Clickbait</i>

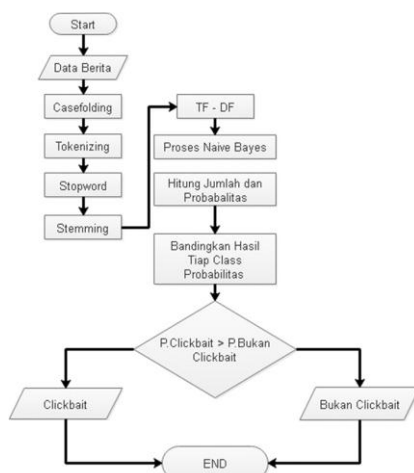
Sehingga dari data pada tabel 3 kata - kata yang seharusnya dihilangkan saat proses *stopword* kata tersebut tetap ditampilkan agar keakuratan data dalam pengklasifikasi kategori meningkat.

Pattern Discovery

Pattern discovery merupakan penentuan pola kata pada suatu kalimat guna menemukan jumlah kata pada masing-masing kategori *clickbait* maupun bukan *clickbait*. Dalam proses ini algoritma Naive Bayes dapat diterapkan, dengan terlebih dahulu menentukan frekuensi kemunculan kata pada dokumen, kemudian dilakukan perhitungan jumlah frekuensi dokumen pada masing-masing kategori *clickbait* dan bukan *clickbait*. Dalam uji penelitian menggunakan TF-DF tingkat akurasi masih rendah sehingga menggunakan bobot kata untuk melihat apakah akurasi data bisa naik ketika diuji akurasi semakin menurun dengan demikian penelitian ini menggunakan TF-DF.

Klasifikasi *Clickbait* dan Bukan *Clickbait*

Hasil nilai TF-DF akan dijadikan sebagai nilai masukan setiap artikel pada tahap pelatihan algoritma. Detail setiap tahapan pelatihan algoritma dapat dilihat pada Gambar 5.



Gambar 5. Sistem Klasifikasi *Clickbait* dan Bukan *clickbait*

Hasil Pengujian Algoritma

Hasil dari beberapa pelatihan dan pengujian yang telah dilakukan dapat dilihat pada tabel berikut:

Tabel 5. Hasil Akurasi Formula Penelitian

Formula	Jumlah Data Latih (RAND OM)	Jumlah Data Uji (RAND OM)	TP	FN	FP	TN	Total CB	Total Bukan CB	Conf. Matrix	Recall	Precision
1	50	100	57	28	10	5	85	15	62%	67%	85%
2	100	100	46	18	24	12	64	36	58%	72%	66%
3	150	100	48	18	24	10	66	34	58%	73%	67%
4	200	100	50	18	23	9	68	32	59%	74%	68%
5	250	100	43	16	27	14	59	41	57%	73%	61%
6	300	100	42	20	24	14	62	38	56%	68%	64%
7	350	100	45	12	30	13	57	43	58%	79%	60%
8	400	100	39	14	33	14	53	47	53%	74%	54%
9	450	100	39	17	30	14	56	44	53%	70%	57%
10	500	100	41	16	30	13	57	43	54%	72%	58%
11	550	100	41	14	27	18	55	45	59%	75%	60%
12	600	100	44	16	22	18	60	40	62%	73%	67%
13	650	100	36	17	28	19	53	47	55%	68%	56%
14	700	100	41	18	28	13	59	41	54%	69%	59%
15	750	100	37	19	31	13	56	44	50%	66%	54%
16	800	100	39	17	27	17	56	44	56%	70%	59%

Dari hasil formula pada Tabel 5 diketahui bahwa tingkat akurasi dalam pengklasifikasian algoritma Naive Bayes tergolong rendah. Berdasarkan hasil perhitungan dalam jurnal (Ishtiaq, 2014) diketahui bahwa nilai tingkat akurasi menggunakan metode Naive Bayes memiliki nilai rata – rata sebesar 80%, sedangkan dalam penelitian ini nilai rata – rata akurasi hanya sebesar 57%, sehingga algoritma ini kurang efektif untuk digunakan dalam klasifikasi. Pemilihan *confusion* matriks sebagai dasar dalam penentuan tingkat akurasi dikarenakan seluruh data digunakan dalam perhitungan. Dari 16 kali percobaan pada data latih dimulai dari data terkecil hingga data yang terbesar menunjukkan bahwa tingkat akurasi pada *confusion* matriks dalam percobaan pertama dan kedua belas memiliki nilai dengan tingkat akurasi yang paling tinggi yaitu 62%.

Perbandingan *Like* dan *Share*

Semua data yang telah dikumpulkan pada penelitian ini, mengambil sampel data dari empat alamat website yaitu detik.com, merdeka.com, tempo.com dan kompas.com. Dari data tersebut dihitung jumlah *like* dan *share* pada kategori *clickbait* dan bukan *clickbait*. Sehingga didapatkan data berikut:

Tabel 6. Pengaruh *Clickbait* pada Perbandingan *Like* dan *Share*

Website	Clickbait			Bukan Clickbait		
	data	like	share	data	like	share
Detik.com	49	6850	13716	125	82941	59199
Jawapos.com	109	14272	4639	10	51648	582
Merdeka.com	84	50368	61084	182	112396	198421
Tempo.com	41	6425	5650	23	22931	4468

Berdasarkan tabel di atas dapat diperoleh jumlah presentase tiap website yang dikategorikan *clickbait* dan bukan *clickbait*. Berikut adalah tabel presentase dari masing masing *website*.

Tabel 7. Tabel Persentase Dampak *Like* dan *Share* Website

Website	Clickbait		Bukan Clickbait	
	%like	%share	%like	%share
Detik.com	45%	33%	55%	67%
Jawapos.com	22%	89%	78%	11%
Merdeka.com	45%	31%	55%	69%
Tempo.com	22%	56%	78%	44%

4. KESIMPULAN DAN SARAN

Setelah dilakukan analisa, perancangan, implementasi, pelatihan algoritma, dan pengujian pada aplikasi pengklasifikasian artikel *clickbait* berbahasa Indonesia menggunakan Naïve Bayes Classifier maka dapat disimpulkan bahwa aplikasi yang dibuat mampu untuk mengklasifikasikan artikel berdasarkan kategorinya yaitu *clickbait* dan bukan *clickbait* secara otomatis. Namun saat uji pelatihan algoritma ini kurang baik dalam pengklasifikasian karena rata – rata akurasi dari data yang sudah dilatih hanya 56% dengan data latih 800 dan data uji 200.

Berdasarkan pengambilan sampel perhitungan yang telah dilakukan, dapat disimpulkan bahwa penggunaan *clickbait* sangat berpengaruh pada jumlah *like* dan *share*. Penggunaan bukan *clickbait* terhadap jumlah *like* berpengaruh 22% lebih tinggi dari pada menggunakan *clickbait*. Dan penggunaan bukan *clickbait* terhadap jumlah *share* 38% lebih tinggi dari pada menggunakan *clickbait*.

Berdasarkan penelitian yang telah dilakukan, sistem yang dibuat masih memiliki beberapa kekurangan. Beberapa saran yang dapat dijadikan acuan untuk pengembangan aplikasi atau sistem lebih lanjut adalah sebagai berikut:

1. Rata – rata akurasi pada penelitian ini dengan menggunakan Naïve Bayes classifier adalah berkisar 56% tergolong sangat rendah dari 16 kali percobaan yang telah dilakukan sehingga perlu dikaji ulang pemilihan fitur, dan formula yang bagus untuk mendeteksi artikel *clickbait*.
2. Pada penelitian ini data yang digunakan adalah data yang dipublikasikan dari berbagai periode yang mengakibatkan jumlah *like* dan *share* tidak signifikan, sehingga data yang di hasilkan kurang begitu akurat, serta ketidakseimbangan jumlah data yang dikumpulkan dari masing-masing situs sehingga membuat perhitungan menjadi berat sebelah.
3. Untuk pengembangan selanjutnya diharapkan penerapan algoritma Naïve Bayes classifier untuk mengklasifikasi artikel *clickbait* dapat diterapkan sebagai *plugin browser* , facebook, dan lain-lain sehingga artikel yang termasuk *clickbait* dapat dihilangkan atau dapat dideteksi oleh sistem.

4. Diharapkan penelitian selanjutnya terkait dengan *clickbait* dapat mendeteksi dan mengklasifikasi apakah judul artikel sesuai dengan konten pada artikel tersebut.

5. REFRENSI

- [1] Akbar, M. I. (2016, Agustus 31). Muhbalak. Retrieved from www.muhalak.com: <http://www.muhalak.com/2016/08/positif-dan-negatif-menggunakan-judul-yang-clickbait.html>
- [2] Basnur, P. W., & Sensue, D. I. (2010). Makara, Teknologi, Vol. 14, No. 1. Pengklasifikasian Otomatis Berbasis Ontologi Untuk, 33.
- [3] Dewobroto, W. 2005. Aplikasi Rekayasa Konstruksi dengan Visual Basic. Jakarta: PT Elex Media Komputindo.
- [4] Feldman, R., & Sanger, J. 2007. The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data. New York: Cambridge University Press.
- [5] Ganesan, K. 2006. Text Mining, Analytics & More. Retrieved from text-analytics101.rxnlp.com: <http://text-analytics101.rxnlp.com/2014/10/all-about-stop-words-for-text-mining.html>.
- [6] Han, J., & Kamber, M. 2006. Data Mining Concepts and Techniques.
- [7] Ishtiaq, A. 2014. Pasca Sarjana. Meachine Learning, 183 - 184.
- [8] Kadir, A. 2013. Pengenalan Algoritma. Yogyakarta: Andi.
- [9] Kunafi, A. 2009. Klasifikasi Email Berbahasa Indonesia Menggunakan Text Mining dan Algoritma KMean. Surabaya: Politeknik Elektronika Negeri Surabaya.
- [10] Pathmanaban, H. 2016. Academia. Retrieved from <https://www.academia.edu/>: https://www.academia.edu/5889117/Text_mining
- [11] Pedia, O. (2015, Februari 21). SI PEDIA. Retrieved from www.si-pedia.com: <http://www.si-pedia.com/2015/02/perbedaan-artikel-dengan-berita.html>
- [12] Prisma, D. (2014, November 1). ILMU PERPUSTAKAAN DAN INFORMASI.

- Retrieved from donyprisma.wordpress.com:
<https://donyprisma.wordpress.com/2014/11/01/recall-dan-precision-dalam-sistem-temu-kembali-informasi/>
- [13] Samodra, J., Sumpeno, S., & Hariadi, M. 2009. Pasca Sarjana. Informatics, 71-74.
- [14] Sanam, S. (2015, may 17). Sohib Sanam's blog. Retrieved from www.sohibsanam.com:
<http://www.sohibsanam.com/14-situs-berita-terkini-online-terpopuler-di-indonesia/>
- [15] Shire, E. (2014, July 14). wikipedia. Retrieved Desember 07, 2016, from <http://www.thedailybeast.com/articles/2014/07/14/saving-us-from-ourselves-the-anti-clickbait-movement.html>
- [16] Sido, F. (2015, Juni 23). Kompasiana. Retrieved from <http://www.kompasiana.com>:
http://www.kompasiana.com/afsee/media-mengenal-jebakan-klik_54f77e62a333112c6f8b45d8
- [17] Sumadiria, A. H. 2011. Menulis Artikel dan Tajuk Rencana: Panduan Praktis Penulis & Jurnalis Profesiona. Bandung: Simbiosis Rekatama Media.
- [18] Tea. (2016, DESEMBER 12). ROMELTEAMEDIA. Retrieved from <http://www.romelteamedia.com/2015/04/jangan-klik-clickbait-judul-berita-jebakan.html>:
<http://www.romelteamedia.com/2015/04/jangan-klik-clickbait-judul-berita-jebakan.html>
- [19] Thompson, D. (2013, November 14). theatlantic. Retrieved Desember 07, 2016, from <http://www.theatlantic.com/business/archive/2013/11/upworthy-i-thought-this-website-was-crazy-but-what-happened-next-changed-everything/281472/>
- [20] Trim, C. (2013, January 24). IBM Developer Works. Retrieved from www.ibm.com:
<https://www.ibm.com/developerworks/community/blogs/nlp/entry/tokenization?lang=en>
- [21] Waldman, K. (2014, May 23). National Post. Retrieved Desember 07, 2016, from <http://news.nationalpost.com/2014/05/23/mind-the-curiosity-gap-how-can-upworthy-be-noble-and-right-when-its-clickbait-headlines-feel-so-wrong/>
- [22] Wijaya, A. P., & Santoso, H. A. (2016). Technology. Naïve Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government, 48 - 55.
- [23] Wiputra, L. (2015, September 30). Judul *Clickbait* dan Kebiasaan Malas Membaca Orang Indonesia. Retrieved desember 12, 2016, from TECHINASIA:
<https://id.techinasia.com/talk/judul-clickbait-dan-kebiasaan-malas-membaca-orang-indonesia>
- [24] Zohar, Y. E. 2002. Introduction to Text Mining: Automated Learning Group National Center For Supercomputing Application. Illinois: University of Illioni.