

Klasifikasi *Tweet* Berbahasa Indonesia Berisi Ujaran Kebencian Menggunakan Metode *Improved K-Nearest Neighbor* dengan Pembobotan BM25F

Nurdifa Febrianti¹, Indriati², Muhammad Tanzil Furqon³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹difadifa@student.ub.ac.id, ²indriati.tif@ub.ac.id, ³m.tanzil.furqon@ub.ac.id

Abstrak

Ujaran kebencian ialah tindakan kebencian verbal yang menargetkan sekelompok orang atau bagian dari komunitas tertentu. Di Indonesia, ujaran kebencian semakin banyak ditemukan, terutama pada media sosial berbasis utama teks seperti Twitter. Sehingga menginspirasi ditulisnya penelitian ini, untuk mengidentifikasi ujaran kebencian di Twitter dengan klasifikasi *tweet*, khususnya yang berbahasa Indonesia. Penulis memilih menggunakan *Improved K-Nearest Neighbor* dengan menggunakan pembobotan kata BM25F, yaitu pembobotan yang mempertimbangkan *field/stream* dalam dokumen. Sehingga *tweet* yang dipilih sebagai dokumen latih dan dokumen uji penelitian, terdiri atas 2 *stream*, yaitu *tweet* dan *hashtag* atau tagar (tanda pagar). Dilakukan pengujian *K-Fold Cross Validation* (dengan $K = 5$) terhadap parameter k untuk klasifikasi IKNN, b_s , v_s , dan k_1 untuk pembobotan BM25F, dengan 400 dokumen latih dan 100 dokumen uji. Hasil pengujian menunjukkan bahwa penentuan nilai bobot stream pada BM25F cukup mempengaruhi hasil klasifikasi IKNN. Sedangkan hasil akhir terbaik untuk *F-Measure*, *Accuracy*, *Precision*, dan *Recall* dari rerata *5-Fold Cross Validation* yang didapatkan ialah 79,77% , 68,80%, 68,80%, dan 89,92% dengan $k = 70$, $b_s = 0,6$, $v_1 = 2$, $v_2 = 5$ dan $k_1 = 2$ sebagai nilai terbaik untuk masing-masing parameternya.

Kata kunci: Ujaran Kebencian, *Tweet*, *Hashtag*, *Improved K-Nearest Neighbor*, BM25F

Abstract

*Hate speech is a verbal hatred act that targets a group of people or parts of a particular community. In Indonesia, hate speech is increasingly found, especially on text-based social media such as Twitter. So that inspired the writing of this research, to identify hate speech on Twitter with the classification of tweets, especially those in Indonesian. The author chooses to use Improved K-Nearest Neighbor by using the BM25F term weighting, which is a weighting that considers the fields/streams in the document. So the tweet chosen as a training document and research test document, consists of 2 streams, the tweet and the hashtag. K-Fold Cross Validation testing (with $K = 5$) was performed on the parameter k for IKNN classification, b_s , v_s , and k_1 for BM25F weighting, with 400 training documents and 100 test documents. The test results show that the determination of stream weight values on BM25F sufficiently influences the results of the IKNN classification. Meanwhile the best final results for the *F-Measure*, *Accuracy*, *Precision*, and *Recall* of the average *5-Fold Cross Validation* obtained were 79.77%, 68.80%, 68.80%, and 89.92% with $k = 70$, $b_s = 0.6$, $v_1 = 2$, $v_2 = 5$ and $k_1 = 2$ as the best value for each parameter.*

Keywords: *Hate Speech*, *Tweet*, *Hashtag*, *Improved K-Nearest Neighbor*, BM25F

1. PENDAHULUAN

Ujaran kebencian merupakan tindakan kebencian verbal yang menargetkan sekelompok orang atau bagian dari komunitas tertentu berdasarkan ras, warna kulit, agama/keyakinan, jenis kelamin, kemampuan

fisik, orientasi seksual, garis keturunan, negara dan suku asal, atau bahkan pandangan politik.

Di Indonesia, ujaran kebencian semakin banyak ditemukan, terutama pada media sosial berbasis utama teks seperti Twitter. Ribuan masyarakat mengunggah *tweet* berisi ujaran

kebencian setiap harinya. Hal ini dapat memicu perseteruan antar golongan di Indonesia.

Bahkan Badan Reserse Kriminal Kepolisian Negara Republik Indonesia (Bareskrim Polri) telah menangani 1.829 kasus ujaran kebencian pada 2016. Setelah dibentuk Direktorat Tindak Pidana Siber Bareskrim tahun 2017, kasus ujaran kebencian yang ditangani oleh Polri menjadi 3.325 kasus (Movanita, 2017).

Hal ini menginspirasi penulis untuk memulai penelitian dalam mengidentifikasi ujaran kebencian di Twitter dengan klasifikasi *tweet*, khususnya yang berbahasa Indonesia.

Salah satu metode yang umum digunakan untuk klasifikasi ialah *K-Nearest Neighbor*. Pada penelitian Wah, dkk. (2016), kinerja KNN dibandingkan dengan metode *Support Vector Machine* dan *Logistic Regression* dalam klasifikasi *Imbalanced Dataset*, dan hasilnya menunjukkan bahwa KNN adalah metode yang menghasilkan akurasi terbaik bagi proses training maupun testing yaitu sebesar 96,7% dan 95,6%, dibandingkan dengan hasil *Support Vector Machine* sebesar 95,9% dan 95,5% serta hasil *Logistic Regression* sebesar 96,2% dan 95,6%.

Namun menurut penelitian Li, dkk. (2003), KNN memiliki kekurangan yang dipengaruhi oleh nilai k yang sama untuk semua kelas, padahal jumlah sampel dari tiap kelas belum tentu merata. Masalah ini diatasi dengan Improved K-Nearest Neighbor (IKNN), yaitu KNN dengan nilai k yang berbeda-beda untuk tiap kelas, dimana nilai k akan disesuaikan dengan proporsi jumlah data yang termasuk dalam masing-masing kelas pada kumpulan data latih untuk mengurangi kemungkinan salah klasifikasi.

Sebelumnya, sudah ada sejumlah penelitian perihal ujaran kebencian dengan objek *tweet*, salah satunya ialah oleh Alfina, dkk (2017). Penelitian ini memilih data berupa *tweet* terkait Pemilihan Gubernur DKI Jakarta 2017. Beberapa kata kunci yang terkait dengan pemilihan tersebut misalnya frasa “Pilkada Jakarta 2017” dan hashtag seperti “#DebatPilkadaDKI” serta “#SidangAhok”.

Klasifikasi IKNN oleh Li, dkk (2003) dan *Naïve Bayes* yang digunakan dalam penelitian Alfina, dkk. (2017) memiliki kesamaan, yaitu penggunaan *term weighting* atau pembobotan

kata pada teks. Kedua penelitian tersebut menggunakan metode pembobotan kata standar, yaitu TF-IDF. Tentunya terdapat beberapa metode lain untuk membobotkan kata.

Salah satu metode selain TF-IDF untuk pembobotan kata ialah BM25F. Metode BM25F ialah *upgrade* dari BM25 yang dapat diimplementasikan pada teks terstruktur dengan lebih dari satu *field* (disebut *stream* dalam BM25F) yang umumnya ialah *head*, *body*, dan sebagainya, dimana bobot TF tiap kata akan dikalikan dengan konstanta unik sesuai *field* masing-masing. Pada penelitian yang dilakukan oleh Perez-Aguera, dkk. (2010), dilakukan pengujian *Mean Average Precision* (MAP), P@5, P@10, *Geometric Mean Average Precision* (GMAP) dan *R-Precision* untuk membandingkan kinerja BM25, BM25F, Lucene, dan LuceneF pada *Semantic Web Search*. BM25F terbukti memiliki kinerja lebih baik pada MAP, GMAP dan R-Prec jika dibandingkan BM25 karena memperhitungkan perbedaan *field*. Salah satu pengujian terbaiknya menghasilkan nilai MAP, P@5, P@10, *Geometric Mean Average Precision* GMAP dan *R-Precision* dari BM25F secara berurutan sebesar 0.1743, 0.4412, 0.3765, 0.1030, dan 0.2172, dimana semua nilai tersebut lebih besar dibandingkan hasil pengujian BM25, Lucene, dan LuceneF.

Berdasarkan penjabaran metode-metode tersebut, penulis memutuskan untuk melakukan klasifikasi dengan metode *Improved K-Nearest Neighbor* dengan proses pembobotan kata menggunakan BM25F untuk *tweet* yang digunakan sebagai data latih dan data uji, dimana *tweet* yang dipilih ialah yang mengandung 2 komponen, yaitu isi *tweet* dan *hashtag*, yang dijadikan 2 *field* dalam BM25F.

2. DASAR TEORI

2.1. Ujaran Kebencian dan Penyebarannya di Media Sosial

Masyarakat menyalahgunakan asas kebebasan dalam berpendapat sebagai pembenaran mereka ketika ingin menjatuhkan maupun mencemarkan nama baik orang lain secara sengaja.

Penyebaran kebencian melalui tulisan saat ini semakin ‘didukung’ dengan adanya media sosial, dimana siapapun hanya perlu mengetik

beberapa kata sesuka hati mereka dan apapun yang tertera akan langsung dapat dibaca oleh jutaan pengguna lain.

Salah satu media sosial yang tercatat paling banyak digunakan untuk menyebarkan ujaran kebencian ialah Twitter.

Buku Saku Komisi Nasional Hak Asasi Manusia (KOMNASHAM RI, 2015) mengungkapkan bahwa penyebab bahayanya ujaran kebencian diantaranya ialah:

- 1) Merendahkan manusia lain: Manusia ialah ciptaan Tuhan dan tidak ada seorang pun yang berhak merendahkan manusia dan kemanusiaan seorang pun yang merupakan ciptaan Tuhan.
- 2) Menimbulkan kerugian materil dan korban manusia: Data penelitian menunjukkan jumlah kerugian material dan korban kekerasan berbasis identitas lebih besar daripada kekerasan lainnya.
- 3) Bisa berdampak pada konflik: Hasutan untuk memusuhi orang atau kelompok bisa menimbulkan konflik, konflik ini bisa antar individu dan meluas menjadi konflik komunal atau antar kelompok.
- 4) Bisa berdampak pada pemusnahan kelompok atau genosida: Hasutan kebencian ini bisa membuat *stereotyping*/pelabelan, stigma, pengucilan, diskriminasi, kekerasan. Pada tingkat yang paling mengerikan bisa menimbulkan kebencian kolektif pembantaian etnis, pembakaran kampung atau pemusnahan (genosida) terhadap kelompok yang menjadi sasaran ujaran kebencian.

2.2. Preprocessing Teks

Preprocessing ialah tahap pertama dalam mengelola teks. Sebelum melakukan klasifikasi dengan IKNN, tahap ini harus dilakukan untuk mendapatkan format teks yang siap untuk dijadikan data latih.

Preprocessing sangat penting dalam memahami makna penyampaian maksud pada kalimat/teks, terutama dalam media sosial dimana komunikasi yang berlangsung sebagian besar menggunakan kata-kata yang tidak formal dan tidak terstruktur serta memiliki noise yang besar (Mujilawati, 2016).

Tahapan *preprocessing* yang digunakan umumnya, dan juga akan digunakan dalam penelitian ini antara lain *Cleaning*, *Case folding*, Tokenisasi, *Filtering*, dan *Stemming*.

Untuk penelitian ini, terdapat penambahan tahap dalam penerapan *preprocessing*, yaitu *Split By Uppercase*, dimana tahap ini ialah inisiatif dari penulis untuk dikenakan pada *field hashtag*.

2.3. Pembobotan BM25F

Pada BM25, semua jenis dokumen dianggap sebagai satu kesatuan teks, tidak terstruktur dan tiap bagian yang ada di dalam teks tidak dibedakan. Maka BM25 tidak bisa digunakan untuk mendapatkan bobot kata dari teks terstruktur, yaitu teks yang isinya terbagi menjadi beberapa bagian.

Dari keterbatasan tersebut, algoritme BM25 dikembangkan menjadi BM25F, yang membagi teks terstruktur menjadi beberapa bagian (bisa disebut juga sebagai *field* atau *stream*). Contoh *stream* dalam teks yang paling umum ialah struktur *title/abstract/body* (Robertson & Zaragoza, 2009). Struktur ini dapat ditemukan di berbagai jenis teks seperti teks berita hingga artikel penelitian.

Menurut Robertson dan Zaragoza (2009), BM25F diawali dengan perhitungan BM25 tiap *term* pada masing-masing *stream*, kemudian hasil dari masing-masing *stream* akan dikombinasikan secara linear, dengan mengalikan hasil tersebut dengan bobot unik untuk tiap *stream* yang disebut dengan *stream weight*. Sehingga *term* yang sama akan memiliki bobot yang berbeda jika kata tersebut terdapat pada *stream* yang berbeda. Semakin suatu *stream* dianggap signifikan dalam proses pembobotan, maka harusnya nilai *stream weight* yang digunakan semakin besar.

BM25F diawali dengan perhitungan normalisasi panjang *stream* dalam dokumen. Normalisasi dilakukan untuk mengatasi dua kecenderungan dalam penulisan, yaitu *Verbosity* dan *Scope*. *Verbosity* ialah kecenderungan penulis menggunakan lebih banyak kata untuk mengatakan hal yang sama. Sedangkan *Scope* ialah kecenderungan menulis satu dokumen yang mencakup topic yang lebih luas. Perhitungan Normalisasi ditampilkan pada Persamaan 1.

$$B_s = \left((1 - b_s) + b_s \frac{sl_s}{avsl_s} \right) \quad (1)$$

Keterangan:

b_s = parameter penentu normalisasi (umumnya $0.5 \leq b_s \leq 0.8$). Bisa bernilai 0 jika tanpa normalisasi, bernilai 1 jika dengan normalisasi penuh.

sl_s = *stream length*, panjang *stream s* (dihitung dari jumlah *term* di masing-masing *stream*)

B_s = Komponen normalisasi panjang dokumen

$avsl_s$ = Rata-rata sl_s dari seluruh dokumen

Selanjutnya akan dihitung total jumlah kemunculan term (*term frequency* atau TF) yang dinormalisasi (*normalized-tf*) pada masing-masing *stream*, yang ditampilkan pada Persamaan 2.

$$\widetilde{tf}_i = \sum_{s=1}^S v_s \frac{tf_{si}}{B_s} \quad (2)$$

Keterangan:

\widetilde{tf}_i = Jumlah *normalized tf* dari seluruh *stream* dalam dokumen

s = *stream*, misal $s = 1$, maksudnya ialah bagian/*field* pertama dalam dokumen.

S = jumlah *stream* yang dimiliki suatu dokumen

B_s = Komponen normalisasi panjang dokumen

v_s = *stream weights* (bobot unik masing-masing *stream s* pada dokumen)

tf_{si} = *tf* dari *term i* dalam *stream s* dalam dokumen

Setelah menghitung *normalized-tf*, maka akan dicari w^{IDF} , yaitu bobot IDF yang telah dimodifikasi, ditampilkan pada Persamaan 3.

$$w_i^{IDF} = \log \frac{N - df_i + 0.5}{df_i + 0.5} \quad (3)$$

Keterangan:

N = Ukuran koleksi atau jumlah dokumen data latih

df_i = Jumlah dokumen dalam data latih yang mengandung kata i

w_i^{IDF} = modifikasi bobot IDF dari kata i

Langkah terakhir ialah perhitungan BM25F dari seluruh bagian dokumen, yang direpresentasikan dalam Persamaan 4.

$$w_d^{BM25F} = \sum_{i=1}^j \frac{\widetilde{tf}_i}{k_1 + \widetilde{tf}_i} w_i^{IDF} \quad (4)$$

Keterangan:

\widetilde{tf}_i = Jumlah *normalized tf* dari seluruh *stream* dalam dokumen

k_1 = konstanta saturasi, umumnya bernilai $1.2 < k_1 \leq 2$

w_i^{IDF} = modifikasi bobot IDF dari *term i*

j = jumlah *term* dari kueri (dokumen uji)

w_d^{BM25F} = Bobot BM25F semua *term* pada kueri dalam dokumen d

2.4. Klasifikasi Improved K-Nearest Neighbor

Improved K-Nearest Neighbor (IKNN), yaitu KNN dengan nilai k yang berbeda-beda untuk tiap kelas, berbeda dengan klasifikasi KNN yang menggunakan nilai k tetap. Secara umum, distribusi dokumen dari berbagai kelas dalam kumpulan data latih tidak merata. Beberapa kelas mungkin memiliki lebih banyak sampel daripada kelas yang lain. Oleh karena itu, sangat mungkin bahwa nilai k yang tetap akan menyebabkan kelas dengan jumlah sampel lebih banyak akan cenderung terpilih, walaupun sebenarnya tidak tepat (Li et al., 2003).

Menurut Li, dkk. (2003) dengan menggunakan tetangga terdekat sebanyak k awal, probabilitas bahwa satu dokumen termasuk dalam kelas tertentu dapat ditentukan dengan menggunakan tetangga terdekat sebanyak n untuk kelas tersebut, di mana n (nilai k baru) berasal dari k awal, sesuai dengan ukuran kelas tersebut dalam kumpulan data latih, dimana n ditampilkan pada Persamaan 5.

$$n = \left\lceil \frac{k \times N(C_m)}{\max\{N(C_j) \mid j=1, \dots, N_c\}} \right\rceil \quad (5)$$

Keterangan:

- n = nilai k baru
- k = nilai k awal
- $N(c_m)$ = Jumlah dokumen yang termasuk kelas m dalam kumpulan data latih
- $\max\{N(C_j) \mid j=1, \dots, N_c\}$ = Jumlah dokumen dalam kelas yang beranggota terbanyak pada kumpulan data latih

Berikutnya, akan dihitung peluang dokumen uji masuk ke masing-masing kelas yang tersedia, yang direpresentasikan oleh Persamaan 6.

$$P(x, C_m) = \operatorname{argsMax}_m \frac{\sum_{i=1}^n \operatorname{sim}(x, d_j) y(d_j, C_m)}{\sum_{i=1}^n \operatorname{sim}(x, d_j)} \quad (6)$$

Keterangan:

- $P(x, C_m)$ = Peluang dokumen x masuk ke kelas C_m
- $\operatorname{sim}(x, d_j)$ = kemiripan dokumen x dengan dokumen latih d_j
- $y(d_j, C_m)$ = parameter indikasi kelas, akan bernilai 1 jika d_j masuk ke kelas C_m , bernilai 0 jika tidak masuk ke kelas C_m
- n = jumlah k hasil modifikasi untuk IKNN
- $\operatorname{argsMax}_m$ = menandakan dipilihnya n data dengan nilai BM25F tertinggi dalam kelas m

3. METODOLOGI

3.1. Pengumpulan Data

Kumpulan dokumen yang dijadikan data latih dan data uji merupakan data primer berupa *tweet* yang disinyalir mengandung ujaran kebencian dan *tweet* yang tidak mengandung ujaran kebencian, agar selanjutnya dapat dikenakan proses klasifikasi. Penulis akan menggunakan 500 *tweet* yang akan dilabeli dengan bantuan Pakar, sebagai kumpulan dokumen latih dan uji. Teknik pengumpulan *tweet* ditempuh dengan 2 cara.

Cara pertama ialah secara manual dengan memanfaatkan fitur/tab *Search* pada website Twitter maupun aplikasi Twitter android, untuk mengumpullkan *tweet* berdasarkan beberapa

kata kunci dan *hashtag* tertentu, antara lain “#DebatPilpres2019”, “UninstallBukalapak”, “Prabohong” dan “#ErinTaulanySakitJiwa”.

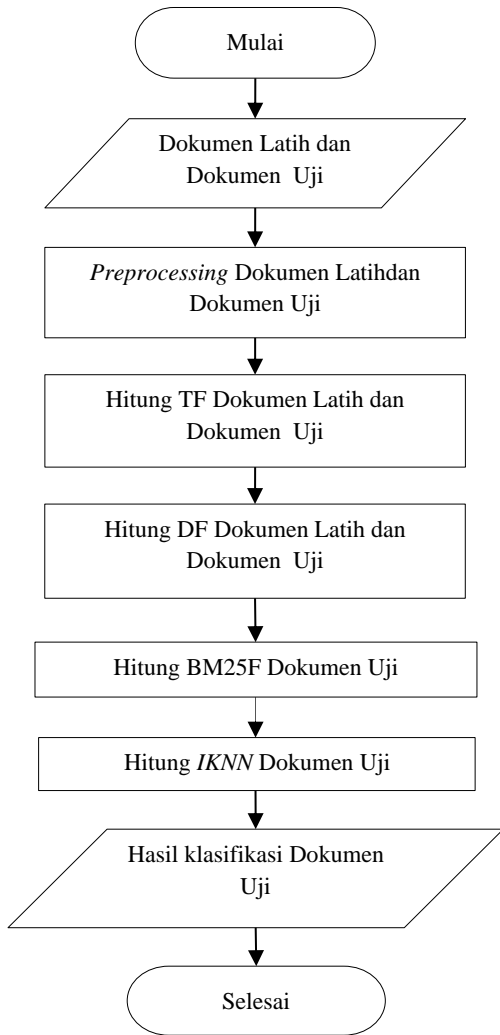
Cara kedua yaitu memanfaatkan aplikasi TAGS, yaitu template Google Sheet gratis yang dapat diatur untuk mengarsipkan *tweet* terbaru selama maksimal 7 hari terakhir berdasarkan tagar dan atau kata kunci, dimana penulis mencoba tagar “#Jokowi2Periode”, “#TetapMasihOposisi” dan kata “tolol”.

Pembagian antara dokumen latih dan uji akan menggunakan *K-Fold Cross Validation*, dimana penulis berencana menggunakan 5 Fold.

3.2. Diagram Alir Sistem

Sistem akan menerima masukan berupa dua dokumen berekstensi .csv, di antaranya ialah dokumen yang berisi 400 *tweet* dokumen latih dan 100 *tweet* data uji. Kedua dokumen memuat 2 kolom, yaitu kolom ‘*tweet*’ yang berisi kalimat *tweet* beserta tagar, dan kolom ‘*label*’ yang berisi label kelas masing-masing *tweet* yang telah ditentukan secara manual oleh penulis dengan bantuan pakar. Semua *tweet* akan dimasukkan ke dalam sistem untuk melalui *preprocessing*. *Preprocessing* yang dilakukan terdapat 5 tahap yaitu *cleaning*, *case folding*, tokenisasi, *split by uppercase*, *filtering* dan *stemming*. Keluaran setelah *preprocessing* berupa *term-term* pada setiap dokumen latih dan uji. Proses dilanjutkan dengan BM25F, yang terdiri atas beberapa tahap, yakni perhitungan *tf*, *df*, panjang dokumen yang dinormalisasi (B_s), *tf* yg dinormalisasi (\widetilde{tf}_i), yaitu *tf* setiap *term* di masing-masing *stream* yang dikalikan dengan bobot unik *stream* tersebut kemudian dibagi dengan nilai B_s , kemudian perhitungan nilai IDF yang sudah dimodifikasi (*wIDF*) dan bobot akhir BM25F masing-masing dokumen latih.

Selanjutnya terdapat klasifikasi IKNN, yang dilakukan dalam 2 tahap, yaitu menentukan nilai k sesuai dengan proporsi jumlah data yang masuk ke masing-masing kelas dalam klasifikasi, dan diikuti dengan proses perhitungan peluang sebuah data uji masuk ke masing-masing kelas yang ada. Alur proses sistem dapat dilihat pada Gambar 1 berikut.



Gambar 1. Alur Proses Sistem

4. PENGUJIAN DAN ANALISIS

4.1. Pengujian Sistem

Pengujian sistem dilakukan berdasarkan perbandingan *F-Measure* dari proses klasifikasi dengan parameter yang telah ditentukan. Pengujian sistem terdiri dari pengujian nilai *k* pada metode IKNN dengan menggunakan *5-Fold Cross Validation*, dan total data yang digunakan sebanyak 500 dokumen, sehingga pengujian dilakukan sebanyak 5 kali pada *5 Fold* data uji dan data latih yang berbeda-beda. Penelitian yang menggunakan *K-Fold Cross Validation* dengan $K = 5$ di antaranya adalah penelitian Yu, dkk. (2008) mengenai prediksi resiko kredit finansial menggunakan *Least-Squares Support Vector Machine*.

Kemudian akan diambil salah satu *k* pada rata-rata semua *Fold*, yang hasil pengujiannya yang memiliki *F-measure* tertinggi. *F-measure*

memperhitungkan *false positive* dan *false negative*, lebih relevan daripada *Accuracy* jika distribusi kelas dataset tidak merata. Nilai *k* tersebut akan digunakan sebagai variabel kontrol untuk menguji variasi penggunaan nilai b_s , v_s , dan k_1 sebagai variabel bebas.

Penelitian oleh Ma, dkk. (2014) menyebutkan tidak ada ketentuan tetap dalam memilih rentang nilai *k* untuk pengujian metode klasifikasi KNN. Sehingga dipilih $k = 3, 4, 5, 10$ dan semua kelipatan 10 hingga 100, sebab kelipatan 10 hingga 100 juga digunakan dalam penelitian oleh Miao, dkk. (2014) mengenai *Improved K-Nearest Neighbor* dengan data latih yang *Imbalanced*.

Hasil rata-rata *5-Fold* untuk variabel *k* ditampilkan di Tabel 1.

Tabel 1. Rata-rata Hasil *5-Fold Cross Validation* untuk nilai *k*

<i>k</i>	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>Accuracy</i>
3	68,57%	73,13%	70,60%	60,80%
4	68,06%	76,31%	71,84%	61,60%
5	69,86%	78,07%	73,62%	64,00%
10	71,14%	83,27%	76,68%	67,40%
20	69,92%	87,64%	77,73%	67,60%
30	70,05%	89,54%	78,56%	68,40%
40	67,85%	90,46%	77,47%	66,00%
50	67,32%	92,95%	78,23%	66,20%
60	67,65%	94,87%	78,90%	67,20%
70	67,66%	96,09%	79,32%	67,60%
80	67,23%	96,09%	79,01%	67,00%
90	66,79%	96,74%	78,92%	66,60%
100	66,73%	97,02%	78,97%	66,60%

Selanjutnya, hasil rata-rata *5-Fold* variabel b_s berdasarkan *k* terbaik berada di Tabel 2.

Tabel 2. Rata-rata Hasil *5-Fold Cross Validation* untuk nilai b_s

b_s	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>	<i>Accuracy</i>
0,5	67,66%	96,13%	79,33%	67,60%
0,6	67,79%	96,09%	79,42%	67,80%
0,7	67,66%	96,09%	79,32%	67,60%
0,8	67,66%	96,09%	79,32%	67,60%

Selanjutnya, hasil rata-rata 5-Fold variabel v_s berdasarkan b_s terbaik berada di Tabel 3.

Tabel 3. Rata-rata Hasil 5-Fold Cross Validation untuk nilai v_s

v_1	v_2	Precision	Recall	F-Measure	Accuracy
2	3	67,92%	95,39%	79,30%	67,80%
2	4	68,01%	95,11%	79,26%	67,80%
2	5	68,30%	95,11%	79,46%	68,20%
3	2	67,78%	95,81%	79,17%	67,40%
3	4	67,71%	95,11%	79,04%	67,40%
3	5	68,02%	95,13%	79,26%	67,80%
4	2	67,97%	95,85%	78,92%	66,80%
4	3	67,45%	96,41%	79,33%	67,40%
4	5	67,72%	95,77%	79,37%	67,60%
5	2	66,72%	95,82%	78,56%	66,20%
5	3	67,79%	96,09%	79,42%	67,80%
5	4	67,59%	96,41%	79,38%	67,60%

Selanjutnya, hasil rata-rata 5-Fold variabel v_s berdasarkan b_s terbaik berada di Tabel 4.

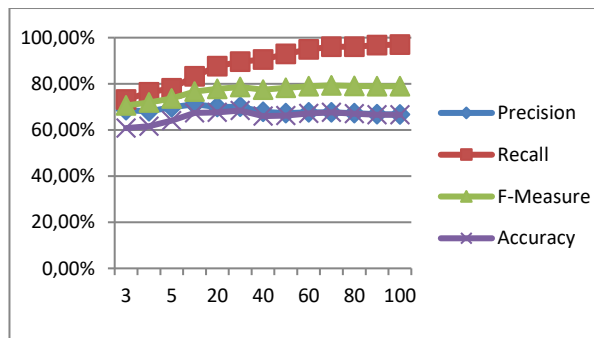
Tabel 4. Rata-rata Hasil 5-Fold Cross Validation untuk nilai k_1

k_1	Precision	Recall	F-Measure	Accuracy
1,3	68,38%	95,43%	79,61%	68,40%
1,4	68,37%	95,11%	79,46%	68,20%
1,5	68,37%	95,11%	79,46%	68,20%
1,6	68,37%	95,11%	79,46%	68,20%
1,7	68,37%	95,11%	79,46%	68,20%
1,8	68,37%	95,11%	79,46%	68,40%
1,9	68,60%	95,11%	79,67%	68,60%
2	68,60%	89,92%	79,77%	68,80%

4.2. Analisis Pengujian Sistem

Dari pengujian, ditemukan bahwa parameter terbaik yang bisa digunakan untuk klasifikasi dengan IKNN dan BM25F ialah $k = 70$, $b_s = 0,6$, $v_1 = 2$, $v_2 = 5$ dan $k_1 = 2$.

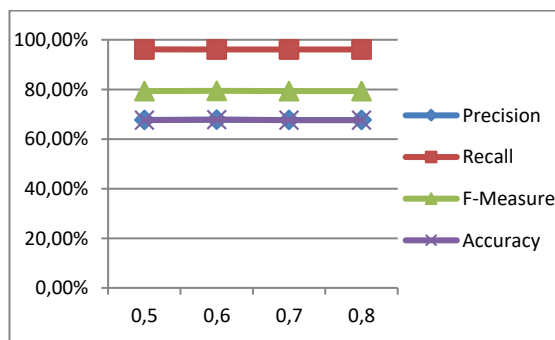
Hasil pengujian nilai k berada di Gambar 2.



Gambar 2. Grafik Rata-rata 5-Fold Cross Validation Nilai k

Nilai ketetangaan 70 akan menghasilkan klasifikasi terbaik. *Precision* mengalami penurunan (semakin banyak Non Ujaran Kebencian yang salah diklasifikasikan sebagai Ujaran Kebencian) dan *Recall* terus meningkat (semakin banyak Ujaran Kebencian yang berhasil diklasifikasikan dengan benar sebagai Ujaran Kebencian) seiring dengan semakin banyaknya tetangga yang diambil dalam klasifikasi. Diperkirakan disebabkan oleh jumlah anggota kelas Non Ujaran Kebencian jauh lebih sedikit (176 dokumen) dari pada Ujaran Kebencian (324 dokumen).

Hasil pengujian nilai b_s berada pada Gambar 3.

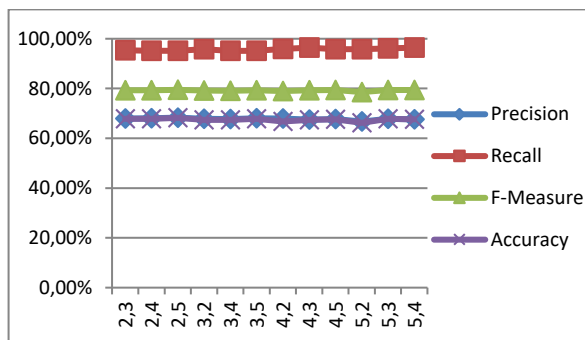


Gambar 3. Grafik Rata-rata 5-Fold Cross Validation Nilai b_s

Tidak ada pola tertentu atau perubahan nilai *F-measure* yang signifikan pada grafik Gambar 3, setiap metrik pengukuran memiliki rerata 5-Fold yang hanya berselisih sangat kecil (semua

berselisih kurang dari 0,5%) antara satu b_s ke b_s yang lain. Meski begitu, pengukuran F -measure, $Precision$ dan $Accuracy$ menunjukkan nilai terbaik pada saat $b_s = 0.6$, menandakan bahwa parameter normalisasi yang bernilai tidak terlalu besar dan tidak terlalu kecil pada panjang $stream$ tiap dokumen akan membuat hasil klasifikasinya semakin baik, walaupun pengaruhnya sangat kecil.

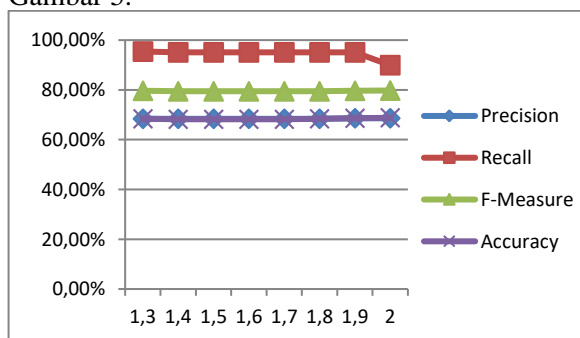
Hasil pengujian nilai v_s berada pada Gambar 4.



Gambar 4. Grafik Rata-rata 5-Fold Cross Validation Nilai v_s

Nilai pengukuran yang didapatkan memiliki pola $Recall$ dan F -measure selalu lebih tinggi dari pada $Accuracy$ dan $Precision$. Bahkan pada beberapa titik v_1 dan v_2 , $Accuracy$ dan $Precision$ hampir memiliki nilai yang sama. Dan sama seperti hasil pengujian b_s , F -measure, $Precision$ dan $Accuracy$ menunjukkan nilai terbaik pada saat bersamaan, yaitu $v_1 = 2$ dan $v_2 = 5$, yang menguatkan dugaan bahwa prioritas antara $stream$ isi $tweet$ dengan tagar memiliki bobot yang cukup berbeda.

Hasil pengujian nilai k_1 berada pada Gambar 5.



Gambar 5. Grafik Rata-rata 5-Fold Cross Validation Nilai k_1

Sama seperti pengujian sebelumnya, $Recall$ dan F -measure selalu lebih tinggi dari pada $Accuracy$ dan $Precision$. Bahkan nilai F -Measure, $Precision$ dan $Accuracy$ juga menunjukkan nilai terbaik pada saat bersamaan, yaitu $k_1 = 2$, namun tidak terlihat pola tertentu pada grafik yang signifikan, masing-masing F -Measure, $Precision$ dan $Accuracy$ pada setiap k_1 berada pada rentang yang sama. Hanya $Recall$ yang justru menunjukkan penurunan yang hampir tak terlihat pada grafik, namun menjadi sangat jelas antara $k_1 = 1,9$ ke $k_1 = 2$, yang menandakan bahwa semakin besar nilai k_1 sebagai parameter saturasi $term$, maka jumlah dokumen ujaran kebencian yang benar-benar diklasifikasikan ke kelas Ujaran Kebencian semakin menurun.

5. KESIMPULAN DAN SARAN

Kesimpulan yang dihasilkan dari penelitian ini ialah pemilihan bobot $stream$ pada BM25F cukup mempengaruhi hasil klasifikasi $tweet$ berbahasa Indonesia berisi Ujaran Kebencian dengan IKNN. F -Measure terbaik didapatkan ketika perbedaan bobot masing-masing $stream$ cukup jauh, dengan bobot untuk tagar sebesar 5, dan bobot untuk $tweet$ sebesar 2. Hal ini disebabkan oleh pelabelan Pakar pada dokumen latih yang lebih menitikberatkan ujaran kebencian melalui tagar yang digunakan, sehingga muncul kecenderungan ini pada hasil pengujiannya. Hasil akhir terbaik untuk F -Measure, $Accuracy$, $Precision$, dan $Recall$ dari rerata 5-Fold Cross Validation yang didapatkan ialah 79,77% , 68,80%, 68,80%, dan 89,92% dengan $k = 70$, $b_s = 0,6$, $v_1 = 2$, $v_2 = 5$ dan $k_1 = 2$. Nilai $Accuracy$ yang tidak tinggi kemungkinan disebabkan oleh beberapa alasan, di antaranya ialah tidak imbangnya jumlah dokumen latih yang berlabel Ujaran Kebencian dengan yang berlabel Non Ujaran Kebencian.

Saran yang diajukan ialah melakukan $oversampling$ atau $undersampling$ apabila menggunakan dataset yang imbalance untuk mengatasi kecenderungan atau bias klasifikasi terhadap kelas mayoritas, sebaiknya kumpulan data dapat di- $shuffle$ terlebih dahulu sebelum dilakukan K -Fold Cross Validation, dan dapat dilakukan proses perbaikan kata tidak baku maupun singkatan kata dalam data latih dan data uji untuk mengurangi adanya term yang bermakna sama namun penulisannya berbeda.

6. DAFTAR PUSTAKA

- Alfina, I., Mulia, R., Fanany, M.I. & Ekanata, Y., 2017. Hate Speech Detection in the Indonesian Language: A Dataset and Preliminary Study. In *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. Bali, 2017. IEEE.
- KOMNASHAM RI, 2015. In *Penanganan Ujaran Kebencian (Hate Speech)*. Jakarta: KOMNASHAM. p.3.
- Li, B., Yu, S. & Lu, Q., 2003. *An Improved k-Nearest Neighbor Algorithm for Text Categorization*. [Online] Shenyang: Cornell University Tersedia di: <https://arxiv.org/abs/cs/0306099> [Diakses 15 Juli 2019].
- Ma, C.M., Yang, W.S., Cheng, B.W, 2014. How the Parameters of K-nearest Neighbor Algorithm Impact on the Best Classification Accuracy: In Case of Parkinson Dataset. *Journal of Applied Sciences*, 14(2), pp. 171-76.
- Miao, Z., Tang, Y., Sun, L., He, Y., Xie, S., 2014. An Improved KNN Algorithm for Imbalanced Data Based on Local Mean. *Journal of Computational Information Systems*, 10(12), pp. 5139-46.
- Movanita, A.N.K., 2017. *Ini Hasil Kerja Polri Perangi Kejahatan Siber Sepanjang 2017*. [Online] PT. Kompas Cyber Media Tersedia di: <https://nasional.kompas.com/read/2017/12/29/17233911/ini-hasil-kerja-polri-perangi-kejahatan-siber-sepanjang-2017> [Diakses 2 Juli 2019].
- Mujilawati, S., 2016. PRE-PROCESSING TEXT MINING PADA DATA TWITTER. In *Seminar Nasional Teknologi Informasi dan Komunikasi (SENTIKA) 2016*. Yogyakarta, 2016. Universitas Islam Lamongan.
- Perez-Aguera, J.R. et al., 2010. Using BM25F for Semantic Search. In *SEMSEARCH '10 Proceedings of the 3rd International Semantic Search Workshop Proceedings of the 3rd International Semantic Search Workshop*. New York, 2010. ACM.
- Robertson, S. & Zaragoza, H., 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends in Information Retrieval*, 3(4), pp.333-89.
- Wah, Y.B., Rahman, H. A. A., He, H. & Bulgiba, A., 2016. Handling imbalanced dataset using SVM and k-NN approach. *AIP Conference Proceedings*, 1750(1).
- Yu, L., Wang, S., Lai, K.K., Zhou, L., 2008. *Bio-Inspired Credit Risk Analysis: Computational Intelligence with Support Vector Machines*. Springer Science & Business Media.