

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/327765350>

# Klasifikasi Berita pada Twitter Menggunakan Metode Naive Bayes dan Query Expansion Hipernim–Hiponim

Article · January 2018

CITATION

1

READS

500

3 authors:



**Fakhruddin Irfani**

Brawijaya University

2 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)



**Mochammad Ali Fauzi**

Brawijaya University

102 PUBLICATIONS 206 CITATIONS

[SEE PROFILE](#)



**Yuita Arum Sari**

Brawijaya University

101 PUBLICATIONS 102 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Smart Wheelchair [View project](#)



Automatic Essay Scoring [View project](#)

## Klasifikasi Berita pada Twitter Menggunakan Metode Naive Bayes dan Query Expansion Hipernim-Hiponim

Fakhruddin Farid Irfani<sup>1</sup>, M. Ali Fauzi<sup>2</sup>, Yuita Arum Sari<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya

Email: <sup>1</sup>udinirfan17@gmail.com, <sup>2</sup>moch.ali.fauzi@ub.ac.id, <sup>3</sup>yuita@ub.ac.id

### Abstrak

Banyaknya jumlah *tweet* yang dipost mengakibatkan *tweet* yang tersebar dan muncul dalam beranda Twitter sangat beragam dan tidak dikelompokkan berdasarkan kategori beritanya seperti kesehatan, olahraga, teknologi, ekonomi, wisata dan lain sebagainya. Tidak adanya pengkategorian menyebabkan pengguna kesulitan jika ingin membaca atau mengambil informasi terkait kategori tertentu yang diinginkan. Salah satu solusi yang dapat dilakukan adalah dengan metode klasifikasi teks, yang dalam proses klasifikasinya mampu mengklasifikasikan secara otomatis terhadap beberapa kategori pada teks tidak terstruktur dengan bahasa alami. Pada penelitian ini dilakukan proses klasifikasi menggunakan metode *Naive Bayes* dengan tambahan *query expansion* untuk menambahkan *term* pada dokumen awal. Penambahan *term* bertujuan untuk mengoptimalkan proses klasifikasi dikarenakan *tweet* merupakan *short text* yang dapat menimbulkan ambiguitas kelas klasifikasi. Penambahan yang dilakukan adalah hiponim dan hipernim dari dokumen asli yang diambil dari WordNet. Metode perhitungan akurasi yang digunakan adalah *k-fold* yang bertujuan untuk menghindari *over fitting* pada dokumen. Akurasi yang didapatkan adalah sebesar 72% untuk klasifikasi tanpa *query expansion*, 65,75% untuk penambahan hiponim dan hipernim, 66,3% untuk penambahan hiponim saja, dan 67,5% untuk penambahan hipernim saja. Dapat disimpulkan bahwa penambahan *query* yang dilakukan kurang efektif untuk meningkatkan akurasi proses klasifikasi.

**Kata kunci:** *Tweet, Naive Bayes, WordNet, Hiponim, Hipernim, K-Fold*

### Abstract

The large number of posted tweets resulted in scattered tweets and appearing on the Twitter homepage very diverse and not classified by categories such as health, sports, technology, economics, tourism and so on. The absence of categorization causes the user difficulty to read or retrieve information related to certain desired categories. Solution that can be done is by the method of text classification, which in the process of classification is able to classify automatically against some categories on unstructured text with natural language. In this research will be done classification process using Naive Bayes method with additional query expansion to add term in initial document. The addition of term aims to optimize the classification process because the tweet is a short text that can lead to ambiguity of classification class. The additions made are hyponym and hypernym from original documents extracted from WordNet. Accuracy calculation method used is *k-fold* that aims to avoid over fitting on the document. The accuracy obtained was 72% for the classification without query expansion, 65.75% for hyponym and hypernym addition, 66.3% for hyponym addition, and 67.5% for hypernym addition. It can be concluded that the addition of queries made less effective to improve the accuracy of the classification process.

**Keywords:** *Tweet, Naive Bayes, WordNet, Hyponym, Hypernym, K-Fold*

### 1. PENDAHULUAN

Pada perkembangan zaman saat ini pengguna internet semakin banyak dan tersebar meluas, salah satu penggunaan internet adalah *electronic news (e-news)* yang merupakan

berita yang disebar dan diberitakan secara online. Seperti halnya berita pada media cetak maupun media-media lain, berita dalam *e-news* sangat beragam seperti olahraga, politik, hiburan dan lain sebagainya. Jika digunakan dengan sesuai dan tepat guna, maka dapat

meningkatkan kualitas dan memajukan dari berbagai bidang seperti pendidikan, bisnis, ilmu pengetahuan, dan sosial.

Bersamaan dengan bertambahnya pengguna internet maka media sosial juga semakin diminati, salah satunya adalah Twitter yang mengalami pertambahan jumlah pengguna setiap tahunnya. *Tweet* yang tersebar dan muncul dalam beranda Twitter salah satunya adalah *tweet* berita yang sangat beragam serta tidak dikelompokkan berdasarkan kategori beritanya seperti kesehatan, olahraga, teknologi, ekonomi, wisata dan lain sebagainya. Tidak adanya pengkategorian menyebabkan pengguna kesulitan jika ingin membaca atau mengambil informasi terkait katogori tertentu yang diinginkan.

Salah satu solusi yang dapat dilakukan adalah dengan metode klasifikasi teks, yang dalam proses klasifikasinya mampu mengklasifikasikan secara otomatis terhadap beberapa kategori pada teks tidak terstruktur dengan bahasa alami. Salah satu metode klasifikasi adalah *Naïve Bayes*. Contoh penerapan metode *Naïve Bayes* adalah pada penelitian yang dilakukan oleh Perdana pada tahun 2013, menunjukkan bahwa *Naïve Bayes* pada klasifikasi *tweet* berbahasa indonesia menghasilkan performa yang baik, menggunakan metode pengukuran akurasi dengan *precision*, *recall*, dan *F1 measure* menghasilkan nilai masing-masing yaitu 80%, 79%, dan 78%. Pada penelitian lain yang dilakukan oleh Buzic dan Dobsa pada tahun 2018, menerapkan *Naïve Bayes* untuk klasifikasi lirik lagu mendapatkan akurasi sebesar 88,4% dan metode pengukuran *precision*, *recall*, dan *F1* menghasilkan nilai masing-masing yaitu 86,9%, 95,2% dan 90,9%. Dari penelitian-penelitian sebelumnya, maka dapat dipahami bahwa *Naïve Bayes* cocok untuk melakukan klasifikasi teks.

Pada proses klasifikasi teks terdapat beberapa permasalahan. Salah satu masalah pada proses klasifikasi teks adalah kata yang menjadi *query* tidak memiliki karakteristik yang cukup untuk menggambarkan sebuah kelas (Sriram, 2010). Salah satu cara untuk menangani permasalahan tersebut adalah dengan menggunakan metode *Query Expansion*. Terdapat beberapa penelitian yang telah menerapkan *Query Expansion*. Salah satu penggunaan *query expansion* adalah penelitian yang dilakukan oleh Bagaskoro 2017, peningkatan akurasi yang dihasilkan dengan

menambahkan *query expansion* yaitu sebesar 2%, sebelum menggunakan *query expansion* akurasi sistem sebesar 90% lalu menjadi 92%. Selain itu pada penelitian lain yang dilakukan oleh Tanjung 2017 menunjukkan peningkatan akurasi yang cukup signifikan dengan penambahan metode *query expansion*. Besaran akurasi tanpa menggunakan *query expansion* adalah 58% sedangkan setelah ditambahkan menjadi 76%. Maka dari itu metode *Query Expansion* terbukti dapat meningkatkan akurasi dari proses klasifikasi.

Ada berbagai macam metode dalam melakukan *query expansion*. Salah satu metode adalah dengan menggunakan WordNet, WordNet merupakan sebuah kamus elektronik yang sudah dikembangkan. Dalam WordNet berisi *synonym set* (*synset*) (Li, Ganguly, & J.F.Jones, 2016). Pada penelitian ini akan dilakukan proses klasifikasi menggunakan metode *Nive Bayes* dengan penambahan *query exansion* yang berbbasis WordNet. Kata yang ditambahkan adalah hipernim dan hiponim dari dokumen asli. Karena WordNet yang digunakan berbahasa inggris maka sebelum dan sesudah dilakukan proses *query expansion* dokumen akan diterjemahkan menggunakan *google translate API*.

## 2. DATA DAN METODE

### 2.1 Dataset

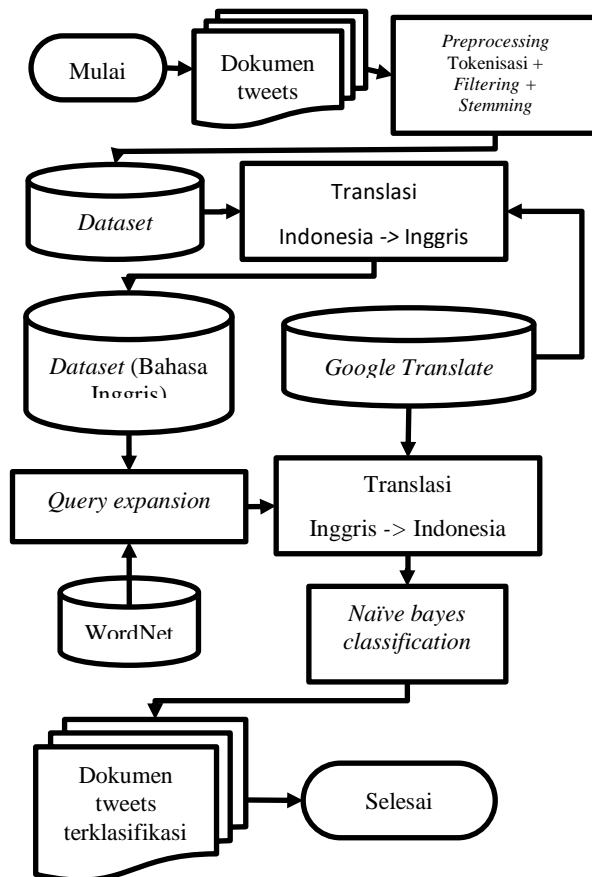
Pada penelitian ini *dataset* yang digunakan adalah data yang telah ada sebelumnya dan telah digunakan dalam penelitian oleh Bagaskoro (2017) dengan jumlah *dataset* sebanyak 400 data. Data *tweet* tersebut didapatkan dari akun Twitter Kompas dan Detik.com (Bagaskoro,Fauzi,Adikara 2017). Berikut adalah contoh dokumen yang digunakan yang digunakan pada Tabel 1.

Tabel 1 Contoh Dataset

No	<i>Tweet</i>	Kelas
1	Awali Tahun 2016, Rupiah Berpeluang Menguat	Ekonomi
2	Tunda Pungutan Dana Ketahanan Energi	Ekonomi
3	Ini 5 Mata Uang Paling Jeblok Tahun 2015	Ekonomi
4	Getarkan DWP 2015, Dipha Barus Gandeng Rinni Wulandari	Entertainment

## 2.2 Metode usulan

Penelitian ini melakukan proses klasifikasi dokumen *tweet* menggunakan metode *Naive Bayes* dengan penambahan metode *query expansion*. Berikut adalah diagram alir dari metode yang diusulkan pada Gambar 1. Proses dimulai dengan *preprocessing* pada seluruh *dataset*. Hasil *preprocessing* kemudian diterjemahkan kedalam bahasa inggris untuk selanjutnya dicari hipernim dan hiponimnya. Setelah didapatkan hipernim dan hiponimnya kemudian ditambahkan kedalam dokumen asli. Dokumen penambahan kemudian diterjemahkan kembali ke dalam bahasa indonesia untuk kemudian dilakukan proses klasifikasi.



Gambar 1 Diagram Alir Metode

### 2.2.1 Preprocessing

Tahapan pertama yang dilakukan pada *dataset* adalah *preprocessing*. Tujuan dari *preprocessing* adalah mengolah data yang sebelumnya tidak terstruktur menjadi lebih terstruktur sehingga dapat diolah lebih lanjut. Tahapan-tahapan yang dilakukan dalam *preprocessing* adalah (Vijayaran, Ilamathi, & Nithya, 2015) :

1. Tokenisasi
2. *Filtering / stopwords removal*
3. *Stemming*

### 2.2.2 Query Expansion

Setelah melalui tahap *preprocessing* pada seluruh *dataset* yang digunakan. Selanjutnya adalah penambahan *query* atau *query expansion* berbasis WordNet. Penambahan *query* dilakukan dengan menambahkan kata lain yang memiliki keterkaitan dengan *query*. (Zhang, Deng, & Li, 2009) Penambahan yang dilakukan berupa relasi kata hipernim dan hiponim. Proses dimulai dengan dokumen yang telah melalui tahap *preprocessing*. Dokumen tersebut kemudian diterjemahkan ke dalam bahasa inggris terlebih dahulu, karena WordNet menggunakan bahasa inggris. Setelah diterjemahkan menggunakan *google translate API* selanjutnya adalah mencari hipernim dan hiponim dari tiap-tiap *term* yang ada pada dokumen. Selanjutnya dilakukan penambahan *query* dengan menambahkan hipernim dan hiponim dari kata yang telah dicari sebelumnya. Terakhir dokumen dikembalikan ke bahasa indonesia untuk diolah dan dilakukan proses klasifikasi.

### 2.2.3 Naive Bayes

Tahapan selanjutnya adalah proses klasifikasi pada dokumen hasil *query expansion*. Proses dimulai dengan melakukan perhitungan nilai *prior* yang merupakan probabilitas kemunculan kelas dibandingkan dengan jumlah seluruh kelas, perhitungan *prior* dilakukan sebanyak kelas yang ada.

Selanjutnya menghitung jumlah *term* unik yang ada pada data latih ( $V$ ). *Term* unik merupakan *term* dari seluruh kelas. Selanjutnya adalah menghitung kemunculan setiap *term* ( $w_i$ ) yang ada pada dokumen, dihitung jumlah kemunculan kata tersebut dalam tiap-tiap kelas ( $c_j$ ). Nilai yang dihitung selanjutnya adalah jumlah kata yang ada pada setiap kelas, kata yang dihitung bukan kata unik melainkan keseluruhan kata yang ada pada kelas ( $\sum_{w \in V} count(w_i, c_j)$ ). Setelah didapatkan nilai-nilai tersebut maka selanjutnya adalah menghitung *probabilitas* setiap kata terhadap setiap kelas yang ada (*likelihood*). Rumus perhitungan *likelihood* dituliskan dalam Persamaan 1 (Rahman, Wiranto, & Doewes, 2017)

$$P(w_i|c_j) = \frac{\text{count}(w_i, c_j) + 1}{(\sum_{w \in V} \text{count}(w, c_j) + |V|)} \quad (1)$$

Nilai *likelihood* yang didapatkan kemudian dikalikan antara satu kata dengan kata lainnya dalam satu dokumen dan dikalikan posterior dari kelas, sehingga dihasilkan probabilitas dokumen terhadap tiap-tiap kelas yang ada (*posterior*). Rumus perhitungan *posterior* dituliskan pada Persamaan 2. (Rahman, Wiranto, & Doewes, 2017)

$$P(c_j|d_i) = P(c) \times P(w_1|c_j) \times P(w_2|c_j) \dots \times P(w_n|c_j) \quad (2)$$

Dari *posterior* yang ada kemudian diambil yang memiliki nilai paling besar dan menjadi kelas dari dokumen tersebut.

### 3. HASIL DAN PEMBAHASAN

Dalam penelitian ada 3 jenis pengujian yang dilakukan.

1. Pengujian perhitungan akurasi tanpa melakukan *query expansion* (*k-fold*)
2. Pengujian variasi *threshold* jumlah kata yang ditambahkan
3. Pengujian jenis kata yang ditambahkan (hipernim dan hiponim)

#### 3.1 Penghitungan Akurasi Menggunakan K-Fold

Pada pengujian ini dilakukan perhitungan akurasi menggunakan metode *k-fold* menggunakan 5 *fold*. Dokumen-dokumen yang terdapat di dalam setiap *fold* dipilih secara manual oleh Penulis. Pemilihan dokumen dilakukan secara acak agar menghindari adanya manipulasi data. Perhitungan akurasi sistem dilakukan dengan merata-rata jumlah dari seluruh akurasi *fold*. Berikut adalah hasil perhitungan akurasi dari masing-masing *fold* yang dituliskan ke dalam Tabel 2.

Tabel 2 Akurasi Sistem Menggunakan K-Fold

K	Akurasi (%)
1	77,5
2	83,75
3	75
4	67,5
5	56,25
Sistem	72

Hasil klasifikasi tanpa menggunakan *query expansion* memiliki akurasi sebesar 72%. Hasil akurasi dari tiap *fold* tidak memengaruhi *fold* yang lain. Rendahnya tingkat akurasi pada proses klasifikasi *tweet* ini disebabkan oleh *dataset* yang digunakan. Persebaran dari dokumen dengan *term* tertentu sangat memengaruhi hasil dari proses klasifikasi, maksudnya jika sebuah *term* hanya muncul pada satu *fold* maka dipastikan proses klasifikasi akan menjadi kurang maksimal. Selain itu, dari *dataset* yang digunakan beberapa dokumen dengan kelas tertentu memiliki satu atau beberapa *term* yang sama dengan kelas lain sehingga juga menyebabkan proses klasifikasi menjadi tidak optimal. Contoh dari dokumen yang memiliki kesamaan *term* dituliskan dalam Tabel 3.

Tabel 3 Dokumen Dengan Kesamaan Term

Dokumen	Kelas	Fold
Pantai Karang Jahe, Wisata Bahari Andalan dari Rembang	Ekonomi	2
Yang beda di Thailand, Wisata ke Kebun Anggur	Travel	2
Pengusaha Pesawat Carter Bidik Sektor Pariwisata dan 'Taksi' Udara	Travel	3

#### 3.2 Pengujian Variasi Threshold

Pada tahap ini dilakukan proses *query expansion* terhadap *term* dokumen asli. Jumlah dari kata yang ditambahkan kemudian divariasi dari range 0,1 - 1,0 untuk mendapatkan akurasi tertinggi berdasarkan jumlah kata ekspansi yang ditambahkan. Pada pengujian ini, kata ekspansi yang ditambahkan merupakan gabungan antara hipernim dan hiponim dari *term*. Hasil dari variasi *threshold* dari kata ekspansi yang ditambahkan ditampilkan ke dalam Tabel 4.

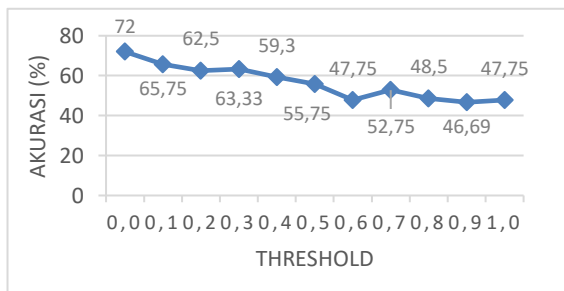
Tabel 4 Akurasi Variasi Threshold

Threshold	Akurasi (%)
0,0	72,0
0,1	65,75
0,2	62,5
0,3	63,3
0,4	59,25
0,5	55,75
0,6	47,75
0,7	52,75
0,8	48,5
0,9	46,69
1,0	47,75



Pada Tabel 4, dimulai dari *threshold* 0,0 atau tidak melalui proses *query expansion* dengan akurasi sebesar 72,0 %. Pada *threshold* selanjutnya akurasi mengalami penurunan sebesar 6,25%, akurasi sistem menjadi 65,75% pada *threshold* 0,1. Nilai akurasi kemudian mulai mengalami kecenderungan penurunan hingga pada *threshold* 1,0.

Untuk lebih jelas dalam melihat peningkatan dan penurunan akurasi maka nilai akurasi ditampilkan dalam bentuk grafik pada Gambar 2.



Gambar 2 Akurasi Variasi *Threshold*

Penambahan berdasarkan *threshold* dilakukan dengan mengalikan panjang dokumen (banyaknya kata ekspansi dalam satu dokumen) dengan *threshold*. Nilai yang didapatkan kemudian menjadi batas indeks kata ekspansi yang akan ditambahkan kedalam dokumen asli. Pada Gambar 2 grafik menunjukkan bahwa dari *threshold* 0,1 ke *threshold* 0,2 mengalami penurunan tingkat akurasi sebesar 5,25%. Hal ini disebabkan saat penambahan kata ekspansi, kata yang ditambahkan pada *threshold* 0,2 memberikan kata ekspansi yang tidak cocok, sehingga akurasi klasifikasi menurun. Contoh dari penambahan *term* yang tidak cocok dengan kelas dituliskan dalam Tabel 5.

Tabel 5 Contoh Penambahan *Term* Tidak Sesuai

Threshold	Dokumen + <i>Query Expansion</i>
0,1	warga,paris,tolak,isis,cafe,nasional,subyek
0,2	warga,paris,tolak,isis,cafe,nasional,subyek,warga,negara,aktif,genus,tanam,putus

Penambahan kata ekspansi saat *threshold* 0,2 memberikan tambahan kata “warga” yang pada *threshold* 0,1 tidak ditambahkan, kata “warga” tersebut terdapat pada data latih kelas *travel*. Sehingga saat perhitungan *likelihood*, nilai probabilitas *term* “warga” meningkat pada

kelas *travel*. Penurunan akurasi terjadi ketika sebuah *term* ditambahkan dan *term* tersebut meningkatkan nilai *posterior* dari kelas lain. Sehingga dokumen dikelompokkan ke dalam kelas yang salah.

### 3.2.1 Permasalahan *Translate* Dokumen

Faktor selanjutnya yang menjadi penentu hasil klasifikasi pada penelitian ini adalah proses *translate* dokumen. Pada penelitian ini digunakan data berupa *tweet* dengan bahasa indonesia. Sedangkan sumber yang digunakan untuk melakukan *query expansion* untuk mendapatkan kata ekspansi merupakan WordNet berbahasa inggris. Pada prosesnya terdapat beberapa kesalahan translasi yang menyebabkan proses klasifikasi kurang maksimal. Beberapa kesalahan translasi dituliskan dalam Tabel 6.

Tabel 6 Kesalahan proses translasi

No	Dokumen Asli	Dokumen setelah <i>translate</i>
1	rapat,tutup,jam,dpr,bahas,rahasia,strategi,kelola,devisa	temu,tutup,jam, dpr,bahas, rahasia,strategi,kelola,tukar
2	city,intip,peluang,arsenal	kota, intip, peluang,arsenal
3	kejut,ponsel,selfie,oppo	syok, telepon, selfie,oppo

Perubahan *term* ini memengaruhi proses klasifikasi dikarenakan *term* yang semula ada pada data uji dan data latih digantikan *term* lain hasil terjemahan yang belum tentu ada pada data latih. Hal ini dapat menyebabkan proses klasifikasi dokumen menjadi kurang tepat.

### 3.2.2 Permasalahan Penambahan *Query*

Hasil dari penambahan yang dilakukan memengaruhi proses klasifikasi yang dilakukan, salah satu dampak yang dihasilkan dari penambahan kata ekspansi adalah meningkatnya akurasi dikarenakan kata ekspansi yang ditambahkan sesuai dan ada pada data latih. Namun pada penelitian ini didapatkan beberapa kata ekspansi yang tidak sesuai dengan dokumen awal dalam hal konteks. Berikut contoh kata ekspansi yang ditambahkan, namun tidak sesuai konteks. Dokumen dan kata ekspansi akan dituliskan dalam Tabel 7.

Tabel 7 Kesalahan Penambahan Term

Dokumen Asli	kata Ekspansi
hijau,daun, rilis,album, negara,asia	hijau,daun,lepas,album,negara, asia,warna,kromatik,warna, kromatik,warna, spektral,warna,spe ktral,hijau,biru,biru,hijau, teal,sistem,bidang,tanah,potong,tan ah,bungkus,tanah,parcel, taman,hibur,pasar,dgn,atraksi, tanah,senang

Kata-kata tersebut memang merupakan hipernim dan hiponim dari *term* awal, namun penambahannya sama sekali tidak memperhatikan konteks. Penambahan yang dilakukan justru menyebabkan hasil klasifikasi salah

### 3.3 Pengujian Jenis Query yang Ditambahkan

Pada pengujian kali ini dilakukan proses *query expansion* dengan menambahkan hipernim atau hiponim saja. Hal ini dilakukan untuk mengetahui pengaruh dari jenis kata ekspansi yang ditambahkan

#### 3.3.1 Penambahan Hiponim

Pada bagian ini dilakukan *query expansion* dengan menambahkan kata ekspansi yang merupakan hiponim dari *term* yang ada pada dokumen uji. Berikut adalah Tabel 8 yang berisi perhitungan akurasi.

Tabel 8 Akurasi penambahan hiponim

K	Akurasi (%)
1	76,25
2	77,5
3	51,81
4	70
5	56,25
Sistem	66,36

Hasil yang didapatkan dari proses klasifikasi dengan penambahan hiponim sebagai kata ekspansi memberikan hasil akurasi sebesar 66,36%. Akurasi yang didapatkan pada proses klasifikasi dengan penambahan hiponim saja menunjukkan bahwa adanya penambahan kata ekspansi menurunkan akurasi, namun belum bisa didapatkan akurasi yang optimal dikarenakan beberapa masalah yang telah dibahas sebelumnya.

#### 3.3.2 Penambahan Hipernim

Selanjutnya dilakukan *query expansion* dengan menambahkan kata ekspansi berupa hipernim dari *term* yang ada pada dokumen uji. Berikut adalah Tabel 9 yang berisi perhitungan akurasi menggunakan *k-fold* pada dokumen setelah ditambahkan hipernim dari *term*.

Tabel 9 Akurasi Penambahan Hipernim

K	Akurasi (%)
1	70
2	81,25
3	67,5
4	61,25
5	57,5
Sistem	67,5

Hasil akurasi yang diperoleh saat penambahan hipernim sebagai kata ekspansi memiliki akurasi 67,5%. Hal ini lebih tinggi dibandingkan daripada akurasi penambahan hiponim saja dan penambahan keduanya sekaligus. Akurasi dari penambahan hipernim lebih tinggi dibandingkan dengan hiponim dikarenakan kebanyakan dari *term* hanya memiliki sedikit atau bahkan tidak memiliki hiponim

## 4. KESIMPULAN

Berdasarkan hasil, analisis serta pembahasan dari penelitian implementasi *Naïve Bayes* dan *query expansion* berbasis WordNet untuk klasifikasi *tweet* berita pada Twitter maka dapat disimpulkan bahwa penambahan *query* pada dokumen sebelum dilakukan proses klasifikasi menurunkan tingkat akurasi. Pada proses klasifikasi tanpa menggunakan *query expansion* didapatkan akurasi sebesar 72%. Penambahan *query* berupa hipernim dan hiponim menghasilkan akurasi paling tinggi sebesar 65,75%, penambahan hipernim saja menghasilkan akurasi 67,5%, dan penambahan hiponim saja sebesar 66,3% semua pada *threshold* 0,1. Penurunan akurasi terjadi karena beberapa hal seperti hasil terjemahan dari *google translate API* mengubah *term* awal menjadi *term* lain yang tidak ada pada data latih dan penambahan *query* sama sekali tidak memperhatikan konteks dokumen.

Untuk penggunaan WordNet berbahasa inggris kurang efektif saat *dataset* berbahasa indonesia karena perlu melakukan proses

*translate*, maka disarankan untuk menggunakan WordNet berbahsa indonesia. Untuk menghasilkan kata ekspansi yang lebih beragam dan akurat maka disarankan untuk encoba menggunakan relasi antar kata selain hiponim dan hipernim.

## 5. DAFTAR PUSTAKA

- Bagaskoro, G. N. (2017). Penerapan Klasifikasi Tweets pada Berita Twitter Menggunakan Metode k-nearest neighbor dan Query Expansion Berbasis Distributional Semantic. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*(10), 3849-3855.
- Buzic, D., & Dobša, J. (2018). Lyrics Classification Using Naive Bayes., (pp. Conference: MIPRO 2018, 41st International Convention on Information and Communication Tachnology, Elecronics and Microelectornics). croatia.
- Li, W., Ganguly, D., & J.F.Jones, G. (2016). Using WordNet for Query Expansion: ADAPT @ FIRE 2016 Microblog Track.
- McDonald, D., & Kelly, U. (2012). *The Value and Benefits of Text Mining*. United Kingdom.
- Perdana, R. S., Suprpto, & Regasari, R. (2013). Pengkategorian Pesan Singkat Berbahasa Indonesia Pada Jejaring Sosial Twitter Dengan Metode Klasifikasi Naïve Bayes. *SI. Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya*.
- Rahman, A., Wiranto, & Doewes, A. (2017). Online News Classification Using Multinomial Naive Bayes. *Jurnal Ilmiah Teknologi dan Informasi*, 32-38.
- Sriram, B. (2010). *Short Text Classification In Twitter To Improve Information Filtering*. ohio: The Ohio State University.
- Taheri, S., & Mammadov, M. (2013). Learning The Naive Bayes Classifier With Optimization Models. *International Journal of Applied Mathematics and Computer Science*, 23(4), 787-795.
- Tanjung, J. P., Fauzi, M. A., & Indriati. (2017). Klasifikasi Tweets Pada Twitter Dengan Menggunakan Metode Fuzzy K-Nearest (Fuzzy K-NN) dan Query Expansion Berbasis Apriori. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1(5), 405-414.
- Vijayaran, S., Ilamathi, J., & Nithya. (2015). Preprocessing Techniques for Text Mining - An Overview . *International Journal of Computer Science & Communication Networks*, 5(1), 7-16.
- Zhang, J., Deng, B., & Li, X. (2009). Concept Based Query Expansion Using WordNet. Beijing: Dept. Electronic Engineering, Tsinghua Univ. Beijing, 100084, China.