

Perbandingan Algoritma Selection Feature Query Expansion Ranking dan Symmetrical Uncertainty Pada Instagram Terkait Sengketa ZEE di Perairan Natuna

Dzurrifqi Aziz^{*1}, Nur Hayatin², Christian Sri Kusuma Aditya³

^{1,2,3}Universitas Muhammadiyah Malang

rifqiaziz874@gmail.com^{*1}, noorhayatin@gmail.com², christianskaditya@umm.ac.id³

Abstrak

Algoritma Naïve Bayes adalah untuk menggabungkan probabilitas kata-kata dan kategori untuk memperkirakan probabilitas kategorikan dokumen. Naïve Bayes ini merupakan algoritma klasifikasi yang sering digunakan dalam sebuah penelitian salah satunya penelitian mengenai Sentimental Analysis. Dengan demikian, penelitian ini mencoba menganalisis ulasan dari masyarakat dengan metode naive bayes classifier dan membandingkan penggunaan seleksi fitur Symmetrical Uncertainty dan query expansion ranking untuk mengoptimalkan kinerja klasifikasi pada penelitian. Hasil pengujian yang diperoleh dengan menggunakan symmetrical uncertainty adalah tingkat accuracy sebesar 76%, precision sebesar 76%, dan recall sebesar 61%. Sedangkan hasil pengujian yang diperoleh dengan menggunakan query expansion ranking adalah tingkat accuracy sebesar 75%, precision sebesar 75%, dan recall sebesar 56%.

Kata Kunci: Analisis Sentimen, Naïve Bayes, Seleksi Fitur, Python

Abstract

Naïve Bayes algorithm is for combining the probability of words and categories to estimate the probability of document categorization. Naïve Bayes is a classification algorithm that is often used in a study, one of which is research on Sentimental Analysis. Thus, this study tries to analyze reviews from the public using the Naïve Bayes Classifier method and compares the use of Symmetrical Uncertainty feature selection and query expansion ranking to optimize classification performance in the study. The test results obtained using symmetrical uncertainty are an accuracy level of 76%, a precision of 76%, and a recall of 61%. While the test results obtained using the query expansion ranking are 75% accuracy level, 75% precision, and 56% recall.

Keywords: Sentiment analysis, Naïve Bayes, Feature Selection, Python

1. Pendahuluan

Algoritma Naïve Bayes adalah Ide dasarnya adalah untuk menggabungkan probabilitas kata-kata dan kategori untuk memperkirakan probabilitas kategori dokumen tersebut. Naïve Bayes ini merupakan algoritma klasifikasi yang sering digunakan dalam sebuah penelitian salah satunya penelitian mengenai Sentimental Analysis of Tweets Using Naïve Bayes Algorithm yang menggunakan klasifikasi Naïve Bayes Ditemukan bahwa jika jumlah fitur meningkat, pembelajaran sentimen Enhanced menggunakan hastag twitter dari fitur yang dipilih juga akan mengalami meningkat. [1]. dengan adanya *preprocessing* atau biasa disebut praproses pada data yang akan diolah sehingga dapat mempertinggi hasil dari data yang diolah.

Feature Selection adalah teknik praproses yang ditemukan untuk proses mengidentifikasi dan menghapus informasi yang tidak penting dan berlebihan sebanyak mungkin. subset minim dari fitur yang dapat menangkap penting properti dataset untuk memungkinkan klasifikasi yang memadai. Data yang sering muncul, dapat berisi tingkat tinggi informasi yang tidak relevan dan berlebihan yang sangat menurunkan kinerja sebuah algoritma. Oleh sebab itu, pemilihan fitur menjadi perlu untuk tugas pembelajaran untuk menghadapi banyaknya data yang sering muncul. Tujuan utama pemilihan fitur Algoritma adalah memilih subset dari fitur yang ada independen satu sama lain dan cukup relevan untuk proses pembelajaran. Feature selection memiliki banyak metode seperti *information gain*, *chi square*, *Query Expansion Ranking (QER)*, *symmetrical uncertainty* dsb. Penelitian sebelumnya tentang sentiment analisis pariwisata di kota Malang, yang memiliki tujuan untuk mengetahui opini masyarakat tentang pariwisata yang ada di kota Malang. Penelitian ini menggunakan *feature selection Query Expansion Ranking*

(*QER*) dengan metode *Naïve Bayes Classifier*. Hasil dari penelitian ini diketahui akurasi terbaik sebesar 86% (fanissa 2018). Dalam penelitian lainnya tentang analisis sentimen kurikulum 2013 pada sosial media twitter yang menggunakan metode K-NN dan *Feature Selection Query Expansion Ranking* sebelum melakukan perhitungan feature selection dilakukannya pembobotan TF-IDF hal dilakukan agar data lebih efektif mereka mendapatkan hasil akurasi yaitu sebesar 96,36% [2].

penelitian yang menggunakan *feature selection* dengan metode *Query Expansion Ranking* menghasilkan nilai akurasi yang cukup tinggi sehingga penelitian ini dilakukan untuk mengetahui perbandingan antara *feature selection* yang jarang digunakan seperti *symmetrical uncertainty*. *symmetrical uncertainty* merupakan turunan dari algoritma *information gain*, sehingga banyak yang belum mengetahui tentang *feature selection* ini. Sehingga penelitian ini membandingkan metode *feature selection* yang jarang digunakan dengan penelitian yang diatas untuk mencari nilai akurasi tertinggi antara kedua *feature selection*.

Pada penelitian sebelumnya mengenai fitur seleksi menggunakan *symmetrical uncertainty* pada cacat perangkat lunak dalam penelitian ini menggunakan metode klasifikasi *Naïve Bayes*. *symmetrical uncertainty* ini digunakan untuk merangking kata yang penting dari hasil pengurangan fitur dan melakukan klasifikasi *Naïve Bayes*. Hasil dari penelitian ini mendapatkan akurasi sebesar 91% [3].

Penelitian lain yang dilakukan oleh Delki Abadi tentang perbandingan algoritme *feature selection information gain* dan *symmetrical uncertainty* pada data ketahanan pangan yang menggunakan algoritma *Decision tree* dan *pruning* dilakukan menjadi 2 perhitungan yaitu *Decision tree* pertama menggunakan algoritme seleksi fitur *information gain*, yang kedua *decision tree* menggunakan algoritme seleksi fitur *symmetrical uncertainty*. Nilai perhitungan akurasi yang pertama memiliki akurasi 52% sedangkan perhitungan yang kedua memiliki nilai akurasi 49% [4].

Penelitian ini mencoba untuk membandingkan tingkat akurasi model pemilihan fitur *Query Expnasion Ranking (QER)* dan *information gain* dengan menggunakan *symmetrical uncertainty* untuk pemilihan fitur. Untuk proses klasifikasi menggunakan algoritma *Naïve Bayes*. Dari penelitian ini diharapkan algoritme *feature selection* yang paling baik digunakan untuk membangun model pemilihan fitur pada suatu data dapat diketahui.

2. Metode Penelitian

2.1 Pengumpulan Data

Dengan cara melakukan pengamatan langsung di kolom instagram platfrom tersebut obyek yang diteliti atau dapat dirumuskan sebagai proses pencatatan pola perilaku subyek (orang), obyek (benda) atau kejadian tanpa adanya pertanyaan atau komunikasi dengan individu-individu yang diteliti. Metode ini dilakukan tanpa perlu memberikan pertanyaan kepada responden pada tanggal 14 Februari 2020 – 24 Juli 2020.

2.2 Preprocessing

Tahap Preprocessing adalah tahapan awal dalam sebelum memproses sebuah data berisi teks Pada preprocessing, data diproses untuk menghilangkan kata yang *noise*. Ada beberapa tahapannya yaitu case folding, tokenizing, dan Filteing.

2.2.1 Case Folding

Preprocessing dokumen terdiri dari *case folding*, yaitu semua huruf dalam dokumen diubah menjadi huruf kecil dan karakter lain kecuali huruf dihilangkan [5]. *Case folding* merupakan proses mengubah sebuah teks yang awalnya huruf besar menjadi huruf kecil sehingga pada keseluruhan data yang akan diproses menjadi huruf kecil.

2.2.2 Tokenizing

Tokenizing merupakan merupakan proses penguraian deskripsi yang awalnya muncul dalam bentuk kalimat yang berisi kata-kata dan pemisah antar kata, seperti titik (.), Koma (,), spasi, dan bentuk pemisahan lainnya sehingga menjadi kata-kata saja baik itu kata-kata penting dan tidak penting. [6]

Tokenizing adalah tugas memotong urutan karakter dan sekumpulan dokumen tertentu menjadi kata atau karakter sesuai dengan persyaratan sistem. Potongan ini disebut token [7].

2.2.3 Filtering

Pada tahap *filtering*, *stop words* digunakan untuk menyaring kata-kata yang dihasilkan pada proses *tokenizing* guna mendapatkan kata-kata penting untuk proses selanjutnya [8] *Stop Word* dapat diartikan sebagai menghapus karakter, tanda baca dan kata umum yang tidak memiliki arti atau informasi yang diinginkan. *Stop word* sering digunakan untuk mengambil informasi, seperti mesin pencari *Google*. Dengan menghilangkan beberapa kata kerja, kata sifat, dan kata keterangan lainnya [9].

2.3 Term Frequency Inverse Document Frequency

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah cara pemberian bobot hubungan suatu kata (*term*) terhadap dokumen atau juga dapat diartikan sebagai sebuah ukuran statistik yang digunakan untuk mengukur seberapa penting sebuah kata di dalam sebuah dokumen [10]. Untuk menghitung TF – IDF dari kata, Persamaan 1.

$$TF - IDF(t, d, D) = tf(t, d) \times idf(t, D) \quad (1)$$

2.4 Feature Selection

Feature selection yang akan digunakan untuk mengurangi data yang kurang relevan dan mengoptimalkan kinerja klasifikasi pada penelitian. Terdapat dua seleksi fitur, yaitu seleksi fitur *Chi Square* dan seleksi fitur *Query Expansion Ranking*.

2.4.1 Symmetrical Uncertainty

Symmetrical Uncertainty merupakan turunan dari *Information Gain*. Sering digunakan untuk menghilangkan bias pada *information gain* [4]. Sebelum menghitung *Symmetrical Uncertainty* terlebih dahulu menghitung *Information Gain* dan *entropy*, ditunjukkan pada Persamaan 2, Persamaan 3, dan Persamaan 4.

$$Entropy(s) = - \sum \frac{|S_i|}{s} \log \frac{|S_i|}{s} \quad (2)$$

Rumus *Information Gain*

$$InfoGain(S, A) = Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{s} Entropy(S_v) \quad (3)$$

Rumus *Symmetrical Uncertainty* (Abadi 2013)

$$SU(A) = \frac{Gain(A)}{\inf o(S) + \inf o(S_v)} \quad (4)$$

2.4.2 Query Expansion Ranking (QER)

Query Expansion Ranking (QER) merupakan salah satu metode seleksi fitur yang digunakan untuk mengurangi kompleksitas komputasi dalam analisis sentimen dan memberikan skor pada setiap kata [11].s Persamaan 5 *Query Expansion Ranking (QER)* untuk menghitung peluang fitur *f* pada dokumen kelas positif.

$$p_f = \frac{df_+^f + 0,5}{n^+ + 1,0} \quad (5)$$

Rumus Persamaan 6 untuk menghitung peluang fitur *f* pada dokumen kelas negatif.

$$qf = \frac{df_-^f + 0,5}{n^- + 0,5} \quad (6)$$

Setelah mendapatkan jumlah dokumen positif dan dokumen negatif, selanjutnya akan menghitung nilai dari *Query Expansion Ranking* pada Persamaan 7.

$$Score_f = \frac{|p_f + q_f|}{|p_f - q_f|} \quad (7)$$

2.5 Klasifikasi Naïve Bayes

Setelah melakukan perhitungan *feature selection*, kemudian akan dilakukan sebuah klasifikasi dengan metode *Naïve Bayes* terhadap data yang sudah diklasifikasikan. *Naïve Bayes* merupakan salah satu metoda pembelajaran mesin yang memanfaatkan perhitungan *probabilitas* dan *statistik*. *Naïve Bayes* adalah teknik prediksi probabilistik sederhana yang didasarkan pada penerapan teorema atau aturan *Bayes*. Teknik ini mengasumsikan independensi yang kuat terhadap fitur, yang berarti bahwa fitur dalam data tidak bergantung pada apakah ada fitur lain dalam data yang sama [12].

Pada proses klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan. Berikut Persamaan 8 yang digunakan dalam proses klasifikasi.

$$P(c|d) = P(c) \prod_{i=1}^n P(w_i|c) \quad (8)$$

2.6 Pengujian

Pengujian dilakukan dengan membagi data latih dan data uji pada tahap klasifikasi di setiap skenario pengujian. Setelah itu akan dilakukan penghapusan fitur pada setiap seleksi fitur dengan menggunakan pembagian data latih dan data uji yang mempunyai akurasi terbaik. Tabel 1 untuk pengukuran akurasi diukur dengan *confusion matrix* dengan Persamaan 9, Persamaan 10, dan Persamaan 11.

Tabel 1. *Confussion Matrix*

	Positif	Negatif
Positif	True Positif	False Positif
Negatif	False Negatif	True Negatif

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

$$Presisi = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

3. Hasil Penelitian dan Pembahasan

Pada bagian ini akan dijelaskan bagaimana proses pengujian berupa perhitungan untuk mengetahui tingkat *accuracy*, *precision*, dan *recall* pada masing-masing metode. Skenario pengujian pada penelitian ini adalah sebanyak tiga. Skenario pertama adalah pengujian dengan hasil seleksi fitur *symmetrical uncertainty*, skenario kedua adalah pengujian dengan hasil seleksi fitur menggunakan *query expansion ranking*, dan skenario yang ketiga adalah grafik hasil perbandingan dari kedua metode. Proses uji coba akan dilakukan dengan beberapa persentase rasio jumlah fitur yang digunakan antara lain 25%, 50%, 75%, dan 100% dari jumlah fitur yang akan diuji yaitu 100 fitur dari keseluruhan fitur yang berjumlah 1522 fitur. Data fitur diseleksi karena 100 fitur tersebut memiliki nilai terbesar diantara fitur-fitur yang lainnya. Data jumlah fitur yang digunakan dengan rasio jumlah fitur sebanyak 25% adalah 25 data. Data jumlah fitur yang digunakan dengan rasio jumlah fitur sebanyak 50% adalah 50 data. Data jumlah fitur yang digunakan dengan rasio jumlah fitur sebanyak 75% adalah 75 data. Data jumlah fitur yang digunakan dengan rasio jumlah fitur sebanyak 100% adalah 100 data.

3.1 Klasifikasi Naïve Bayes Classifier dan Menggunakan Feature Symmetrical Uncertainty

Pada Tabel 2, skenario pertama adalah pengujian dengan hasil seleksi fitur *symmetrical uncertainty*, proses uji coba akan dilakukan dengan beberapa persentase rasio jumlah fitur yang digunakan antara lain 25%, 50%, 75%, dan 100% dari jumlah fitur yang akan diuji yaitu 100 fitur.

Tabel 2. Hasil Perhitungan Nilai *Accuracy*, *Precision*, dan *Recall Symmetrical Uncertainty*

Rasio Fitur	Accuracy	Precision	Recall
25%	81%	81%	65%
50%	79%	78%	63%
75%	75%	75%	60%
100%	70%	71%	57%
Rata-Rata	76%	76%	61%

3.2 Klasifikasi Naïve Bayes Classifier dan Fitur Rasio Menggunakan Feature Selection Query Expansion Ranking

Pada Tabel 3, skenario kedua adalah pengujian dengan hasil seleksi fitur menggunakan *query expansion ranking*, proses uji coba akan dilakukan dengan beberapa persentase rasio jumlah fitur yang digunakan antara lain 25%, 50%, 75%, dan 100% dari jumlah fitur yang akan diuji yaitu 100 fitur skenario yang ketiga adalah grafik hasil perbandingan dari kedua metode.

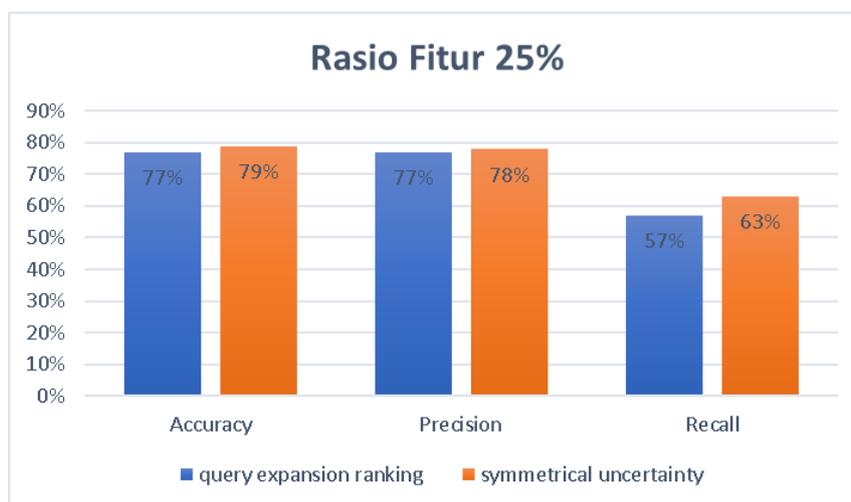
Tabel 3. Hasil Perhitungan Nilai *Accuracy*, *Precision*, dan *Query Expansion Ranking*

Rasio Fitur	Accuracy	Precision	Recall
25%	81%	81%	61%
50%	77%	77%	57%
75%	73%	73%	54%
100%	69%	69%	53%
Rata-Rata	75%	75%	56%

3.3 Perbandingan Seleksi Fitur

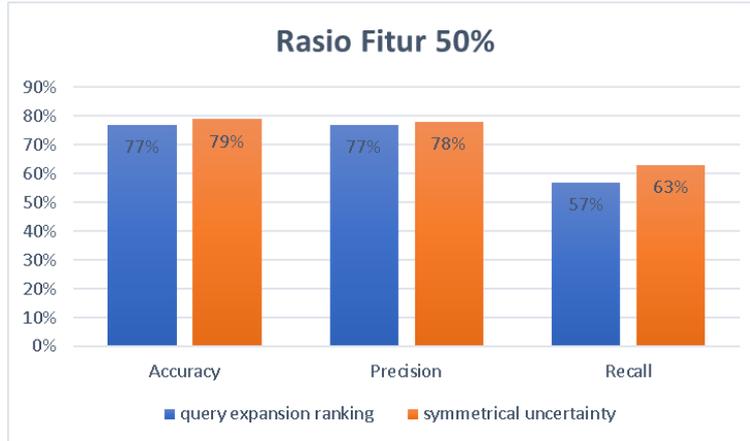
Pada tahap ini merupakan tahap melakukan perbandingan hasil klasifikasi dari penggunaan metode *symmetrical uncertainty* dan metode *query expansion ranking*. Untuk visualisasi tersebut digunakannya *chart* menggunakan Microsoft excel. Perbandingan hasil pengujian tersebut dibedakan dengan beberapa rasio fitur yaitu 25%, 50%, 75%, dan 100% data fitur.

Hasil perbandingan seleksi fitur dengan nilai *accuracy*, *precision*, dan *recall* antara metode *symmetrical uncertainty* dan metode *query expansion ranking* pada masing-masing rasio fitur dalam penelitian ini disajikan seperti pada Gambar 1.



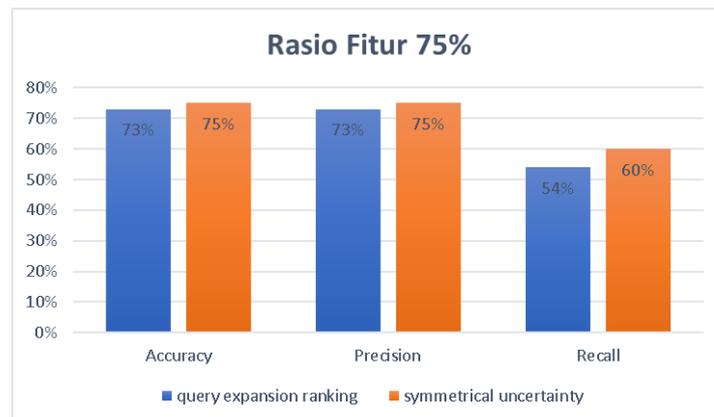
Gambar 1. Hasil Perbandingan Seleksi Fitur dengan Rasio Fitur 25%

Pada rasio fitur sebesar 25% atau dalam artian jumlah fitur yang diuji adalah sebesar 25 fitur dihasilkan tingkat *accuracy*, *precision*, dan *recall* dari metode *symmetrical uncertainty* yang lebih tinggi daripada metode *query expansion ranking* yaitu *accuracy* sebesar 79%, *precision* sebesar 78%, dan *recall* sebesar 63%.



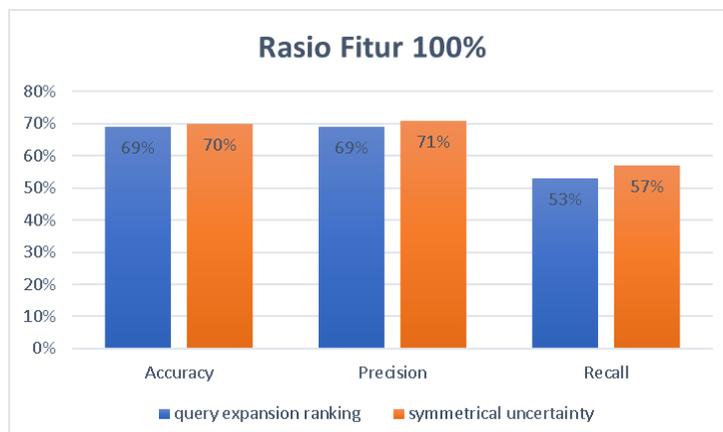
Gambar 2. Hasil Perbandingan Seleksi Fitur dengan Rasio Fitur 50%

Pada Gambar 2, rasio fitur sebesar 50% atau dalam artian jumlah fitur yang diuji adalah sebesar 50 fitur dihasilkan tingkat *accuracy*, *precision*, dan *recall* dari metode *symmetrical uncertainty* yang lebih tinggi daripada metode *query expansion ranking* yaitu *accuracy* sebesar 79%, *precision* sebesar 78%, dan *recall* sebesar 63%.



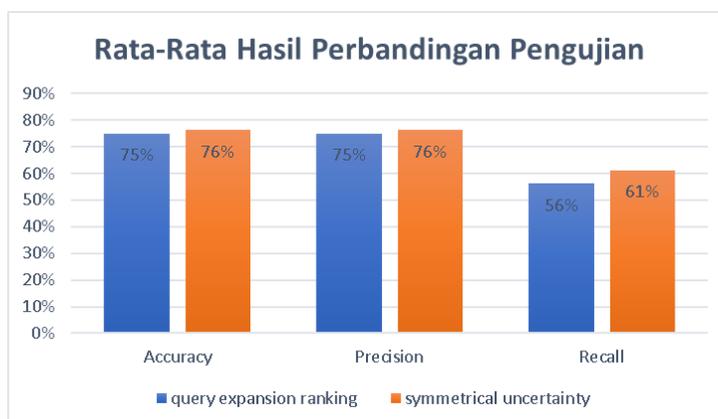
Gambar 3. Hasil Perbandingan Seleksi Fitur dengan Rasio Fitur 75%

Pada Gambar 3, rasio fitur sebesar 75% atau dalam artian jumlah fitur yang diuji adalah sebesar 75 fitur dihasilkan tingkat *accuracy*, *precision*, dan *recall* dari metode *symmetrical uncertainty* yang lebih tinggi daripada metode *query expansion ranking* yaitu *accuracy* sebesar 75%, *precision* sebesar 75%, dan *recall* sebesar 60%.



Gambar 4. Hasil Perbandingan Seleksi Fitur dengan Rasio Fitur 100%

Pada Gambar 4, rasio fitur sebesar 100% atau dalam artian jumlah fitur yang diuji adalah sebesar 100 fitur dihasilkan tingkat *accuracy*, *precision*, dan *recall* dari metode *symmetrical uncertainty* yang lebih tinggi daripada metode *query expansion ranking* yaitu *accuracy* sebesar 70%, *precision* sebesar 71%, dan *recall* sebesar 57%.



Gambar 5. Rata-Rata Hasil Perbandingan Seleksi Fitur

Pada Gambar 5, hasil pengujian yang diperoleh dengan menggunakan *symmetrical uncertainty* adalah tingkat *accuracy* sebesar 76%, *precision* sebesar 76%, dan *recall* sebesar 61%. Sedangkan hasil pengujian yang diperoleh dengan menggunakan *query expansion ranking* adalah tingkat *accuracy* sebesar 75%, *precision* sebesar 75%, dan *recall* sebesar 56%.

Dari keempat kategori perbandingan berdasarkan rasio fitur yang berbeda, terlihat bahwa seleksi fitur menggunakan *symmetrical uncertainty* bernilai lebih tinggi daripada seleksi fitur menggunakan *query expansion ranking*. Hal ini dikarenakan pada seleksi fitur menggunakan *symmetrical uncertainty* terdapat proses perhitungan *entropy* terlebih dahulu sehingga hasil pengujiannya lebih tinggi, sedangkan di seleksi fitur *query expansion ranking* tidak ada. Perhitungan *entropy* sangat berpengaruh terhadap kesuksesan seleksi fitur yang dilakukan oleh metode *symmetrical uncertainty*.

4. Kesimpulan

Berdasarkan pada penelitian yang dilakukan, klasifikasi *naïve bayes classifier* dengan menggunakan *symmetrical uncertainty* adalah tingkat *accuracy* sebesar 76%, *precision* sebesar 76%, dan *recall* sebesar 61%. Sedangkan hasil pengujian yang diperoleh dengan menggunakan *query expansion ranking* adalah tingkat *accuracy* sebesar 75%, *precision* sebesar 75%, dan *recall* sebesar 56%.

Seleksi fitur menggunakan *symmetrical uncertainty* bernilai lebih tinggi daripada seleksi fitur menggunakan *query expansion ranking* karena pada seleksi fitur menggunakan *symmetrical uncertainty* terdapat proses perhitungan *entropy* terlebih dahulu sehingga hasil pengujiannya lebih tinggi, sedangkan di seleksi fitur *query expansion ranking* tidak ada. Perhitungan *entropy* sangat berpengaruh terhadap kesuksesan seleksi fitur yang dilakukan oleh metode *symmetrical uncertainty*.

Referensi

- [1] Vadivukarassi, M., Puviarasan, N., & Aruna, P. (2017). Sentimental analysis of tweets using Naive Bayes algorithm. *World Applied Sciences Journal*, 35(1), 54-59.
- [2] Mentari, N. D., Fauzi, M. A., & Muflikhah, L. (2018). Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN*, 2548, 964X.
- [3] Roihan, A. (2018). *Seleksi fitur menggunakan Symmetrical Uncertainty pada prediksi cacat perangkat lunak* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- [4] Abadi, D. (2013). Perbandingan Algoritme Feature Selection Information Gain Dan Symmetrical Uncertainty Pada Data Ketahanan Pangan.

-
- [5] Yudiarta, N. G., Sudarma, M., & Ariastina, W. G. (2018). Penerapan Metode Clustering Text Mining Untuk Pengelompokan Berita Pada Unstructured Textual Data. *Majalah Ilmiah Teknologi Elektro*, 17(3), 339-344.
- [6] Melindah, M. (2020). *Pembangunan Aplikasi Chatbot Sparepart Handphone Sebagai Media Informasi Dan Promosi Di Beyoungkeyrocks Workshop Berbasis Website* (Doctoral dissertation, Universitas Komputer Indonesia).
- [7] Schütze, H., Manning, C. D., & Raghavan, P. (2008). *Introduction to information retrieval* (Vol. 39, pp. 1041-4347). Cambridge: Cambridge University Press.
- [8] Rahmad, A. N., & Pribadi, F. S. (2015). Pemilihan Feature Dengan Chi Square Dalam Algoritma Naïve Bayes Untuk Klasifikasi Berita. *Edu Komputika Journal*, 2(1).
- [9] Ling, J., Kencana, I. P. E. N., & Oka, T. B. (2014). Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, 3(3), 92-99.
- [10] Melita, R., Amrizal, V., Suseno, H. B., & Dirjam, T. (2018). Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Syarah Umdatil Ahkam). *Jurnal Teknik Informatika*, 11(2), 149-164.
- [11] Fanissa, S., Fauzi, M. A., & Adinugroho, S. (2018). Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN, 2548, 964X*.
- [12] Nugroho, Y. S. (2014). Data Mining Menggunakan Algoritma Naive Bayes Untuk Klasifikasi Kelulusan Mahasiswa Universitas Dian Nuswantoro. *Dian Nuswantoro Fakultas Ilmu Komputer Skripsi*.