Jurnal JUPITER, Vol.14, No. 2, Bulan Oktober Tahun 2 0 2 2 , Hal. 44-52

Klasifikasi Pertanyaan Berbahasa Indonesia Menggunakan Algoritma *Support Vector Machine* dan Seleksi Fitur *Mutual Information*

Syechky Al Qodrin[1)], Novi Yusliani[2)], Alvi Syahrini[3)]
[*1,2,3], Program Studi Teknik Informatika, Universitas Sriwijaya,
Jl. Palembang – Prabumulih Km.32 Indralaya, Ogan Ilir, Sumatera Selatan
e-mail: [*1]syechkyal@gmail.com, [2]novi_yusliani@unsri.ac.id, [3]alvisyahrini@ilkom.unsri.ac.id

Abstrak

Klasifikasi teks dapat digunakan untuk menyusun, mengatur dan melakukan kategori terhadap suatu teks. Klasifikasi teks dapat digunakan untuk semua dokumen teks meskipun suatu teks tersebut memiliki jumlah fitur yang banyak. Namun, banyaknya jumlah fitur dapat menyebabkan bekurangnya akurasi pada hasil kinerja sistem klasifikasi dikarenakan terdapat sebagian fitur yang memiliki relevansi yang kurang terhadap suatu kategori teks. Metode seleksi fitur Mutual Information yang dikombinasikan dengan algoritma Support Vector Machine (SVM) digunakan untuk meningkatkan hasil kinerja pada proses klasifikasi terhadap dokumen pertanyaan berbahasa indonesia dengan cara menghilangkan fitur dengan bobot dibawah nilai ambang batas. Hasil penelitian menunjukkan bahwa dengan penggunaan metode seleksi fitur Mutual Information pada algoritma klasifikasi SVM mampu menghasilkan kinerja terbaik dengan nilai accuracy sebesar 0.92 , precision : 0.93, recall : 0.89, f-measure : 0.9, waktu komputasi : 7 s dan jumlah fitur : 240.

Kata kunci*—Klasifikasi Teks, Seleksi Fitur, Support Vector Machine, Mutual Information*

Abstract

Text classification can be used to organize, arrange and categorize a text. Text classification can be used for all text documents even if a text has a large number of features. However, the large number of features can cause reduced accuracy in the performance results of the classification system because there are some features that have less relevance to a text category. The Mutual Information feature selection method combined with the Support Vector Machine (SVM) algorithm is used to improve performance results in the classification process for Indonesian question documents by eliminating features with weights below the threshold. The results showed that the use of the Mutual Information feature selection method on the SVM classification algorithm was able to produce the best performance with an accuracy value of 0.92, precision: 0.93, recall: 0.89, f-measure: 0.9, computation time: 7 s and number of features: 240.

Keywords*— Text Classification, Feature Selection, Support Vector Machine, Mutual Information*

1. PENDAHULUAN

Kalimat tanya ialah susunan kata yang digunakan untuk mendapatkan jawaban atau informasi terkait suatu hal. Untuk mendapatkan suatu jawaban yang tepat, maka diperlukan analisa untuk memahami suatu konteks pertanyaan. Pertanyaan berbahasa

Indonesia dapat dikategorikan menjadi 3 kategori berdasarkan analisis dari jenis jawaban yang diberikan. Pertanyaan *factoid* ditujukkan untuk pertanyaan dengan jawaban berupa fakta yang singkat. Pertanyaan *non-factoid* ditujukkan untuk jawaban yang disertai penjelasan terkait suatu hal. Pertanyaan *others* ditujukkan untuk jenis pertanyaan diluar pertanyaan *factoid* dan *non-factoid* [1].

Teknik klasifikasi teks dapat digunakan untuk menganalisa dan menentukan kategori dari dokumen berupa teks pertanyaan berbahasa Indonesia. Salah satu teknik klasifikasi teks yang cukup banyak digunakan ialah menggunakan algoritma *Support Vector Machine* (SVM). Namun, teknik klasifikasi teks memiliki permasalahan dalam pemilihan fitur, dimana semakin banyaknya fitur yang kurang relevan pada suatu teks dapat mengakibatkan berkurangnya akurasi dan waktu komputasi yang lebih lama. Penggunaan metode seleksi fitur digunakan untuk menghilangkan fitur yang memiliki tingkat relevansi yang rendah.

Metode seleksi fitur *Mutual Information* merupakan suatu metode perhitungan yang dapat digunakan sebagai metode seleksi fitur dengan cara melakukan perhitungan terhadap jumlah informasi yang terdapat pada suatu term dan juga menghitung kontribusi term tersebut dalam menetukan hasil dari keputusan kelas kata pada proses klasifikasi [2].

Penerapan metode seleksi fitur *Mutual Information* telah dilakukan dibeberapa bidang penelitian, Penggunaan metode seleksi fitur Mutual Information pada klasifikasi multi label hadis bukhari yang dikombinasikan dengan algoritma k-Nearest Neighbor mampu meningkatkan hasil kinerja akurasi dari 91.86% menjadi 93.14% [2]. Penelitian lain, menggunakan metode *Mutual Information* untuk mengklasifikasi pertanyaan berbahasa indonesia menggunakan algoritma *Naïve Bayes Classifier* terjadi penurunan akurasi dari 80% menjadi 70% dikarenakan ada fitur penting yang tereliminasi pada proses seleksi fitur [3]. Dan pada penelitian lain dengan topik yang sama mengkombinasikan metode *Mutual Information* dan algoritma klasifikasi *Support Vector Machine* (SVM) menunjukkan hasil kinerja terbaik sebesar 94.24% [4].

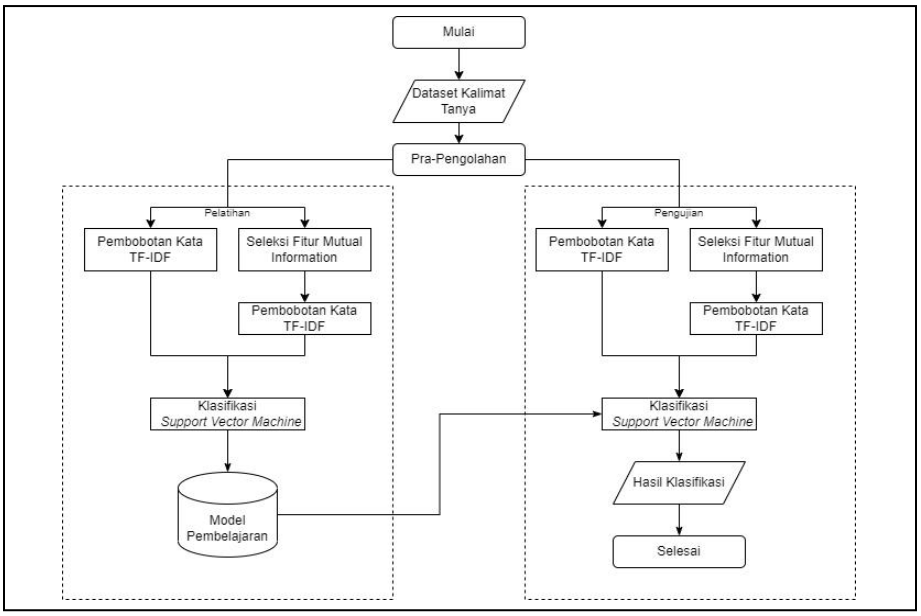
Penelitian ini bertujuan untuk membangun perangkat lunak klasifikasi pertanyaan berbahasa Indonesia menggunakan algoritma *Support Vector Machine*. Selanjutnya, akan dilakukan analisa untuk mengetahui pengaruh dari penggunaan metode seleksi fitur *Mutual Information* terhadap klasifikasi pertanyaan berbahasa indonesia menggunakan algoritma *Support Vector Machine.*

2. METODE PENELITIAN

Terdapat beberapa tahapan yang dilakukan pada penelitian ini. Tahapan dilakukan dengan melakukan pengumpulan data penelitian terlebih dahulu. Kemudian, dilakukan perancangan penelitian. Selanjutnya, dilakukan implementasi terhadap perancangan yang telah dibuat. Terakhir, dilakukan pengujian terhadap implementasi yang telah dilakukan. Tahapan akan dijelaskan secara mendetail pada subbab setelah ini.

2.1. Arsitektur Sistem

Penelitian ini terdiri dari beberapa tahapan arsitektur sistem yang dimulai dari pengumpulan data, pra-pengolahan data*, splitting* data menjadi data pelatihan dan data uji, seleksi fitur menggunakan metode *Mutual Information*, pembobotan pada fitur menggunakan TF-IDF, terakhir pemodelan dan klasifikasi pertanyaan dilakukan menggunakan algoritma *support vector machine*. Aristektur sistem dapat dilihat pada Gambar 1.



Gambar 1. Arsitektur Sistem

Tahapan pengumpulan data dilakukan dengan menggunakan data sekunder berupa kumpulan pertanyaan berbahasa Indonesia dengan kategori *factoid*, *non-factoid* dan *others*. Selanjutnya data yang telah didapatkan diolah menggunakan proses pra-pengolahan data. Pra-pengolahan data merupakan tahapan dalam mengubah teks berupa data yang tidak terstruktur menjadi data terstruktur yang dapat diproses [5]. Proses pra-pengolahan data yang dilakukan pada penelitian ini terdiri dari proses *Case Folding*, *Noise Removal* dan Tokenisasi. Selanjutnya, dilakukan pemisahan data menggunakan metode *cross validation*. Data yang telah dipisah akan dilakukan proses seleksi fitur. Seleksi fitur berfungsi untuk meningkatkan efisiensi dan hasil kinerja dengan melakukan eliminasi terhadap fitur yang tidak relevan [6]. Setelah dilakukan tahapan seleksi fitur dilakukan pembobotan kata menggunakan TF-IDF. TF-IDF merupakan teknik yang memperhatikan kemunculan term pada dokumen yang mewakili suatu kategori, kemudian memberi bobot pada tiap kata [7]. Setelah dilakukan pembobotan data maka akan dilakukan proses klasifikasi data menggunakan algoritma *Support Vector Machine*.

2.2. Support Vector Machine

Support Vector Machine termasuk kedalam metode *supervised learning* yang masih sering digunakan sebagai algoritma klasifikasi. SVM berkerja dengan menemukan *hyperlane* pada *margin* yang terbesar. Semakin besar nilai margin dapat membuat hyperlane yang dicari menjadi lebih baik [8]. SVM mengkategorikan hasil klasifikasi berdasarkan hasil *training* data melalui batas keputusan atau *decision boundary* dengan tujuan untuk menemukan *hyperlane* teroptimal [9]. Perhitungan untuk mencari *hyperlane* teroptimal ditampilkan pada persamaan 1.

$$f(x) = sgn(\sum_{i=1}^{nsv} \alpha_i \, y_i K(x_i, x_d) + b) \quad (1)$$

Fungsi f(x) merupakan fungsi pemisah optimal. Variabel α merupakan nilai *Lagrange Multiplier* data *support vector*. Dan b merupakan nilai bias yang diperlukan untuk mencari nilai *hyperlane*. Untuk mencari nilai dari variabel b dapat menggunakan persamaan 2.

$$b = \frac{1}{NSV} \sum_{x_j \in SV} \left(\frac{1}{y_i} - \sum_{x_j \in SV} (\alpha_j y_j K(x_j, x_i))\right) \quad (2)$$

2.3. Mutual Information

Mutual Information adalah salah satu metode perhitungan yang dapat digunakan untuk melakukan seleksi fitur. Metode ini melakukan perhitungan dengan cara memperhatikan seberapa banyak informasi suatu fitur dan kontribusinya dalam menentukan hasil keputusan klasifikasi suatu term pada suatu kategori kata [2].

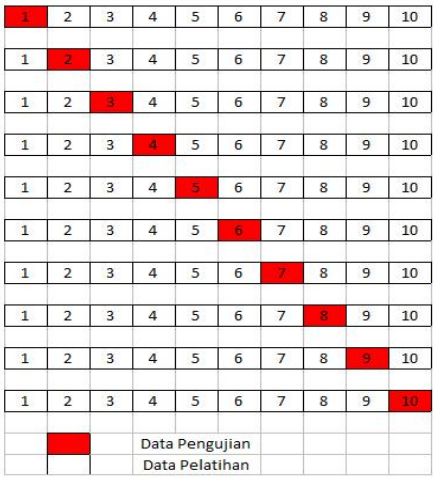
Perhitungan nilai MI pada tiap kategori suatu kata dapat ditampilkan pada persamaan 3.

$$I(U,C) = \frac{N_{11}}{N} log_2 \frac{N.N_{11}}{N_1.N_1} + \frac{N_{01}}{N} log_2 \frac{N.N_{01}}{N_0.N_1} + \frac{N_{10}}{N} log_2 \frac{N.N_{10}}{N_1.N_0} + \frac{N_{00}}{N} log_2 \frac{N.N_{00}}{N_0.N_0} \quad (3)$$

Nilai N merupakan variabel yang menunjukkan jumlah dokumen yang memiliki *et* dan *ec*. Selanjutnya hasil perhitungan yang telah didapat pada tiap kategori akan dipilih hasil terbesar dan akan dilakukan pengurutan pada setiap fitur berdasarkan hasil perhitungan seleksi dari mulai terbesar hingga terkecil [4].

2.4. Cross Validation

Cross Validation adalah suatu metode statistik yang ditujukkan untuk mengevaluasi kinerja algoritma klasifikasi dengan cara membagi data latih dan uji ke dalam beberapa fold. Penelitian ini menggunakan metode 10 k-fold *Cross Validation*, metode ini bekerja dengan cara membagi data menjadi 10 bagian secara acak. Dengan rincian berupa data bagian ke-k menjadi data uji dan sisa bagian lain sebagai data uji. Percobaan dilakukan sesuai dengan jumlah k yang digunakan [10]. Ilustrasi penggunaan gambar 10 k-fold *Cross Validation* ditampilkan pada Gambar 2.



Gambar 2. Ilustrasi 10 k-fold *Cross Validation*

2.5. Confusion Matrix

Confusion Matrix adalah suatu metode yang digunakan untuk mengukur hasil kinerja dari suatu model klasifikasi. Dalam penelitian ini menggunakan matrix dengan ukuran 3x3. Matrix 3x3 dibentuk melalui 3 kelas yang terdapat pada *Confusion Matrix* dan menghasilkan 9 sel matrix [11]. Model *Confusion Matrix* dalam penelitian ini ditampilkan pada Tabel 1.

Tabel 1. Model *Confusion Matrix*

Fakta	Prediksi		
	A	**B**	**C**
A	TA	FB1	FC1
B	FA1	TB	FC2
C	FA2	FB2	TC

Hasil kinerja yang diukur dalam *Confusion Matrix* terdiri dari *accuracy, precision, recall* dan *f-measure*. *Accuracy* digunakan untuk menghitung ketepatan suatu model klasifikasi. Persamaan dalam menghitung nilai *accuracy* ditampilkan pada persamaan 4.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

Recall digunakan untuk menghitung kinerja model dalam menemukan informasi yang relevan. Persamaan dalam menghitung *recall* ditampilkan pada persamaan 5.

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Precision digunakan untuk menghitung perbandingan dari total dokumen yang memiliki hubungan dengan total keseluruhan dokumen. Persamaan dalam menghitung *precision* ditampilkan pada persamaan 6.

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

F-measure digunakan untuk menghitung nilai keseimbangan anatara nilai *precision* dan *recall*. Persamaan dalam menghitung *F-measure* ditampilkan pada persamaan 7.

$$F-measure = \frac{2\,X\,(Precision\,X\,Recall)}{Precision+Recall} \quad (7)$$

3. HASIL DAN PEMBAHASAN

Bab ini akan membahas mengenai uraian dari hasil penelitian pada sistem klasifikasi pertanyaan berbahasa Indonesia menggunakan algoritma *Support Vector Machine* (SVM) dan metode seleksi fitur *Mutual Information*.

3.1. Skenario Pengujian

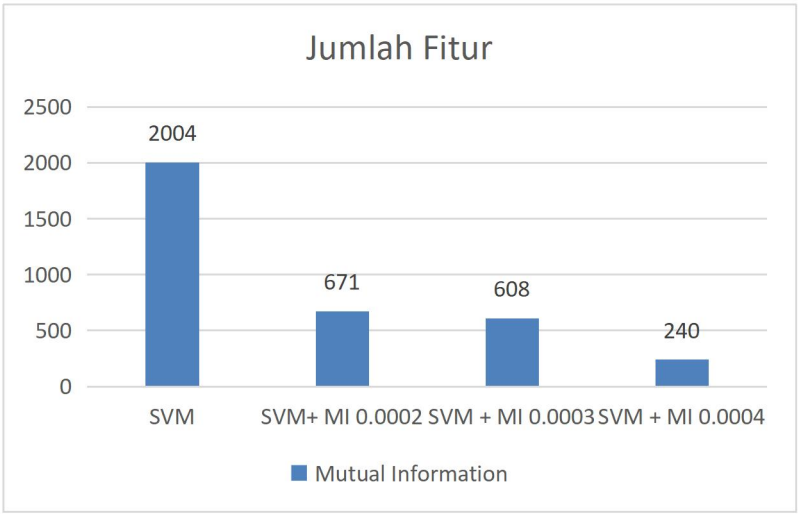
Untuk mengetahui hasil kinerja pada penggunaan model pengklasifikasi, maka dibuat rancangan skenario pengujian menggunakan metode 10 k- *Fold Cross Validation* yang membagi data menjadi 9 data latih dan 1 data uji sebanyak 10 kali percobaan secara acak. Data latih merupakan data yang digunakan untuk melatih model pengetahuan berdasarkan data latih berupa pertanyaan berbahasa Indonesia untuk setiap kategori pertanyaan yang terdiri dari kelas *factoid, non-factoid* dan *other*. Kemudian, data uji merupakan data yang digunakan untuk mengukur hasil kinerja dari model klasifikasi.

Skenario pengujian dibagi menjadi dua bagian, yakni skenario pengujian untuk model klasifikasi pertanyaan menggunakan algoritma *Support Vector Machine* tanpa seleksi fitur. Dan skenario kedua pengujian untuk model klasifikasi pertanyaan menggunakan algoritma

Support Vector Machine dengan seleksi fitur *Mutual Information*. Pada skenario pengujian juga akan digunakan beberapa parameter berupa kernel yang digunakan sebagai pengoptimasi algoritma SVM dalam mengubah data kedalam ruang fitur yang berdimensi tinggi. Parameter kernel yang digunakan meliputi kernel *Linear, Polynomial* dan Rbf. Selanjutnya, digunakan parameter berupa nilai C yang digunakan untuk mengontrol nilai *trade off* pada error yang berguna untuk memberikan informasi seberapa banyak kesalahan klasifikasi yang ingin dihindari. Parameter nilai C yang digunakan meliputi nilai 0,1, 1, dan 10. Kemudian, digunakan parameter *treshold* yang digunakan untuk menentukan nilai ambang batas dari bobot hasil seleksi fitur yang akan dieliminasi. Hasil pengujian dari setiap skenario pengujian akan dilakukan perbandingan dan analisa untuk melihat pengaruh dari penggunaan metode seleksi fitur pada algoritma klasifikasi SVM.

3.2. Hasil dan Analisis Pengujian

Penggunaan parameter berupa *treshold* digunakan untuk menentukan nilai ambang batas pada metode seleksi fitur *Mutual Information*. Pada penelitian ini, penggunaan parameter *treshold* dengan nilai 0.0002, 0.0003 dan 0.0004 dapat mereduksi jumlah fitur dengan total 2004 fitur menjadi sejumlah fitur yang memiliki bobot nilai diatas masing-masing nilai ambang batas. Hasil reduksi ditampilkan pada grafik jumlah fitur pada Gambar 3.



Gambar 3. Grafik Hasil Perbandingan Jumlah Fitur

Berdasarkan tampilan grafik diatas dapat diketahui bahwa semakin besar nilai *treshold* yang digunakan maka jumlah fitur yang terpilih menjadi semakin sedikit. Penggunaan parameter dengan *treshold* 0.0004 berhasil mereduksi jumlah fitur menjadi 240 fitur. Selanjutnya, penggunaan parameter tersebut dapat menghasilkan kinerja dan efisiensi waktu komputasi yang lebih baik dibandingkan penggunaan *treshold* lain sehingga *treshold* dengan nilai 0.0004 akan digunakan pada skenario pengujian pada model algoritma SVM dan Metode Seleksi Fitur *Mutual Information*.

Pada subbab sebelumnya diuraikan bahwa skenario pengujian dilakukan menjadi dua bagian, yaitu skenario pertama untuk model SVM tanpa seleksi fitur. Dan skenario kedua untuk model SVM dengan metode seleksi fitur *Mutual Information*. Hasil kinerja dari kedua skenario tersebut akan ditampilkan pada tabel perbandingan untuk setiap parameter SVM berupa *accuracy, precision, recall, f-measure* dan jumlah fitur. Hasil perbandingan dari skenario pengujian untuk setiap parameter SVM ditampilkan pada Tabel 2, Tabel 3 dan Tabel 4.

Tabel 2. Hasil Skenario Pengujian pada Kernel Linear

Algoritma	Nilai C	Nilai *Treshold*	Kernel : Linear				
			Accuracy	*Precision*	*Recall*	*F-Measure*	Waktu Komputasi
SVM	0.1	-	0.75	0.51	0.59	0.54	0.14
	1		0.91	0.92	0.86	0.88	0.12
	10		0.89	0.9	0.86	0.87	0.14
SVM+MI	0.1	0.0004	0.79	0.53	0.62	0.57	0.07
	1		0.92	0.92	0.89	0.9	0.07
	10		0.9	0.9	0.88	0.89	0.05

Tabel 3. Hasil Skenario Pengujian pada Kernel Polynomial

Algoritma	Nilai C	Nilai *Treshold*	Kernel : Polynomial				
			Accuracy	*Precision*	*Recall*	*F-Measure*	Waktu Komputasi
SVM	0.1	-	0.43	0.15	0.33	0.2	0.18
	1		0.75	0.84	0.61	0.59	0.19
	10		0.76	0.77	0.63	0.62	0.18
SVM+MI	0.1	0.0004	0.76	0.51	0.6	0.55	0.09
	1		0.89	0.9	0.84	0.86	0.1
	10		0.89	0.88	0.85	0.86	0.11

Tabel 4. Hasil Skenario Pengujian pada Kernel Rbf

Algoritma	Nilai C	Nilai *Treshold*	Kernel : Rbf				
			Accuracy	*Precision*	*Recall*	*F-Measure*	Waktu Komputasi
SVM	0.1	-	0.44	0.35	0.34	0.21	0.16
	1		0.87	0.9	0.84	0.81	0.18
	10		0.88	0.9	0.82	0.84	0.19
SVM+MI	0.1	0.0004	0.78	0.52	0.62	0.56	0.08
	1		0.91	0.92	0.87	0.89	0.09
	10		0.91	0.9	0.88	0.89	0.09

Tabel hasil skenario pengujian model klasifikasi diatas menunjukkan hasil kinerja yang diperoleh setiap model klasifikasi. Pada model klasifikasi menggunakan algoritma SVM tanpa seleksi fitur menghasilkan kinerja yang cukup baik namun kurang stabil pada parameter SVM tertentu. Model ini dapat menghasilkan hasil kinerja terbaik dengan nilai akurasi sebesar 91% dan waktu komputasi sebesar 19 s pada parameter kernel *Linear* dan nilai C : 1. Sedangkan, penggunaan metode seleksi fitur *Mutual Information* pada algoritma klasifikasi SVM menunjukkan hasil kinerja yang sangat baik dan relatif stabil pada setiap parameter SVM. Peningkatan hasil dan efisiensi kinerja yang cukup signifikan terlihat pada parameter kernel polynomial dengan nilai C : 1. Dimana penggunaan model SVM dengan metode seleksi fitur *Mutual Information* mampu meningkatkan akurasi dari 75% menjadi 89%, mampu mempercepat waktu komputasi dari 19 s menjadi 10 s dan mampu mengurangi jumlah fitur dari 2004 menjadi 240. Selanjutnya, penggunaan model SVM dengan metode seleksi fitur *Mutual*

Information mampu mendapatkan hasil kinerja terbaik pada kernel *Linear* dengan nilai akurasi 92 % dan waktu komputasi sebesar 7 s pada kernel *Linear* dan nilai C : 1.

4. KESIMPULAN

Berdasarkan uraian yang dimuat pada bab pembahasan dan hasil penelitan sebelumnya, maka uraian tersebut dapat disumpulkan menjadi sebagai berikut.

1. Proses pada system pengklasifikasian pertanyaan berbahasa Indonesia menggunakan algoritma Support Vector Machine (SVM) dan metode seleksi fitur *Information Gain, Chi Square* dan *Mutual Information* berhasil diimplementasikan.
2. Penggunaan metode seleksi fitur *Mutual Information* yang dikombinasikan dengan algoritma klasifikasi SVM pada kernel *linear* dengan parameter berupa nilai C :1 dan treshold : 3.5 berhasil mendapatkan hasil kinerja terbaik dengan nilai *accuracy* : 0.92, *precision* : 0.93, *recall* : 0.89, *f-measure* : 0.9,waktu komputasi : 7 s dan jumlah fitur : 240.

5. SARAN

Pada penelitian ini masih memiliki beberapa kekurangan yang dapat dikembangkan pada penelitian selanjutnya. Beberapa saran yang harus dilakukan adalah melakukan perbandingan terhadap metode seleksi fitur lain baik bertipe filter, wrapper dan embedded selector. Kemudian, melakukan penambahan data pada setiap kategori pertanyaan berbahasa Indonesia untuk mencegah terjadinya data tidak seimbang.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada redaksi jurnal JUPITER atas kesempatan yang telah diberikan ke penulis sehingga penerbitan jurnal ini dapat terlaksana.

DAFTAR PUSTAKA

[1] A. Purwarianti and N. Yusliani, “Sistem Question Answering Bahasa Indonesia Untuk Pertanyaan Non-Factoid,” *J. Ilmu Komput. dan Inf.*, vol. 4, no. 1, pp. 10–14, 2011, doi: 10.21609/jiki.v4i1.151.

[2] A. Hanafi, A. Adiwijaya, and W. Astuti, “Klasifikasi Multi Label pada Hadis Bukhari Terjemahan Bahasa Indonesia Menggunakan Mutual Information dan k-Nearest Neighbor,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, pp. 357–364, 2020, doi: 10.32736/sisfokom.v9i3.980.

[3] S. A. Karunia, “O nline News Classification Using Naive Bayes Classifier with Mutual Information for Feature Selection,” vol. 6, no. 1, 2017.

[4] L. G. Irham, A. Adiwijaya, and U. N. Wisesty, “Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine,” *J. Media Inform. Budidarma*, vol. 3, no. 4, p. 284, 2019, doi: 10.30865/mib.v3i4.1410.

[5] Saniyah, *Named Entity Recognition pada Teks Berita menggunakan Support Vector Machine*. 2019.

[6] D. Buani, “Penerapan Algoritma Naïve Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung,” vol. 9, no. 2, pp. 43–48, 2021.

[7] D. Zhafira, B. Rahayudi, and Indriati, “ANALISIS SENTIMEN KEBIJAKAN

KAMPUS MERDEKA MENGGUNAKAN NAIVE BAYES DAN PEMBOBOTAN TF-IDF SENTIMENT ANLAYSIS OF KAMPUS MERDEKA POLICY USING NAIVE BAYES AND TF-IDF TERM WEIGHTING BASED ON YOUTUBE," vol. 2, no. 1, pp. 55–63, 2021.

[8] I. Yulietha, S. Faraby, and Adiwijaya, "Klasifikasi Sentimen Review Film Menggunakan Algoritma Support Vector Machine Sentiment Classification of Movie Reviews Using Algorithm Support Vector Machine," vol. 4, no. 3, pp. 4740–4750, 2017.

[9] L. Mutawalli, M. T. A. Zaen, and W. Bagye, "KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto)," *J. Inform. dan Rekayasa Elektron.*, vol. 2, no. 2, p. 43, 2019, doi: 10.36595/jire.v2i2.117.

[10] I. A. M. SUPARTINI, I. K. G. SUKARSA, and I. G. A. M. SRINADI, "Analisis Diskriminan Pada Klasifikasi Desa Di Kabupaten Tabanan Menggunakan Metode K-Fold Cross Validation," *E-Jurnal Mat.*, vol. 6, no. 2, p. 106, 2017, doi: 10.24843/mtk.2017.v06.i02.p154.

[11] H. Irmanda and R. Astriratma, "Klasifikasi Jenis Pantun Dengan Metode Support Vector Machines (SVM)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 5, pp. 915–922, 2020, doi: 10.29207/resti.v4i5.2313.