

Klasifikasi Kalimat Tanya Berdasarkan Taksonomi Bloom Menggunakan Support Vector Machine

Setio Basuki^{*1}, Zakiyah Rakhmawati², Galih Wasis Wicaksono³

^{1,2,3}Teknik Informatika/Universitas Muhammadiyah Malang

setiobasuki@umm.ac.id^{*1}, rakhma.zakiyah@webmail.umm.ac.id², galih.w.w@umm.ac.id³

Abstrak

Klasifikasi adalah suatu teknik dalam melakukan kategori pada dokumen yang memiliki jumlah isi dokumen sangat banyak. Pada penelitian ini mengusulkan sebuah metode yang dapat menghasilkan klasifikasi kalimat tanya sesuai dengan level untuk ranah kognitif pada Taksonomi Bloom. Proses identifikasi soal dilakukan berdasarkan ekstraksi fitur sintaktik dan Bag-of-Words (BoW). Hasil dari ekstraksi fitur tersebut kemudian akan diklasifikasikan menggunakan algoritma klasifikasi SVM. Tugas akhir ini bertujuan untuk melakukan hasil klasifikasi kalimat tanya berdasarkan taksonomi bloom untuk ranah kognitif yang terbagi dalam 6 level yaitu, mengingat, memahami, menerapkan, menganalisis, mengevaluasi, dan membuat. Algoritma yang digunakan pada penelitian ini adalah Support Vector Machine (SVM). Algoritma tersebut dipilih karena mampu mengklasifikasikan data berdimensi tinggi yang dalam konteks tugas akhir ini adalah data berupa teks. Berdasarkan skenario pengujian yang sudah dilakukan, hasil rata-rata akurasi yang diperoleh pada proses pengklasifikasian kalimat menggunakan fitur sintaktik pada data uji sebesar 89.2%, sedangkan untuk fitur Bag-of-Words (BoW) tingkat akurasi yang diperoleh sebesar 100%. Berdasarkan hasil pengujian terhadap sistem klasifikasi yang telah dilakukan, rata-rata nilai akurasi didapatkan sangat baik, sehingga sistem dapat dikatakan atau layak untuk dikembangkan selanjutnya dengan berbagai variasi fitur atau algoritma pendukung lainnya.

Kata Kunci: Klasifikasi, Support Vector Machine, Taksonomi Bloom, Sintaktik, Bag Of Words

Abstract

Classification is a technique in the category of documents that have a lot of document content. In this study proposes a method that can generate the classification of sentences in accordance with the level for cognitive domain in Bloom's Taxonomy. The problem identification process is performed based on syntactic feature extraction and Bag-of-Words (BoW). The results of the feature extraction will then be classified using the SVM classification algorithm. This final project aims to perform the classification of sentences based on taxonomy of bloom for the cognitive domain which is divided into 6 levels namely, remembering, understanding, applying, analyzing, evaluating, and making. The algorithm used in this research is Support Vector Machine (SVM). The algorithm is chosen because it is able to classify high-dimensional data which in the context of this final task is text data. Based on the test scenario that has been done, the average result of accuracy obtained in the process of classification of sentences using syntactic features in the test data of 89.2%, while for the Bag-of-Words (BoW) 100% accuracy rate is obtained. Based on the results of testing against the classification system that has been done, the average accuracy value obtained very well, so the system can be said or eligible to be developed further with variations of features or other supporting algorithms.

Keywords: Classification, Support Vector Machine, Bloom's Taxonomy, Syntax, Bag of Words

1. Pendahuluan

Sebagai penunjang dalam pembelajaran, siswa memiliki pegangan yaitu buku pelajaran untuk melakukan pelatihan dalam menyelesaikan setiap persoalan dalam soal.

Peraturan perundang-undangan telah disebutkan bahwa pedoman utama untuk menunjang siswa dalam melakukan proses pembelajaran terletak pada buku yang sudah disediakan oleh pengajar sebagai sumber paling utama pendukung siswa dalam belajar. [1] yang sudah melakukan praktik pada penelitiannya. Salah satu kelemahan yang ditemukan dalam penelitian [2] adalah rendahnya proporsi soal-soal dalam buku teks yang bertujuan agar siswa mampu menggunakan kemampuan penalaran mereka dalam mengerjakan atau menyelesaikan setiap soal yang diberikan oleh pendidik dalam semua kategori soal yang diberikan. Sehingga

perlu dilakukan proses identifikasi terhadap kemampuan kognitif siswa untuk mengetahui sejauh mana pemahaman siswa terhadap pembelajaran yang sudah diajarkan. Hasil dari identifikasi tersebut nantinya yang akan dijadikan sebagai pedoman untuk dasar penilaian.

Taksonomi Bloom, terbagi menjadi tiga ranah ("domain"), yaitu ranah kognitif, ranah afektif dan ranah psikomotor. Untuk ranah kognitif Taksonomi Bloom, telah dilakukan perubahan yang mengakibatkan terjadi pemisahan antara pengetahuan dari proses kognitif. Sehingga, pada Taksonomi Bloom yang telah dilakukan perubahan, terbagi dalam dua dimensi yaitu dimensi pengetahuan dan dimensi proses kognitif [3]. Didakannya teknik pengklasifikasian ini bertujuan untuk memberikan kemudahan pada siswa dalam menyelesaikan setiap persoalan dalam soal berdasarkan tahapan untuk kategori soal termudah hingga tersulit. Pengklasifikasian pertanyaan tersebut menggunakan taksonomi pendidikan yang telah dilakukan oleh Bloom [4]. Menurut [5] menyatakan bahwa dalam praktiknya, tingkat kesulitan soal akan mengikuti hirarki taksonomi kognitif dari Bloom. Penulisan tugas akhir ini menggunakan salah satu algoritma klasifikasi yaitu *Support Vector Machine (SVM)*. *Support Vector Machine* yaitu algoritma yang dalam klasifikasi bekerja pada prinsip *linear classifier* yang selanjutnya dikembangkan untuk dapat melakukan atau mengklasifikasi dalam persoalan yang *non-linear*. Pada penelitian [6] digunakan algoritma SVM untuk kategorisasi teks berbahasa Indonesia.

2. Metode Penelitian

2.1 Studi Pustaka

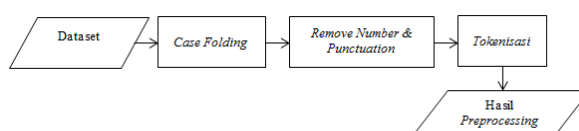
Mengumpulkan semua referensi dan memahami konsep dari system yang akan dibangun tentang studi kasus yang diambil, yaitu mengenai algoritma *Support Vector Machine*. Studi pustaka diperoleh dari berbagai referensi seperti jurnal, website resmi, dan laporan Tugas Akhir (TA), dsb.

2.2 Persiapan Data

Data yang disiapkan dalam tugas akhir ini yaitu kalimat tanya Berbahasa Indonesia, dimana kalimat tanya tersebut didapat dari soal-soal pada buku seperti Psikotes dan Lembar Kerja Siswa (LKS). Selain itu sebagai penunjang literatur pengambilan data yang lain, data kalimat tanya juga diambil dari situs bukusekolahdigital.com.

2.3 Preprocessing Data

Preprocessing Data dilakukan untuk membersihkan data kalimat tanya. Pada Gambar 1, yang dilakukan pada proses *preprocessing data* antara lain *case folding*, *remove number & punctuation* serta tokenisasi.



Gambar 1. Flowchart Sistem Preprocessing

2.4 Ekstraksi Fitur

Ekstraksi fitur dilakukan untuk mendapatkan karakteristik pembeda antar kelas untuk mempermudah proses klasifikasi. Fitur yang digunakan dalam penelitian ini adalah fitur sintaktik dan fitur *bag of words*.

2.4.1 Fitur Sintaktik

Fitur sintaktik adalah fitur dari sebuah soal yang diekstrak berdasarkan susunan kata pada soal tersebut [7]. Dalam penyusunan pola terhadap kalimat, terdapat beberapa kata kunci taksonomi bloom dan kata kunci pendukung lainnya yang dianalisis pada *dataset* untuk memperkuat hasil klasifikasi. Berikut beberapa daftar ekstraksi fitur pada Tabel 1.

Tabel 1. Daftar Ekstraksi Fitur Sintaktik

No	Fitur	Deskripsi
1	Kata WH	<true, false> Kata Kunci 5W + 1H
2	Kata Kerja	<true,false> Kata Kunci Kerja
3	Kata Perbandingan	<true,false> Kata Kunci Perbandingan

4	Kata Definisi	<true,false> Kata Kunci Definisi
5	Kata Kausalitas	<true,false> Kata Kunci Kausalitas
6	Kata Penyebutan	<true,false> Kata Kunci Penyebutan
7	Kata Preposisi	<true,false> Kata Kunci Preposisi
8	Kata Penjelas	<true,false> Kata Kunci Penjelas
9	Kata Pilihan	<true,false> Kata Kunci Pilihan
10	Kata Tujuan	<true,false> Kata Kunci Tujuan
11	Kata Cara	<true,false> Kata Kunci Cara
12	Kata Waktu	<true,false> Kata Kunci Waktu
13	Kata Tambahan	<true,false> Kata Kunci Tambahan
14	Kata Mengingat	<true,false> Kata Kunci Mengingat
15	Kata Memahami	<true,false> Kata Kunci Memahami
16	Kata Menerapkan	<true,false> Kata Kunci Menerapkan
17	Kata Menganalisis	<true,false> Kata Kunci Menganalisis
18	Kata Mengevaluasi	<true,false> Kata Kunci Mengevaluasi
19	Kata Membuat	<true,false> Kata Kunci Membuat
20	Class	<nominal> mengingat, memahami, menerapkan, menganalisis, mengevaluasi, membuat

2.4.2 Fitur Leksikal (Bag of Words)

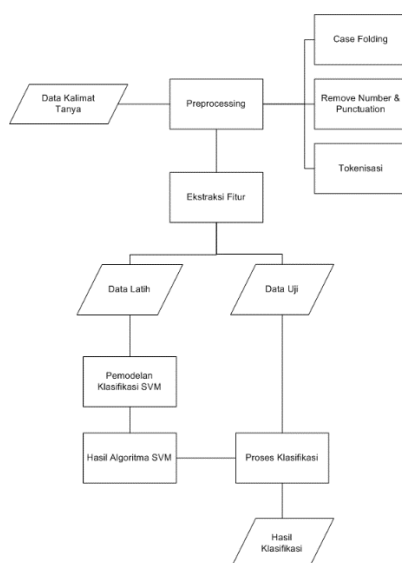
Bag-of-Words adalah pemodelan yang dilakukan berdasarkan objek secara umum atau global seperti halnya kalimat teks atau dokumen yang digunakan sebagai bag atau dengan kata lain multiset kata tanpa melihat susunan tata bahasa maupun urutan kata untuk menjaga variasi kata [8]. Dengan kata lain, BoW merupakan sekumpulan kata-kata dalam teks dokumen untuk membentuk urutan agar dapat dihitung frekuensi kemunculan kata dalam domen tersebut. Tabel 2 merupakan contoh dari fitur Leksikal sebagai berikut.

Tabel 2. Ekstraksi Fitur

Kalimat : “jelaskan perbedaan kehidupan pada pranata penjara, pranata rumah sakit, dan pranata tempat pemukiman orang lanjut usia”

Fitur	Ekstraksi
Bag of Words	{(jelaskan,1), (perbedaan,1), (kehidupan,1), (pada,1), (pranata,3), (penjara,1), (rumah,1), (sakit,1), (dan,1), (tempat,1), (pemukiman,1), (orang,1), (lanjut, 1), (usia, 1)}

2.5 Klasifikasi menggunakan Support Vector Machine



Gambar 2. Alur Sistem Klasifikasi

Model klasifikasi yang digunakan pada tugas akhir ini yaitu menggunakan algoritma klasifikasi *Support Vector Machine*. Sebelum melakukan tahap klasifikasi dilakukan tahap *preprocessing* dan ekstraksi fitur, seperti pada Gambar 2.

Konsep SVM berusaha menemukan fungsi pemisah (*hyperplane*) untuk mencari bidang pemisah terbaik diantara fungsi yang tidak terbatas jumlahnya. Menemukan *Hyperplane* untuk mendapatkan pemisah terbaik antara kedua kelas dapat dilakukan dengan mengukur dari margin *hyperplane* tersebut agar dapat dicari titik maksimalnya. Adapun data yang berada pada bidang pembatas disebut *support vector* [9]. Pada dasarnya, konsep dasar dari algoritma SVM ditunjukkan pada Persamaan 1 dan Persamaan 2.

$$\min \frac{1}{2} |w|^2 \quad (1) \quad [9]$$

$$\text{s.t } y_i(x_i \cdot w + b) - 1 \geq 0 \quad (2)$$

dimana $(x_i \cdot w + b) \geq 1$ untuk kelas 1, dan $(x_i \cdot w + b) \leq -1$ untuk kelas 2,

Keterangan:

x_i adalah data set

y_i adalah output dari data x_i , dan

w, b adalah parameter yang dicari nilainya.

2.6 Pengujian

Pengujian yang dilakukan pada penelitian tugas akhir ini, yaitu pengujian terhadap hasil klasifikasi kalimat tanya berdasarkan target kelas. Tabel 3 berikut merupakan scenario pengujian.

Tabel 3. Skenario Pengujian

Tahap	Parameter Pengujian	Parameter Data
I	Data Latih	Kalimat Tanya
II	Hasil Klasifikasi	Kalimat Tanya

untuk setiap klasifikasi SVM di hitung nilai akurasi dengan Persamaan 3.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4) \quad [9]$$

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

3. Hasil Penelitian dan Pembahasan

3.1 Persiapan Data

Pengerjaan tugas akhir ini data yang digunakan adalah kalimat tanya Berbahasa Indonesia, dimana kalimat tanya tersebut diambil dari soal-soal pada buku seperti Psikotes dan Lembar Kerja Siswa (LKS). Selain itu sebagai penunjang literatur pengambilan data yang lain, data kalimat tanya juga diambil dari situs bukusekolahdigital.com. Data yang terkumpul kemudian diklasifikasikan secara manual terlebih dahulu untuk mengetahui setiap target kelas dari masing-masing kalimat tanya. Jumlah data yang digunakan dalam penelitian ± 600 data kalimat tanya dan data dari peneliti sebelumnya ± 900 . Data tugas akhir disimpan dalam format *Comma Separated Value* (CSV). Berikut contoh dari data tugas akhir seperti pada Tabel 4.

Tabel 4. Data Kalimat Tanya

Identifikasikan sebab terjadinya kemajemukan oleh beragam suku bangsa yang tersebar di wilayah Indonesia?

Mengapa sifat unsur non logam semakin bertambah pada pergeseran dari kiri ke kanan dalam tabel periodik? Jelaskan!

Apakah situasi dan kondisi kehidupan masyarakat Indonesia dari hari ke hari kian memburuk merupakan salah satu dampak dari globalisasi? Kemukakan pendapatmu!

Temukan perbedaan pelaksanaan Politik Bebas Aktif Indonesia sejak masa awal kemerdekaan hingga masa Reformasi!
 Buktikan pernyataan bahwa Sosiologi merupakan ilmu sosial, bersifat heterogen, ilmu murni, dan abstrak!
 Kelompokkan tumbuhan-tumbuhan dikotil dan monokotil berdasarkan persamaan dan perbedaan ciri yang dimiliki!

3.2 Preprocessing Data

Langkah pertama yang dilakukan sistem sebelum melakukan klasifikasi yaitu dilakukan pembersihan pada data kalimat tanya. Gambar 3 berikut merupakan hasil dari *preprocessing data*.

- _____ tulislah faktor penghambat yang dihadapi untuk melaksanakan berbagai aktivitas ekonomi yang dominan berkembang di daerah tempat tinggalmu _____
- _____ identifikasilah hal apa saja yang diperlukan untuk mengatasi berbagai masalah yang muncul dalam kehidupan bermasyarakat _____
- _____ deskripsikan akibat yang terjadi jika kemerdekaan dalam menyampaikan pendapat tidak ada peraturan _____
- _____ bagaimana persamaan kalor untuk menaikkan suhu benda dan kalor untuk mengubah wujud benda coba jelaskan _____

Gambar 3. Hasil Preprocessing Data

3.3 Ekstraksi Fitur

3.3.1 Fitur Sintaktik

Setelah selesai dilakukan proses pembersihan data, pengguna kemudian melakukan tahap ekstraksi fitur. Sebelum dilakukan ekstraksi fitur, pengguna terlebih dahulu melakukan pelabelan pada data kalimat tanya ke dalam beberapa target kelas taksonomi bloom untuk ranah kognitif. Pelabelan data tersebut dilakukan secara manual, dan setelah data selesai dilakukan pelabelan kemudian dapat disimpan dengan format *.csv. Gambar 4 berikut merupakan contoh data setelah dilakukan pelabelan.

mengingat,sebutkan pula termasuk dalam kategori apa cara penyajian fakta sejarah dalam artikel tersebut										
mengingat,sebutkan contoh ketampakan bentuk muka bumi sebagai hasil proses tenaga endogen yang ada atau dapat diamati dari tempat tinggal anda										
mengingat,sebutkan daerahdaerah di indonesia yang merupakan daerah bayangan hujan dan jelaskan apa akibatnya										
mengingat,sebutkan beberapa peran tokoh pengembang agama islam di indonesia										
mengingat,sebutkan prosesproses dalam siklus air										
mengingat,sebutkan halhal apa saja yang dalam bentuk rintangan sering terjadi dalam perjalanan yang dilalui oleh seseorang										
mengingat,sebutkan pula termasuk dalam kategori apa cara penyajian fakta sejarah dalam artikel tersebut										

Gambar 4. Data Pelabelan Target Kelas

Ekstraksi fitur sintaktik digunakan 19 fitur untuk mengekstraksi kalimat agar mendapatkan hasil klasifikasi. Gambar 5 berikut merupakan hasil dari ekstraksi fitur Sintaktik.

1,1,0,0,0,0,1,0,0,0,1,0,1,1,1,1,0,1,1,mengingat
 1,1,0,1,0,0,1,0,1,0,0,0,1,1,1,1,0,0,1,mengingat
 1,1,0,1,0,0,1,0,0,0,0,0,1,1,1,1,0,0,0,mengingat
 1,1,0,0,0,1,1,0,0,0,0,0,1,1,1,1,0,0,0,mengingat
 0,1,0,0,0,0,1,0,0,0,0,0,1,1,1,1,0,0,0,mengingat
 1,1,0,0,0,0,1,0,0,0,0,1,1,1,1,1,0,0,1,mengingat
 1,1,0,0,0,0,1,0,0,0,1,0,1,1,1,1,0,1,1,mengingat

Gambar 5. Hasil Data Ekstraksi Fitur Sintaktik

3.3.2 Fitur Leksikal (Bag of Words)

Tahap ekstraksi fitur *Bag-of-Words*, untuk data awal berbeda dengan data awal yang ada pada ekstraksi fitur sintaktik. Dalam melakukan ekstraksi fitur untuk *Bag-of-Words*, data yang sudah terlabeli diubah dan disesuaikan dengan bentuk data dalam format *.arff. Berikut merupakan Gambar 6 contoh untuk data *Bag-of-Words* sebelum melakukan tahap ekstraksi.

```

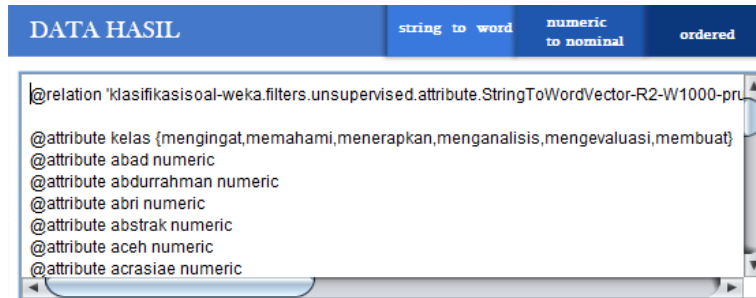
@relation klasifikasisoal
@attribute kelas {mengingat,memahami,menerapkan,menganalisis,mengevaluasi,membuat}
@attribute kalimat String

@data
mengingat,'sebutkan pula termasuk dalam kategori apa cara penyajian fakta sejarah dalam artikel tersebut'
mengingat,'sebutkan contoh ketampakan bentuk muka bumi sebagai hasil proses tenaga endogen yang ada atau dapat diamati dari tempat tinggal anda'
mengingat,'sebutkan daerahdaerah di Indonesia yang merupakan daerah bayangan hujan dan jelaskan apa akibatnya'
mengingat,'sebutkan beberapa peran tokoh pengembang agama islam di Indonesia'

```

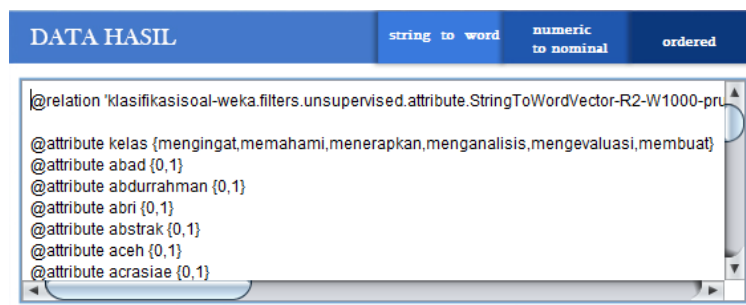
Gambar 6. Data Awal Bag-of-Words

Selanjutnya adalah proses filter terhadap *dataset*, yaitu tipe data pada data awal berbentuk *String* akan diubah menjadi bentuk *Numeric*. Proses pembentukan data tersebut sudah tersedia pada WEKA yaitu dengan *keyword* *StringToWordVector*. Gambar 7 merupakan hasil dari *StringToWordVector*.



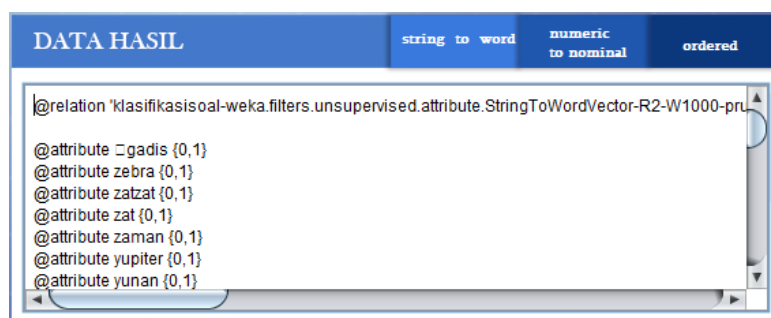
Gambar 7. Hasil Filter StringToWordVector

Setelah dilakukan filter *StringToWordVector*, kemudian dilakukan filter data *NumericToNominal*. Hal ini dikarenakan algoritma *Support Vector Machine* tidak mendukung tipe data *numeric*, sehingga perlu dilakukan filter data *NumericToNominal*. Gambar 8 dibawah ini menunjukkan hasil filter data *NumericToNominal*.



Gambar 8. Hasil Filter NumericToNominal

Setelah dilakukannya tahap filter *NumericToNominal*, tahap akhir dari proses ekstraksi fitur *Bag-of-Words* yaitu filter *Ordered*. Dimana pada tahap *Ordered* ini dilakukan untuk mengubah letak dari target kelas yang semula pada tahap *NumericToNominal* berada di *attribute* paling atas, dengan adanya filter *Ordered* ini *attribute* target kelas diubah ke *attribute* paling bawah untuk mempermudah algoritma melakukan klasifikasi nantinya. Berikut hasil filter data *Ordered* ditunjukkan pada Gambar 9.



Gambar 9. Hasil Filter Ordered

3.4 Klasifikasi

3.4.1 Pembentukan Model dari Fitur Sintaktik dan Fitur *Bag of Words*

Pengujian untuk model dari *data train* yang telah diekstraksi. Model tersebut dibentuk berdasarkan algoritma *Support Vector Machine* dan untuk melakukan pengujian terhadap model menggunakan *K-Cross Validation* dengan iterasi 10 kali. Gambar 10 dan Gambar 11 merupakan hasil pengujian model dari data train sebagai berikut.

Cross Validation		
Correctly Classified Instances	769	80.1042 %
Incorrectly Classified Instances	191	19.8958 %
Kappa statistic	0.7613	
K&B Relative Info Score	34344.1907 %	
K&B Information Score	887.7845 bits	0.9248 bits/instance
Class complexity order 0	2481.564 bits	2.585 bits/instance
Class complexity scheme	2666.09 bits	2.7772 bits/instance
Complexity improvement (Sf)	-184.526 bits	-0.1922 bits/instance
Mean absolute error	0.2273	
Root mean squared error	0.3178	
Relative absolute error	81.8333 %	
Root relative squared error	85.2623 %	
Total Number of Instances	960	

Gambar 10. Hasil Pengujian Data Latih Fitur Sintaktik

Cross Validation		
Correctly Classified Instances	1164	97 %
Incorrectly Classified Instances	36	3 %
Kappa statistic	0.964	
K&B Relative Info Score	45385.7423 %	
K&B Information Score	1173.2044 bits	0.9777 bits/instance
Class complexity order 0	3101.955 bits	2.585 bits/instance
Class complexity scheme	6221.6359 bits	5.1847 bits/instance
Complexity improvement (Sf)	-3119.6809 bits	-2.5997 bits/instance
Mean absolute error	0.2238	
Root mean squared error	0.3126	
Relative absolute error	80.58 %	
Root relative squared error	83.8899 %	
Total Number of Instances	1200	

Gambar 11. Fitur *Bag-of-Words*

Berdasarkan hasil pengujian pada Gambar 12 dan Gambar 13, diperoleh hasil yang sangat baik dari jumlah data latih (*data train*) 480 data dengan nilai akurasi pada fitur sintaktik sebesar 80% dan nilai akurasi pada fitur *Bag-of-Words* sebesar 97%.

3.4.2 Pengujian Klasifikasi

Pengujian klasifikasi dilakukan dengan memasukkan data uji sejumlah 120 data uji (*data test*) yang kemudian akan diuji dengan model yang sudah dibentuk untuk memperoleh hasil akurasi. Berikut merupakan hasil pengujian dari data uji yang telah dilakukan.

Summary		
Correctly Classified Instances	107	89.1667 %
Incorrectly Classified Instances	13	10.8333 %
Kappa statistic	0.87	
K&B Relative Info Score	4448.2216 %	
K&B Information Score	114.9849 bits	0.9582 bits/instance
Class complexity order 0	310.1955 bits	2.585 bits/instance
Class complexity scheme	195.2106 bits	1.6268 bits/instance
Complexity improvement (Sf)	114.9849 bits	0.9582 bits/instance
Mean absolute error	0.225	
Root mean squared error	0.314	
Relative absolute error	81 %	
Root relative squared error	84.243 %	
Total Number of Instances	120	

Gambar 12. Hasil Pengujian Data Test Fitur Sintaktik

Summary			
Correctly Classified Instances	120	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
K&B Relative Info Score	4642.2337		%
K&B Information Score	120	bits	1
bits/instance			
Class complexity order 0	310.1955	bits	2.585
bits/instance			
Class complexity scheme	190.1955	bits	1.585
bits/instance			
Complexity improvement (Sf)	120	bits	1
bits/instance			
Mean absolute error	0.2222		
Root mean squared error	0.3103		
Relative absolute error	80		%
Root relative squared error	83.2613		%
Total Number of Instances	120		

Gambar 13. Fitur Bag-of-Words

Mendapatkan akurasi, *precision*, dan *recall* harus diketahui *confusion matrix* dari hasil klasifikasi. Berikut *Confusion Matrix* dari hasil klasifikasi untuk masing-masing fitur:

Confusion Matrix						
a	b	c	d	e	f	<-- classified as
20	0	0	0	0	0	a = mengingat
1	14	1	0	3	1	b = memahami
0	5	13	0	2	0	c = menerapkan
0	0	0	20	0	0	d = menganalisis
0	0	0	0	20	0	e = mengevaluasi
0	0	0	0	0	20	f = membuat

Gambar 14. Confusion Matrix Tes Klasifikasi Fitur Sintaktik

Confusion Matrix						
a	b	c	d	e	f	<-- classified as
20	0	0	0	0	0	a = mengingat
0	20	0	0	0	0	b = memahami
0	0	20	0	0	0	c = menerapkan
0	0	0	20	0	0	d = menganalisis
0	0	0	0	20	0	e = mengevaluasi
0	0	0	0	0	20	f = membuat

Gambar 15. Fitur Bag of Words

Berdasarkan Gambar 14 dan Gambar 15 hasil confusion matrix diatas, maka dapat diketahui hasil dari akurasi, *precision*, dan *recall* dari masing-masing target kelas. Berikut merupakan Tabel 5 dan Tabel 6, detail dari hasil akurasi, *precision* dan *recall*.

Tabel 5. Detail Akurasi Tes Klasifikasi Fitur Sintaktik

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1	0.01	0.952	1	0.976	0.995	mengingat
0.7	0.05	0.737	0.7	0.718	0.894	memahami
0.65	0.01	0.929	0.65	0.765	0.905	menerapkan
1	0	1	1	1		menganalisis
1	0.05	0.8	1	0.889	0.975	mengevaluasi
1	0.01	0.952	1	0.976	0.995	membuat
<i>Weighted Avg.</i>	<i>0.892</i>	<i>0.022</i>	<i>0.895</i>	<i>0.892</i>	<i>0.887</i>	<i>0.961</i>

Tabel 6. Detail Akurasi Tes Klasifikasi Fitur Bag-Of-Words

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0	1	1	1	1	mengingat
	1	0	1	1	1	1	memahami
	1	0	1	1	1	1	menerapkan
	1	0	1	1	1	1	menganalisis
	1	0	1	1	1	1	mengevaluasi
	1	0	1	1	1	1	membuat
<i>Weighted Avg.</i>	1	0	1	1	1	1	

4. Kesimpulan

Berdasarkan hasil pengujian klasifikasi yang sudah dilakukan dengan metode Support Vector Machine (SVM) menunjukkan bahwa hasil akurasi, presisi, dan recall pada masing-masing fitur yaitu fitur sintaktik dan fitur *Bag-of-Words* memberikan hasil akurasi yang cukup baik, sehingga dapat diambil kesimpulan dari pengerjaan tugas akhir ini, klasifikasi kalimat tanya berdasarkan taksonomi bloom menggunakan *support vector machine* (SVM) sistem dan model yang telah dibangun dapat bekerja secara optimal. Berikut beberapa pembahasan dari tugas akhir ini antara lain:

1. Berdasarkan hasil pengujian dari 600 *dataset*, 120 data digunakan sebagai data uji (data test).
2. Pada proses pengujian klasifikasi dilakukan 2 jenis pengujian yaitu pengujian akurasi menggunakan fitur sintaktik dan pengujian akurasi menggunakan fitur *Bag-of-Words*.
3. Hasil akurasi masing-masing fitur berbeda, pada pengujian menggunakan ekstraksi fitur sintaktik diperoleh hasil akurasi, *precision*, dan *recall* sebesar 89.2%, 89.5%, dan 89.2%. Sedangkan untuk pengujian menggunakan fitur *Bag-of-Words* diperoleh hasil akurasi, *precision*, dan *recall* sebesar 100%, 100%, dan 100%.
4. Dari hasil akurasi yang diperoleh pada masing-masing fitur, dapat disimpulkan bahwa pembuatan model pada algoritma *Support Vector Machine* memiliki kelayakan sangat baik dalam melakukan klasifikasi data berupa teks.

Pengembangan pada penelitian selanjutnya, diharapkan kalimat tanya tidak hanya diklasifikasikan kedalam ranah kognitif saja, tetapi dapat dikembangkan untuk klasifikasi pada ranah afektif dan psikomotor, sehingga semua ranah pada konsep Taksonomi Bloom dapat diklasifikasikan berdasarkan tingkat kesulitan pada masing-masing penyusun hierarkinya.

Referensi

- [1] C. H. Giani, Zulkardi, "Analisis Tingkat Kognitif Soal-Soal Buku Teks Matematika Kelas VII Berdasarkan Taksonomi Bloom," vol. 1, 2012.
- [2] A. Masduki. Subandriah, M.R. Irawan, D.Y. Prihantoro, "Level Kognitif soal-soal Pada Buku Teks Matematika SMP Kelas VII," 2013.
- [3] T. T. Pembelajaran, "Taksonomi Tujuan Pembelajaran Ari Widodo Jurusan Pendidikan Biologi FPMIPA – Universitas Pendidikan Indonesia Jl . Dr . Setiabudhi 229 Bandung Email : widodo@upi.edu September 2005 Dari penulis Furst , Hill dan Krathwohl (1956), sudah sejak lama digunak," vol. 4, pp. 61–69, 2005.
- [4] A. H. Nasrulloh, "Analisis Tingkat Kognitif Tes Kompetensi pada Buku Sekolah Elektronik (Bse) Matematika Smp/Mts Kelas IX berdasarkan Berdasarkan Taksonomi Bloom," 2011.
- [5] A. Sudarsyah, "Manajemen Implementasi Kurikulum: Sisi Lain Ujian Nasional," 2013. [Online]. Available: <http://m.kompasiana.com/post/re ad/543693/1/sisi-l>.
- [6] D. Wulandini, F., "A Study on Text Classification for Webmining Based Spatio Temporal Analysis of the Spread of Tropical Diseases.," 2010.
- [7] B. Loni, "Enhanced Question Classification With Optimal Combination of Features. Delft University of Technology," 2011.
- [8] "Bag of words model." [Online]. Available: http://en.wikipedia.org/wiki/Bag-of-words_model. [Accessed: 05-Dec-2014]
- [9] D. Widiastuti, "Analisa Perbandingan Algoritma SVM, Naïve Bayes, dan Decission Tree dalam Mengklasifikasikan Serangan (Attack) pada Sistem Pendeteksi Intrusi," *Dwi Widiastuti*, p. 3, 2007.

