

IMPLEMENTASI KLASIFIKASI SOAL BERDASARKAN TAKSONOMI BLOOM MENGUNAKAN ALGORITMA SVM

Selvia Ferdiana Kusuma¹, Agustono Heriadi², Mohammad Farid Naufal³

^{1,2}Teknik Informatika, Politeknik Kediri

³Teknik Informatika, Fakultas Teknik, Universitas Surabaya

E-mail: *¹selvia@poltek-kediri.ac.id, ²agustono@poltek-kediri.ac.id,

³faridnaufal@staff.ubaya.ac.id

Abstrak – Proses penilaian merupakan aspek penting dalam pembelajaran. Penilaian harus dilakukan secara benar agar dapat mengukur kemampuan peserta didik. Pada soal-soal yang digunakan untuk ujian pada program Studi Teknik Informatika Politeknik Kediri belum dilakukan pengklasifikasian soal berdasar tingkat kesulitannya. Sehingga pada proses penilaian tidak didasarkan atas pemberian soal-soal dengan tingkat kesulitan yang berbeda. Pada tahun 1995, Benjamin Bloom telah memperkenalkan adanya proses pengklasifikasian soal berdasarkan tingkat kesulitannya, metode tersebut dinamakan Taksonomi Bloom. Proses pengklasifikasian soal sesuai level pada taksonomi bloom tidaklah mudah jika dilakukan secara manual. Proses otomatisasi klasifikasi perlu dilakukan ketika akan melakukan klasifikasi soal dalam jumlah yang banyak, misalkan pada proses pengklasifikasian soal pada bank soal. Otomatisasi dilakukan selain untuk mempersingkat waktu juga untuk mengurangi tendensi dari ahli pada pengklasifikasian soal. Proses klasifikasi dilakukan dengan pengidentifikasian fitur leksikal dan sintaktik sebagai proses ekstraksi fitur, kemudian hasil ekstraksi fitur diklasifikasikan menggunakan algoritma SVM. Penelitian ini menghasilkan sebuah aplikasi yang dapat melakukan pengklasifikasian sejumlah soal berdasarkan taksonomi bloom menggunakan algoritma SVM. Aplikasi ini memiliki akurasi klasifikasi soal sebesar 86%.

Kata Kunci — Taksonomi Bloom, Algoritma SVM, leksikal, sintaktik, ekstraksi fitur

Abstract – The assessment process is an important aspect in learning. Assessment must be done correctly in order to measure the ability of learners. Exam questions in Informatics Engineering of Polytechnic Kediri has not been classified based on the difficulty levels. It cause the assessment process is not based on the provision of the problems with different difficulty levels. In 1995, Benjamin Bloom introduced the classification process based on the difficulty levels, the method is called Bloom's Taxonomy. The process of classifying questions based on difficulty levels in bloom taxonomy is not an easy task if it is done manually. Classification automation process needs to be done when we classify questions in large amounts for example classification process in question bank. Automation process not only for time saving but also to reduce the tendency of experts in classifying questions. The classification process is done by extracting lexical and syntactic features then classified using SVM algorithm. This research produce an application that can classifies a number of questions based on Bloom Taxonomy using SVM algorithm. The applications has a classification accuracy of about 86%.

Keywords — *Bloom Taxonomy, SVM Algorithm, lexical, syntactic, feature extraction.*

1. PENDAHULUAN

Proses penilaian merupakan aspek penting dalam pembelajaran. Penilaian harus dilakukan secara benar agar dapat mengukur kemampuan peserta didik dengan baik. Soal-soal yang digunakan untuk ujian pada program Studi Teknik Informatika Politeknik Kediri belum diklasifikasikan berdasar tingkat kesulitannya. Sehingga pada beberapa proses penilaian belum didasarkan atas pemberian soal-soal dengan tingkat kesulitan yang berbeda.

Pada tahun 1995, Benjamin Bloom telah memperkenalkan proses pengklasifikasian soal berdasarkan tingkat kesulitann, metode tersebut dinamakan Taksonomi Bloom. Konsep ini terbagi menjadi tiga domain/ranah kemampuan intelektual yaitu kognitif, afektif dan psikomotorik. Aspek kognitif berkaitan erat dengan pengetahuan dan pengembangan kemampuan intelektual [1]. Aspek kognitif dibagi menjadi 6 level yaitu pengetahuan, pemahaman, penerapan, analisa, sintesa dan evaluasi. Tahun 1994 Lorin Anderson Krathwohl dan para ahli psikologi aliran kognitivisme memperbaiki taksonomi bloom agar sesuai dengan kemajuan zaman [2]. Perubahan tersebut membuat level kognitif Taksonomi Bloom berubah menjadi mengingat, memahami, mengaplikasikan, menganalisis, mengevaluasi dan menciptakan.

Proses pengklasifikasian soal sesuai level pada taksonomi bloom tidaklah mudah jika dilakukan secara manual [3]. Proses otomatisasi klasifikasi perlu dilakukan ketika akan melakukan klasifikasi soal dalam jumlah yang banyak, misalkan pada proses pengklasifikasian soal pada bank soal. Berbeda dengan pengklasifikasian teks yang berfokus pada level dokumen, pengklasifikasian soal akan berfokus pada hal yang lebih sempit lagi yaitu kalimat dan kata proses tersebut memiliki tingkat kesulitan tersendiri [4]. Menurut penelitian penelitian Kusuma, Daniel dan Yuhana tahun 2015 proses klasifikasi soal berdasarkan Taksonomi Bloom dapat dilakukan dengan pengidentifikasi fitur leksikal dan sintaktik

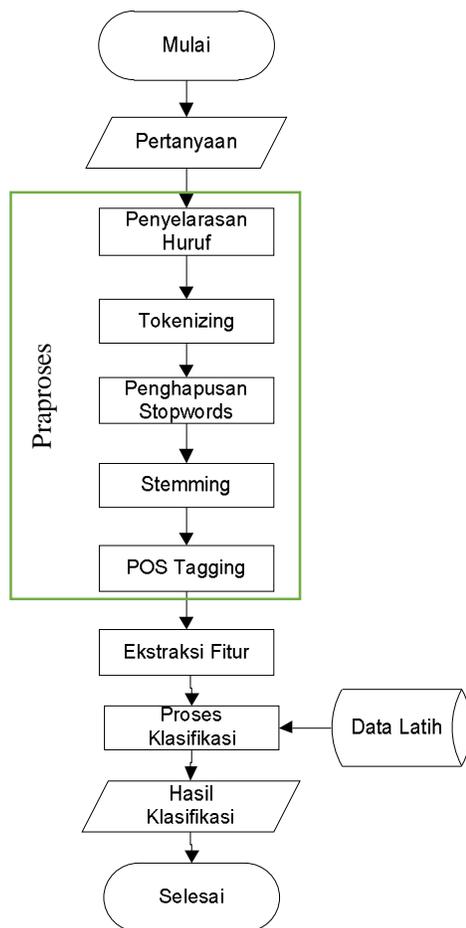
pada setiap soal [3]. Proses identifikasi tersebut digunakan sebagai proses ekstraksi fitur yang kemudian hasil ekstraksi fitur tersebut digunakan sebagai dasar proses pengklasifikasian. Penelitian terdahulu yang dijadikan dasar penelitian ini adalah penelitian Kusuma, Daniel dan Yuhana tahun 2015 yang berjudul Automatic Indonesia's Questions Classification Based On Bloom's Taxonomy Using Natural Language Processing [3]. Penelitian tersebut dipilih menjadi dasar penelitian karena memiliki jenis kesamaan dalam bahasa yang digunakan yaitu Bahasa Indonesia. Selain penelitian tersebut sebenarnya telah ada penelitian terdahulu yang membahas tentang pengklasifikasian soal berdasarkan Taksonomi Bloom namun penelitian-penelitian tersebut ditujukan untuk soal-soal yang menggunakan Bahasa Inggris. Penelitian tersebut diantaranya adalah Yahya & Osman yang mengklasifikasikan pertanyaan sesuai taksonomi bloom menggunakan Support Vector Machine [1]. Klasifikasi yang dilakukan menghasilkan akurasi 87,4%. Penelitian ini tidak menggunakan fitur semantik maupun sintaktik, akurasi yang didapatkan hanya berdasarkan Bag Of Word (BOW). Penelitian kedua adalah milik Omar yang dilakukan pada tahun 2011. Penelitian tersebut menggunakan beberapa aturan yang dibuat berdasarkan struktur soal yang digunakan untuk melakukan klasifikasi. Selain itu Omar juga melakukan pembobotan kategori untuk menangani overlapping kata kunci pada taksonomi bloom yang memungkinkan adanya kesamaan kata kunci pada level yang berbeda [5]. Penelitian ketiga adalah Haris & Omar pada tahun 2012 yang melakukan klasifikasi berdasarkan aturan yang telah dibuat. Aturan dibuat berdasarkan pola pada data latih. Hasil akurasi yang didapatkan pada penelitian Haris & Omar (2012) sebesar 77% [6]. Penelitian keempat adalah milik Abduljabbar & Omar pada tahun 2015 yang menggunakan 3 metode seleksi fitur yaitu mutual information, odd ratio and chi-square untuk meningkatkan hasil akurasi klasifikasi [7]. Penelitian tersebut merupakan perbaikan metode dari penelitian yang pernah dilakukan Haris & Omar pada tahun 2012 dengan perbaikan akurasi klasifikasi sebesar 12% .

Melalui penelitian ini diharapkan implementasi yang dilakukan dapat

digunakan untuk mengklasifikasikan soal-soal di Program Studi Teknik Informatika Politeknik Kediri berdasarkan taksonomi bloom secara otomatis sehingga proses pengklasifikasian soal dapat dilakukan secara lebih cepat.

2. METODE PENELITIAN

Tahapan metodologi penelitian pada penelitian ini meliputi pengumpulan dataset, praproses data, ekstraksi fitur, kemudian proses klasifikasi. Alur tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Metodologi penelitian

Secara lebih detail alur pada gambar 1 akan dijelaskan sebagai berikut.

2.1. Dataset

Penelitian ini menggunakan dataset yang berasal dari soal-soal Ujian Tengah Semester (UTS) dan Ujian Akhir Semester (UAS) Program Studi Teknik Informatika Politeknik Kediri. Dataset yang digunakan berjumlah 120 soal. 90 soal akan digunakan sebagai data latih dan 30 soal akan digunakan sebagai data uji. Contoh soal sesuai dengan Taksonomi Bloom ditunjukkan pada Tabel 1.

Tabel 1. Contoh Soal Sesuai Taksonomi Bloom

No	Level Taksonomi Bloom	Contoh Soal
1	Mengingat	Sebutkan 3 fungsi dari proses partisi
2	Memahami	jelaskan pola kerja AJAX
3	Menerapkan	Hitunglah waktu tunggu dan turn around data di atas dengan menggunakan algoritma First Come First Served
4	Menganalisis	Bagaimana mekanisme kode hamming mampu mendeteksi posisi bit yang berubah tersebut
5	Mengevaluasi	Apakah program yang menggunakan perulangan do while akan menjalankan programnya minimal 1 kali
6	Menciptakan	Buatlah program untuk memperjelas gambar dengan melakukan proses kontras

2.2. Praproses

Praproses adalah langkah awal sebelum melakukan klasifikasi. Terdapat 5 proses dalam lingkup praproses yaitu penyalarsan huruf, *tokenizing*, penghapusan *stopwords*, *stemming*, dan yang terakhir POS tagging. Proses penyalarsan huruf merupakan proses pengubahan huruf kapital menjadi huruf non-

kapital. Proses ini diperlukan karena proses identifikasi string biasanya bersifat case sensitive, sehingga sebaiknya semua huruf memiliki jenis yang sama. Proses tokenizing digunakan untuk menguraikan sebuah string menjadi satuan kata. Setelah string berubah menjadi satuan kata kemudian dilakukan proses penghapusan *stopwords*. Proses selanjutnya adalah stemming kata, proses ini dilakukan untuk memperoleh kata dasar dari setiap kata. Proses penghapusan stopwords dan proses stemming dapat dilakukan untuk mereduksi fitur, dan mengurangi kompleksitas dari vektor fitur. Hal ini akan meningkatkan kualitas klasifikasi [8]. Kata dasar yang telah didapatkan kemudian ditentukan jenis katanya melalui proses POS *tagging*. Contoh hasil praproses sampai tahap *stemming* ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil Praproses Sampai Tahap Stemming

No	Contoh Soal	Hasil Praproses Sampai Tahap Stemming
P1	Sebutkan 3 fungsi dari proses partisi	sebut -fungsi- proses -partisi
P2	jelaskan pola kerja AJAX	jelas -pola-kerja AJAX
P3	Hitunglah waktu tunggu dan turn around data di atas dengan menggunakan algoritma First Come First Served	hitung -tunggu-turn-around-data- guna -algoritma-First-Come-First-Served
P4	Bagaimana mekanisme kode hamming mampu mendeteksi posisi bit yang berubah tersebut	bagaimana-mekanisme- kode -hamming- deteksi -posisi-bit- ubah - sebut
P5	Apakah program yang menggunakan perulangan do while akan menjalankan programnya minimal 1 kali	apakah-program- guna - ulang -do-while-jalan-program-minimal-kali

2.3. Ekstraksi Fitur

Guna dapat melakukan klasifikasi pada sebuah kalimat soal, maka diperlukan

ekstraksi fitur. Tujuan dari ekstraksi fitur adalah mendapatkan informasi unik yang nantinya dapat digunakan untuk melakukan klasifikasi. Fitur yang digunakan pada penelitian ini adalah fitur leksikal dan sintaktik. Penggunaan kedua fitur tersebut dipilih karena penggunaan kata kunci untuk pengidentifikasian level soal ini juga belum tentu efisien karena terkadang dalam satu soal juga mengandung kata kunci ganda, misalkan pada soal “Sebutkan 3 fungsi dari **proses** partisi! ”. Soal tersebut mengandung 2 kata kunci yang berada pada 2 level yang berbeda yaitu kata sebutkan yang mengacu pada kata kunci level mengingat, dan kata proses yang mengacu pada level mengaplikasikan. Hal tersebut membuktikan bahwa apabila hanya menggunakan kata kunci saja sebagai metode untuk melakukan otomatisasi klasifikasi soal belumlah cukup.

2.3.1 Fitur Leksikal

Fitur leksikal yang digunakan pada penelitian ini diperoleh dari kata tanya yang digunakan. Proses ekstraksi fitur leksikal dilakukan dengan menghitung jumlah kata tanya yang digunakan dalam satu soal. Penggunaan jumlah kata tanya sebagai fitur karena kata tanya tersebut dianggap mampu merepresentasikan level taksonomi bloom. Misalkan pada soal “Siapakah penemu algoritma SVM?” kata tanya siapa mengacu pada taksonomi bloom level mengingat. Sedangkan pada soal “Mengapa perulangan do while akan menjalankan programnya minimal 1 kali?” kata tanya mengapa mengacu pada taksonomi bloom level menganalisis. Contoh perhitungan fitur leksikal untuk perhitungan kata tanya ditunjukkan pada Tabel 3. Perhitungan fitur tersebut berdasarkan contoh soal yang digunakan pada Tabel 1.

Tabel 3. Contoh Ekstraksi Fitur Leksikal

No	Keterangan	P1	P2	P3	P4	P5
1	Apa (<i>What</i>)	0	0	0	0	1
2	Dimana (<i>Where</i>)	0	0	0	0	0
3	Siapa (<i>Who</i>)	0	0	0	0	0
4	Kapan (<i>When</i>)	0	0	0	0	0
5		0	0	0	0	0

- 6 Mengapa (*Why*) 0 0 0 1 0
7 Bagaimana
(How)

2.3.2 Fitur Sintaktik

Fitur sintaktik didapatkan dari hasil proses ekstraksi pada kata kunci dan POS tagging. Proses ekstraksi fitur kata kunci dilakukan dengan menghitung frekuensi kemunculan kata kunci yang ada pada soal lalu memetakannya sesuai dengan letak kata kunci pada level Taksonomi Bloom. Kemudian informasi tersebut akan ditabelkan agar mudah untuk diklasifikasikan. Penggunaan banyaknya jenis kata dan perhitungan frekuensi kemunculan kata kunci digunakan sebagai fitur karena banyaknya jenis kata yang digunakan dan jumlah kata kunci dalam suatu soal dapat membantu dalam proses klasifikasi. Soal yang memiliki kemiripan akan menghasilkan pola yang cenderung sama. Contoh perhitungan kata kunci ditunjukkan pada Tabel 4, sedangkan contoh perhitungan jenis kata yang digunakan ditunjukkan pada Tabel 5. Perhitungan fitur tersebut berdasarkan contoh soal yang digunakan pada Tabel 1.

Tabel 4. Contoh Ekstraksi Fitur Kata Kunci

No	Keterangan	P1	P2	P3	P4	P5
1	Mengingat	1	1	0	2	1
2	Memahami	0	2	1	1	0
3	Menerapkan	1	0	2	0	1
4	Menganalisis	0	0	0	1	0
5	Mengevaluasi	0	1	0	0	0
6	Menciptakan	0	1	0	1	0

Tabel 5. Contoh Ekstraksi Fitur Pos Tagging

No	Keterangan	P1	P2	P3	P4	P5
1	Jumlah kata benda	3	2	2	4	5
2	Jumlah kata sifat	0	1	0	0	1
3	Jumlah kata kerja	1	0	2	2	1
4	Jumlah kata selain yang disebutkan	0	1	7	3	3

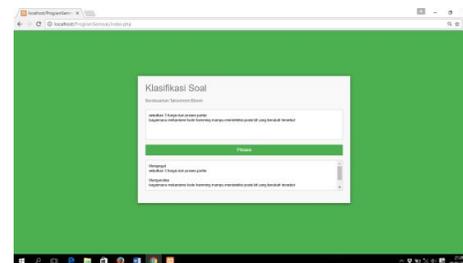
2.4. Klasifikasi

Proses klasifikasi adalah proses menentukan sebuah objek ke dalam suatu kategori yang telah ditentukan [9]. Proses klasifikasi pada penelitian ini dilakukan terhadap 120 dataset. 2/3 dari dataset tersebut akan digunakan sebagai data latih dan 1/3 nya akan digunakan sebagai data uji. Sehingga dari 120 data tersebut ada 90 data yang digunakan sebagai data latih dan 30 soal yang digunakan sebagai data uji. Proses klasifikasi pada penelitian ini dilakukan menggunakan algoritma SVM dengan teknik SMO. Algoritma tersebut dipilih karena menurut beberapa penelitian terdahulu algoritma tersebut memiliki akurasi yang tinggi [1] [3] [7]. Proses klasifikasi dilakukan menggunakan *library* weka yang digabungkan dengan aplikasi berbasis web yang telah dibuat. Guna memvalidasi keberhasilan proses klasifikasi yang telah dilakukan diperlukan sebuah proses penilaian. Proses penilaian dapat dilakukan menggunakan perhitungan akurasi. Nilai akurasi adalah proporsi jumlah prediksi data yang benar diklasifikasikan dari keseluruhan dataset. Persamaan yang digunakan untuk menghitung akurasi ditunjukkan pada persamaan 1.

$$\text{Akurasi} = \frac{\text{Data Benar}}{\text{Jumlah Data}} \quad (1)$$

3. HASIL DAN PEMBAHASAN

Implementasi proses klasifikasi dilakukan menggunakan aplikasi berbasis web. Pengguna dapat memasukkan soal yang akan diklasifikasikan pada kolom *input* soal. Setelah diproses maka hasil klasifikasi akan muncul di bagian kolom hasil. Contoh proses pengklasifikasian ditunjukkan pada Gambar 2.



Gambar 2. Tampilan Proses Klasifikasi

Proses uji coba klasifikasi dilakukan terhadap 30 soal yang digunakan sebagai data uji. Secara detail hasil klasifikasi ditunjukkan pada Tabel 6.

Tabel 6. Hasil Confusion Matrix

a	b	c	d	e	f	←---- Klasifikasi
5	0	0	0	0	0	a Mengingat
0	5	0	0	0	0	b Memahami
0	0	3	0	0	2	c Menerapkan
0	0	0	4	1	0	d Menganalisis
0	0	0	0	5	0	e Mengevaluasi
0	0	1	0	0	4	f Menciptakan

Berdasarkan informasi dari Tabel 6 diketahui bahwa ada 2 soal level mengaplikasikan diklasifikasikan sebagai level membuat, 1 soal level menganalisis yang diklasifikasikan sebagai level mengevaluasi dan 1 soal level membuat diklasifikasikan sebagai level mengaplikasikan. Sehingga secara keseluruhan ada 26 soal yang benar diklasifikasikan dan ada 4 soal yang salah diklasifikasikan. Nilai akurasi klasifikasi dihitung menggunakan persamaan 1 dan didapatkan hasil klasifikasi sebesar 86%.

4. SIMPULAN

Berdasarkan implementasi klasifikasi soal yang telah dilakukan dapat disimpulkan bahwa:

1. Proses klasifikasi soal-soal mata kuliah Teknik Informatika berdasarkan Taksonomi Bloom dapat dilakukan menggunakan fitur leksikal dan fitur sintaktik dari soal-soal tersebut.
2. Proses klasifikasi soal-soal berdasarkan Taksonomi Bloom menggunakan algoritma SVM memiliki akurasi yang tinggi dengan tingkat kebenaran 86%.

5. SARAN

Kesalahan klasifikasi kemungkinan disebabkan masih terbatasnya data latih yang digunakan sehingga pola yang dihasilkan belum mampu mengenali keseluruhan pola data uji. Jadi sebaiknya untuk penelitian kedepannya jumlah data latih bisa diperbanyak. Selain proses pengklasifikasian soal, sebaiknya kedepannya sistem ini

dikembangkan untuk dapat membangkitkan soal dari beberapa jenis pertanyaan yang telah dimasukkan. Sehingga jenis soal yang dihasilkan bisa lebih bervariasi dan proses pembuatan soal tidak dilakukan secara manual lagi.

DAFTAR PUSTAKA

- [1] A. A. Yahya and A. Osman, "Automatic Classification Of Questions Into Bloom's Cognitive Levels Using Support Vector Machines," pp. 1-6, 2011.
- [2] M. Taher, "Urgensi Taksonomi Bloom Domain Kognitif Versi Baru Dalam Kurikulum," Balai Diklat Keagamaan Medan, Medan, 2013.
- [3] S. F. Kusuma, D. Siahaan and U. L. Yuhana, "Automatic Indonesia's Questions Classification Based On Bloom's Taxonomy Using Natural Language Processing," in *International Conference on Information Technology Systems and Innovation (ICITSI)*, Bandung, 2015.
- [4] A. Sangodiah, R. Ahmad and W. F. Ahmad, "A Review in Feature Extraction Approach in Question Classification Using Support Vector Machine," *IEEE*, pp. 536-541, 2014.
- [5] N. Omar, S. S. Haris, R. Hassan, H. Arshad, M. Rahmat, N. F. A. Zainal and R. Zulkifli, "Automated analysis of exam questions according to bloom's taxonomy," *Procedia - Social and Behavioral Sciences*, pp. 297-303, 2012.
- [6] S. S. Haris and N. Omar, "A Rule-based Approach in Bloom's Taxonomy Question Classification through Natural Language Processing".
- [7] D. A. Abduljabbar and N. Omar, "Exam Questions Classification Based On Bloom's Taxonomy Cognitive Level Using Classifiers Combination," *Journal of Theoretical and Applied Information Technology*, pp. 447-455, 2015.
- [8] N. Yusof and C. J. Hui, "Determination of Bloom's Cognitive Level of Question Items Using Artificial Neural Network," *IEEE*, pp. 866-869, 2010.
- [9] S. Raharjo and E. Winarko, "Klasterisasi,