

Klasifikasi Teks Soal Ujian Berbahasa Indonesia Berdasarkan Ranah Kognitif Taksonomi Bloom

Written exam is the common type of exam to measure student's learning achievement, from Elementary School, Junior High School and Senior High School. The difficulty level of the exam may vary from one question to another, so that the exam result could be analyzed further by observing in which difficulty students can or cannot answer correctly. Bloom's Taxonomy has a cognitive domain that can be used as reference for determining the exam questions' difficulty. The cognitive domain has 6 different classes which in order of them are remember, understand, apply, analyze, evaluate, create. This final project aims to do a classification of exam questions into 6 classes of Bloom's Taxonomy cognitive field. The data used will be a text in Bahasa Indonesia from Elementary School, Junior High School and Senior High School. Methods used in this final project are Support Vector Machine and Naive Bayes, since both are proven in previous study to perform well in the same task. As for the feature extraction, this final project will be using TF-IDF that has been modified based on weight value from POS Tag. Such feature extraction method is already proven in previous study to perform better than the regular TF-IDF.

Latar Belakang

Ujian tulis dapat diberikan pada siswa dengan tujuan untuk mengukur capaian belajar siswa dan memastikan pemahaman siswa atas materi yang sudah diberikan [1]. Tingkat kesulitan pada soal dapat beragam, dan mengacu pada ranah kognitif Taksonomi Bloom [2].

Klasifikasi secara manual membutuhkan waktu yang selaras dengan jumlah soal yang ada [2]. Selain itu, rentan akan perbedaan persepsi antar pengajar yang mengakibatkan perbedaan dari hasil klasifikasi [1].

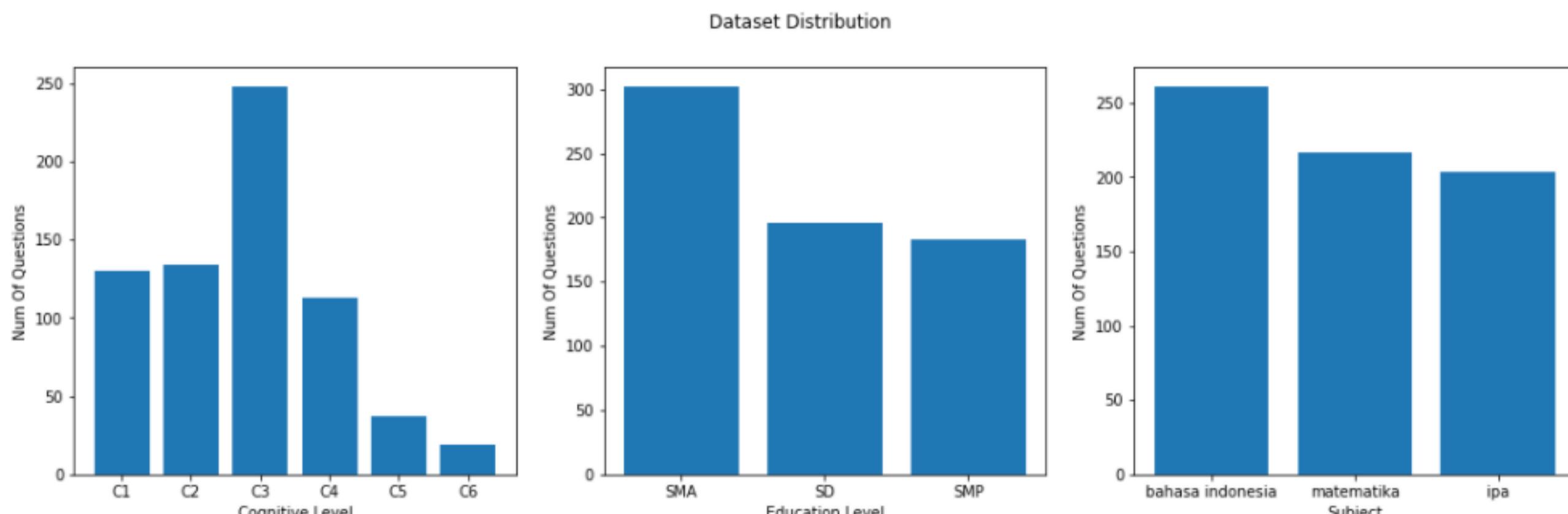
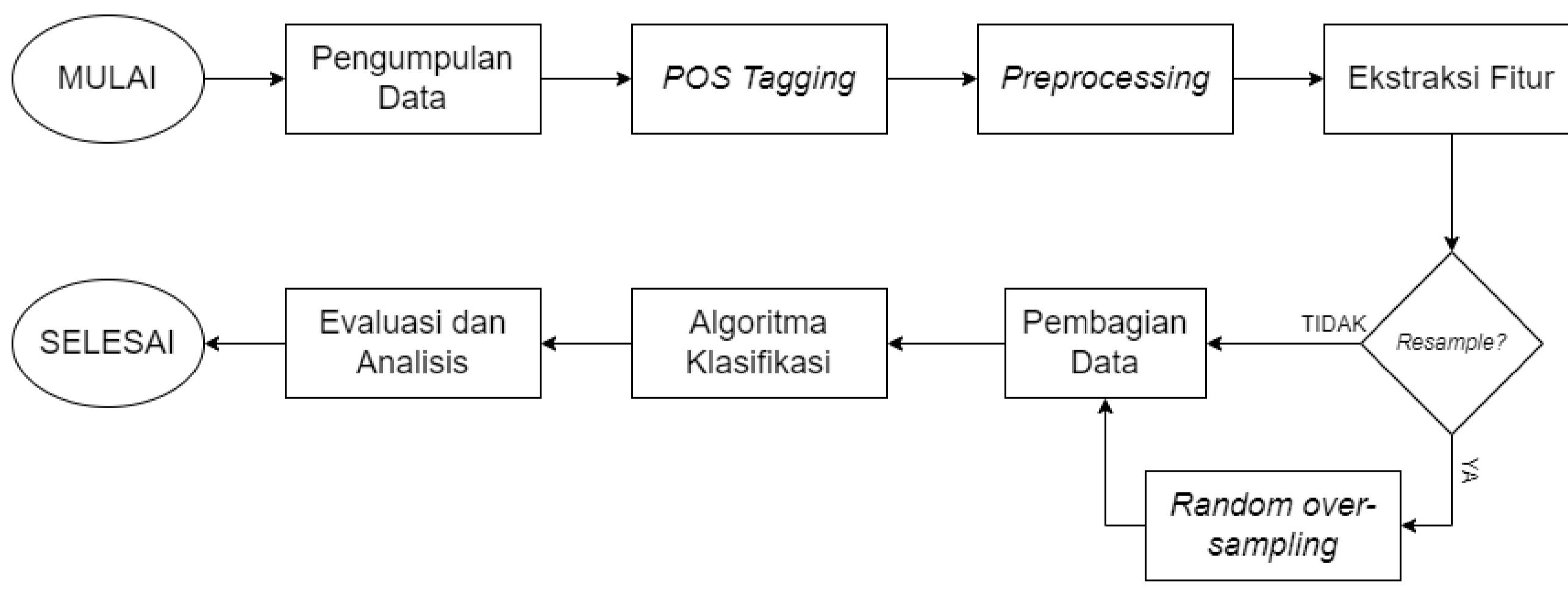
Penelitian ini melakukan klasifikasi otomatis menggunakan Support Vector Machine (SVM) dan Naive Bayes (NB). Dataset yang digunakan diambil dari mata pelajaran Bahasa Indonesia, Matematika dan Ilmu Pengetahuan. Pengambilan dataset dari mata pelajaran berbeda dilakukan untuk melihat apakah algoritma klasifikasi dapat bekerja dengan baik pada dataset dengan karakteristik yang beragam.

Topik dan Batasan Masalah

Topik yang dibahas dalam penelitian ini adalah bagaimana melakukan klasifikasi teks berdasarkan Taksonomi Bloom dengan metode SVM dan NB serta mengukur performansi dari metode yang digunakan.

Batasan masalah pada penelitian ini adalah sebagai berikut: Pertama, data yang digunakan berupa teks Berbahasa Indonesia. Kedua, mata pelajaran yang digunakan hanya Bahasa Indonesia, Matematika dan Ilmu Pengetahuan Alam dari jenjang pendidikan SD, SMP, dan SMA. Ketiga, klasifikasi yang dilakukan hanya untuk menentukan tingkatan kognitif yang sesuai dari sebuah soal berdasarkan Taksonomi Bloom.

Perancangan Sistem



Program diimplementasikan menggunakan bahasa pemrograman Python. Stopwords yang dihapus pada tahap preprocessing dimodifikasi dari daftar library PySastrawi. Dokumen pada data dilakukan POS Tagging dengan bantuan library FlairNLP. Ekstraksi fitur menggunakan TF-IDF dan TFPOS-IDF. TFPOS-IDF bekerja dengan memberikan bobot pada setiap kata berdasarkan POS Tag masing-masing [3]. Karena jumlah kelas pada dataset tidak berimbang, maka akan dilakukan random oversampling dengan strategi *not majority* sehingga semua kelas akan diduplikasi hingga jumlahnya menyamai kelas mayoritas. Data dibagi untuk pelatihan dan pengujian dengan rasio 80:20 [4]. Setelah itu, algoritma SVM dan NB akan melakukan klasifikasi dan hasilnya akan dievaluasi dan dianalisis.

Skenario	Algoritma	Feature Extraction	Stopwords	Random Over-Sampling	Akurasi & F1-Measure	Akurasi & F1-Measure optimized
1	SVM	TF-IDF	Default	N	0.453 0.429	0.445 0.438
	NB	TF-IDF	Default	N	0.423 0.421	0.445 0.434
2	SVM	TF-IDF	Modifikasi	N	0.474 0.452	0.438 0.434
	NB	TF-IDF	Modifikasi	N	0.401 0.421	0.467 0.479
3	SVM	TF-IDF	Default	Y	0.799 0.798	0.839 0.837
	NB	TF-IDF	Default	Y	0.772 0.772	0.815 0.815
4	SVM	TF-IDF	Modifikasi	Y	0.819 0.818	0.842 0.842
	NB	TF-IDF	Modifikasi	Y	0.782 0.781	0.829 0.829
5	SVM	TFPOS-IDF	Default	N	0.438 0.43	0.438 0.43
	NB	TFPOS-IDF	Default	N	0.431 0.463	0.453 0.45
6	SVM	TFPOS-IDF	Modifikasi	N	0.445 0.431	0.504 0.491
	NB	TFPOS-IDF	Modifikasi	N	0.401 0.43	0.467 0.472
7	SVM	TFPOS-IDF	Default	Y	0.815 0.814	0.836 0.836
	NB	TFPOS-IDF	Default	Y	0.735 0.732	0.795 0.793
8	SVM	TFPOS-IDF	Modifikasi	Y	0.826 0.825	0.846 0.846
	NB	TFPOS-IDF	Modifikasi	Y	0.752 0.748	0.812 0.81

Analisis Hasil Pengujian

Beberapa stopwords dapat berdampak signifikan dalam menentukan tingkat kesulitan sebuah soal [2], terbukti pada hasil pengujian dimana kedua algoritma sama-sama memiliki performa terbaik saat menggunakan stopwords versi modifikasi. Ekstraksi fitur menggunakan metode TFPOS-IDF juga secara rata-rata berdampak baik pada hasil F1-measure dibandingkan dengan metode TF-IDF.

Dataset yang melalui proses random oversampling mampu menghasilkan skor F1-measure dan akurasi yang lebih baik pada semua skenario pengujian dibandingkan data yang tidak. Hal ini dikarenakan algoritma klasifikasi dapat dilatih dengan data yang lebih banyak, sehingga dapat melakukan klasifikasi pada data pengujian secara lebih baik.

Kesimpulan

Pada penelitian ini, SVM menghasilkan performa paling baik pada skenario 8, mengungguli NB dengan performa terbaiknya pada skenario 4. Kendati demikian, kedua algoritma sama-sama memiliki performa yang baik karena skor akurasi dan F1-measure lebih mendekati 1 dibandingkan mendekati 0 [4].

[1] A. Aninditya, M. A. Hasibuan and E. Sutoyo, "Text Mining Approach Using TF-IDF and Naive Bayes for Classification of Exam Questions Based on Cognitive Level of Bloom's Taxonomy," 2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS), 2019, pp. 112-117, doi: 10.1109/IoTaIS47347.2019.8980428.

[2] S. F. Kusuma, D. Siahaan and U. L. Yuhana, "Automatic Indonesia's questions classification based on bloom's taxonomy using Natural Language Processing a preliminary study," 2015 International Conference on Information Technology Systems and Innovation (ICITSI), 2015, pp. 1-6, doi: 10.1109/ICITSI.2015.7437696.

[3] Mohammed M, Omar N (2020) Question classification based on Bloom's taxonomy cognitive domain using modified TF-IDF and word2vec. PLoS ONE 15(3): e0230442. <https://doi.org/10.1371/journal.pone.0230442>

[4] Gholamy, Afshin; Kreinovich, Vladik; and Kosheleva, Olga, "Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation" (2018). Departmental Technical Reports (CS). 1209.

Penghargaan

Penelitian ini dapat diselesaikan dengan bantuan dan bimbingan dari Dr. Ade Romadhony S.T., M.T. sebagai dosen pembimbing, serta program studi S1 Informatika, Fakultas Informatika dan Universitas Telkom yang telah mendukung dan memfasilitasi penulis dalam melaksanakan penelitian ini.