JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 1, Des 2021, hlm. 11 - 18 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i1.1185
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

PERBANDINGAN ALGORITMA KLASIFIKASI *SUPPORT VECTOR MACHINE* DAN *NAIVE BAYES* PADA *IMBALANCE* DATA

Chika Enggar Puspita[1*], Oktariani Nurul Pratiwi[1], Edi Sutoyo[1]
[1]Sistem Informasi, Telkom University
email: [*]chikapuspita@student.telkomuniversity.ac.id

Abstract: Question classification is a computer science system, which aims to analyze questions and can label each question based on existing categories. Questions can be collected from several materials or topics that are many and different. Therefore, the researcher intends to create a classification system for quiz questions Data Warehouse and Business Intelligence which can be grouped into topics Data Warehouse, Business Intelligence, Data Analytics, and Performance Measurement. One way to solve this problem is by approach machine learning. In this study, researchers used a comparison of machine learning algorithms, namely the algorithm NaïveBayes and SupportVectorMachine using SMOTE and methods Cross-Validation The results of this study show the best accuracy results and are very helpful. The results obtained in the method cross-validation before SMOTE resulted in an accuracy rate of 82.02% for the results after going through the SMOTE stage of 94.79% on the algorithm Naïve Bayes, while the algorithm SupportVectorMachine get accuracy of 81.39% in the process before SMOTE for the results after going through SMOTE of 96.52%.

Keywords: Cross-Validation; Machine Learning; Naive Bayes; Support Vector Machine; Question Classification

Abstrak: Klasifikasi pertanyaan merupakan sebuah sistem ilmu komputer, yang bertujuan untuk menganalisis pertanyaan serta dapat memberi label pada setiap pertanyaan berdasarkan kategori yang ada. Pertanyaan soal dapat dikumpulkan dari beberapa materi atau topik yang banyak dan berbeda. Oleh karena itu, bermaksud untuk membuat sistem klasifikasi pertanyaan soal kuis *Data Warehouse* dan *Business Intelligence* yang dapat dikelompokkan menjadi topik *Data Warehouse*, *Business Intelligence*, Data Analitik, dan Pengukuran Kinerja. Cara yang dapat dilakukan untuk permasalahan ini dengan menggunakan pendekatan *MachineLearning*. Pada penelitian kali ini menggunakan perbandingan algoritma *MachineLearning* yaitu algoritma *NaïveBayes* dan *SupportVectorMachine* menggunakan metode SMOTE dan *Cross*-Validation. Hasil penelitian ini menunjukkan hasil akurasi yang terbaik dan sangat membantu. Hasil yang diperoleh pada metode *cross-validation* sebelum SMOTE menghasilkan tingkat akurasi sebesar 82.02% untuk hasil sesudah melalui tahap SMOTE sebesar 94.79 % pada algoritma *Naïve Bayes*, sedangkan pada algoritma *Support Vector Machine* menghasilkan akurasi sebesar pada proses sebelum SMOTE 81.39% untuk hasil sesudah melalui SMOTE sebesar 96.52%.

Kata kunci: Klasifikasi Pertanyaan; Pembelajaran Mesin; *Naive Bayes*; *Support Vector Machine*; *Cross-Validation*

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 1, Des 2021, hlm. 11 - 18 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i1.1185
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

PENDAHULUAN

E-learning atau *Electronic-learning* merupakan konsep belajar mengajar yang dapat dilakukan oleh mahasiswa dimana saja dan kapan saja. *E-learning* sendiri juga dapat mengembangkan kemandirian mahasiswa seperti pada pemahaman materi pembelajaran melalui soal-soal kuis yang nantinya di kumpulkan menjadi bank soal [1]. Universitas yang sudah menerapkan *e-learning* salah satunya adalah Universitas Telkom dalam proses belajar mengajarnya. Salah satu program studi di Universitas Telkom yaitu S1 Sistem Informasi memiliki mata kuliah wajib yaitu *Data Warehouse* dan *Business Intelligence* (DWBI). Dari mata kuliah tersebut terdapat beberapa materi topik seperti *Business Intelligence*, *Data Warehouse*, Data Analitik, dan Pengukuran Kinerja . Dengan banyaknya topik yang dipelajari beberapa mahasiswa terkadang sulit memahami materi yang diberikan dosen Mata Kuliah tersebut. Hal tersebut membuat mahasiswa sulit untuk mencari jawaban dari soal kuis dengan topik yang berbeda-beda. Maka dari itu, dibutuhkan sistem yang dapat melakukan klasifikasi kategori topik soal secara otomatis dan sistematis. Dengan harapan sistem tersebut nantinya akan mempermudah para mahasiswa mencari jawaban soal kuis tanpa melihat semua topik yang telah diajarkan melalui *platform e-learning* [2]. Sistem ini juga diharapkan dapat mempermudah para dosen pengampu dalam mengukur mahasiswa dalam menjawab soal kuis, berdasarkan pemahaman topik atau materi yang telah diajarkan.

Banyak pendekatan yang dilakukan untuk membuat *question classification*, dan memperoleh hasil yang wajar. Salah satu cara yang dapat dilakukan dalam klasifikasi soal berdasarkan kategori topik secara otomatis yaitu dengan menggunakan pendekatan *machine learning*. Dari beberapa metode pada *machine learning* yang cocok digunakan dalam *question classification* ini yaitu *Natural Language Processing* (NLP). NLP merupakan upaya untuk mengekstrak representasi lebih lanjut dari teks bebas, tujuan dari NLP sendiri untuk membuat model perhitungan Bahasa sehingga manusia dengan komputer dapat berinteraksi melalui perantara bahasa alami [3]. NLP juga dapat digunakan konsep bahasa seperti kata benda dan kata sifat [4]. Dengan menggunakan pendekatan ini serangkaian soal kuis dimasukkan terlebih dahulu ke dalam pengklasifikasian berdasarkan topik soal kuis mata kuliah DWBI [5]. Pada metode *machine learning* terdapat beberapa algoritma, pada penelitian ini digunakan algoritma *NaïveBayes* dan *Support Vector Machine* dalam klasifikasi soal kuis berdasarkan topik. Perbandingan dilakukan dengan mengukur akurasi kedua algoritma menggunakan pendekatan *Cross Validation*. Metode ini digunakan apabila data yang dimiliki terbatas [6].

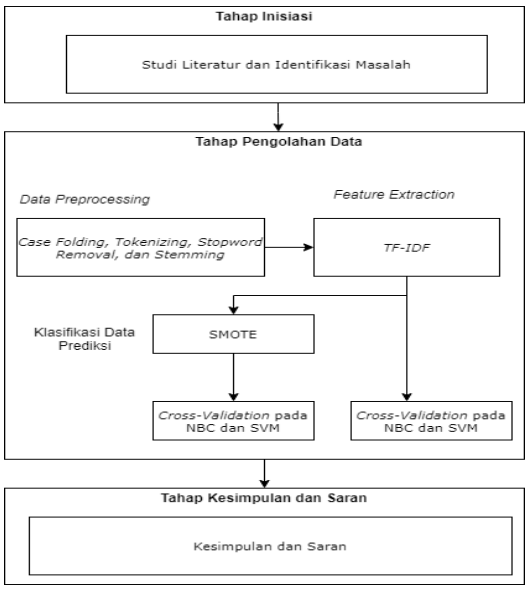
Berdasarkan penelitian jurnal yang berjudul "*Question Classification using Machine Learning Approaches"* dengan algoritma yang digunakan yaitu *Naïve Bayes Classifier* dan *Support VectorMachine*, dinyatakan bahwa pengklasifikasian menggunakan algoritma *Naïve Bayes* memiliki kesederhanaan dan sangat populer karena memiliki komputasi yang efisien dan kinerja yang sangat bagus dalam menyelesaikan masalah pada dunia nyata. Hasil percobaan pada penelitian tersebut juga menyatakan algoritma *Support Vector*

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 1, Des 2021, hlm. 11 - 18 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i1.1185
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Machine secara konsisten mempunyai kinerja yang baik pada klasifikasi pertanyaan, mengungguli algoritma *Naïve* Bayes [7]. Sedangkan pada penelitian lain yang berjudul *"Topic Classification of Islamic Question and Answer Using Naïve Bayes Classifier"* menyatakan algoritma *Naïve Bayes* dengan nilai akurasi 0,97 dapat digunakan untuk klasifikasi pertanyaan *"* [8]. Sehingga jurnal yang dibuat bermaksud membandingkan hasil klasifikasi serta akurasi pada soal kuis mata kuliah DWBI menggunakan algoritma *machine learning* yaitu *Naïve Bayes* dan *Support Vector Machine*.

METODE

Penelitian ini dikerjakan dengan menggunakan pendekatan *machine learning* dengan perhitungan akurasi algoritma menggunakan *cross-validation*. Tahapan penelitian ditunjukkan pada Gambar 1, dan selanjutnya menjelaskan masing-masing tahapan yang dilakukan.



Gambar 1 Sistematika Penulisan

Tahap yang dilakukan pertama pada penelitian ini yaitu melakukan studi literatur serta identifikasi masalah mengenai klasifikasi pertanyaan berdasarkan kategori topik soal otomatis, sehingga dapat memudahkan mahasiswa dalam menjawab soal kuis, dan memahami materi atau topik yang dipelajari [9]. Metode yang digunakan yaitu dengan membandingan hasil akurasi dari algoritma klasifikasi *machine learning*, seperti *NaïveBayes* dan *Support Vector Machine*. Selan-jutnya, peneliti melakukan pengumpulan data - data untuk dijadikan dataset dan pemberian label berdasarkan topik pada setiap soal. Data yang digunakan adalah soal kuis mata kuliah *Data Warehouse* dan *Business Intelligence*. Berkas tersebut merupakan kumpulan data kuis dari empat kelas mahasiswa S1 Sistem Informasi Telkom University Angkatan 2018 yang telah diekstrak dan disimpan kedalam *Google Drive* dengan masing-masing kelas memiliki 11 file excel yang sama, berdasarkan jumlah materi yang diberikan. Dari semua data yang sudah didapat menggabungkannya menjadi satu sehingga memperoleh jumlah soal sebanyak 3135 soal kuis. Soal kuis tersebut kemudian dilakukan pemilahan untuk menghilangkan data duplikat dan data tidak relevan, dari hasil pemilahan data tersebut diperoleh jumlah soal sebanyak 160 soal kuis. Sebelas file excel tadi diolah lagi menjadi empat topik yang lebih luas lagi seperti, *Business* Intelligence, *Data Warehouse*, Data Analitik, dan Pengukuran Kinerja yang dapat ditunjuk pada Tabel 1.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 1, Des 2021, hlm. 11 - 18 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i1.1185
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

Tabel 1. Labelling

Soal	Label
Fokusnya terhadap Enterprise/Executive IS adalah sejarah business intelligence pada tahun	*Business Intelligence*
Data terorganisir berdasarkan subjectnya yaitu pengertian karakteristik data warehouse berupa	*Data Warehouse*
Jenis KPI yang menjelaskan target diberikan kerangka waktu yang harus dipenuhi adalah	Pengukuran Kinerja
Big data challenge yang Kemampuan untuk memproses data dengan cepat, seperti yang ditangkap adalah	Data Analitik

Tahap dari pengolahan data yang selanjutnya yaitu *Data Preprocessing*. *Preprocessing* merupakan Teknik untuk mengubah teks menjadi data yang siap diolah yang bertujuan untuk menghilangkan *noise*, dan mengambil fitur penting pada sebuah teks. Tahapan yang dilakukan pertama kali dalam *Preprocessing* yaitu *Casefolding*. *Case-folding* itu merupakan langkah yang dilakukan untuk ubah semua huruf yang terdapat pada suatu dokumen teks semua huruf kecil,[10]. Setelah selesai melakukan tahap *case folding*, peneliti melakukan *tokenizing* yang merupakan proses pemecahan kalimat menjadi bagian seperti token [10]. Tahap selanjutnya yaitu *stopword removal* yang merupakan proses menghilangkan kata-kata yang tidak mengandung makna [11], [12]. Langkah terakhir pada tahapan *preprocessing* yaitu *stemming* dengan mengembalikan kata-kata ke dalam bentuk kata dasarnya dengan cara menghapus imbuhan awal, akhir, atau kedua-duanya [13].

Tahapan selanjutnya yaitu pada penghitungan bobot kata pada dokumen yaitu TF-IDF [14]. Kemudian dilakukan klasifikasi dan prediksi data dengan menerapkan algoritma *Naïve Bayes* dan *Support Vector Machine* dengan metode SMOTE dan *Cross-Validation* untuk menghitung performa kedua algoritma tersebut. SMOTE merupakan salah satu solusi yang diusulkan untuk menangani data yang tidak seimbang atau data *imbalance*. SMOTE dapat membuat data manual, kelas kecil setara dengan kelas mayor [15],[16]. Algoritma *Naïve Bayes* bertujuan untuk menemukan nilai probabilitas tertinggi pada data testing [17]. Secara umum untuk rumus teorema bayes yang digunakan

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \qquad [1]$$

Keterangan:
P(C|X) : Diberikan sebuah fakta atau rekaman X, asumsikan probabilitas Ci
P(X|C) : Menemukan nilai parameter yang memberikan probabilitas terbesar
P(C) : Probabilitas sebelumnya dari X
P(X) : jumlah tupel probabilitas kemunculan

Selain algoritma *NaïveBayes*, algoritma klasifikasi *Support Vector Machine* juga merupakan algoritma yang meraih kesuksesan besar pada berbagai masalah klasifikasi [18]. Dalam pemodelan matematisnya SVM diberi poin N dari pertanyaan dalam data training

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 1, Des 2021, hlm. 11 - 18 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i1.1185
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

$$\{x_i, y_i\}_{i=1}^{N} \quad [2]$$

di mana:
x_i= *input* ke-*i*
y_i = *output* label kelas ke-*i*

Persamaan kernel linear SVM dapat ditulis sebagai berikut [19] :

$$f(x) = sign\ [\Sigma_{i=1}^{n} \alpha_i y_i \theta(x, \chi_i) + b] \quad [3]$$

di mana:
n = nomor *support vectors*
x_i = pola pertanyaan ke-*i*
y_i = label kelas pertanyaan ke-*i*
$\theta(x, \chi_i)$ = fungsi kernel linier

Penggunaan flip terbaik adalah melakukan verifikasi 10 kali lipatan dalam model. Gambar 2 merupakan contoh proses 10-*fold* pembagian dataset menggunakan *cross-validation*. Cara Kerja *Crossvalidation* yaitu total *instance* dibagi menjadi N bagian, ketika bagian pertama data testing dilipat dan sisanya menjadi data latih, begitu seterusnya hingga *fold* ke-10 [20].

Gambar 2. *Cross-Validation*

Selanjutnya, menghitung berdasarkan keakuratan bagian ini dari data. Akurasi dihitung menggunakan persamaan:

$$Akurasi = \frac{\Sigma\ data\ testing\ klasifikasi\ benar}{\Sigma total\ data\ testing} \text{ x } 100 \quad [4]$$

Setelah mendapatkan hasil perbandingan akurasi dilakukan evaluasi model. Kesimpulan dilakukan untuk merangkum semua hasil yang sudah dibahas pada penelitian ini, serta memberikan saran sebagai bahan pertimbangan untuk evaluasi penelitian selanjutnya.

HASIL DAN PEMBAHASAN

Data *preprocessing* dilakukan melalui banyak tahap agar data tersebut dapat diterima atau digunakan oleh model. Tabel 2 merupakan hasil data *preprocessing* pada tahap pengolahan data.
Setelah melalui tahap *preprocessing* masuk kedalam TF-IDF pada tahap ini setiap kata pada dokumen diperhitungkan nilai bobotnya. Perhitungan bobot diambil dari hasil *stemming* yang dapat dilihat pada Tabel 3:

Tabel 2. *Preprocessing*

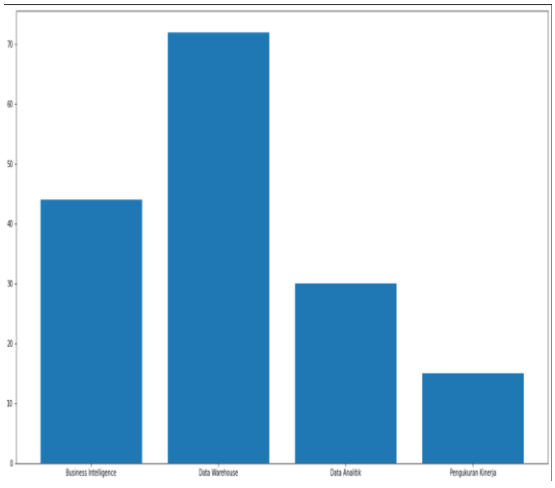
Preprocessing	Hasil *Preprocessing*
ebelum *Preprocessing*	Aktor yang bertanggung jawab atas analisis prediktif, analisis statistik, dan tools analitik yang lebih canggih serta algoritmanya adalah
Case Folding	aktor yang bertanggung jawab atas analisis prediktif, analisis statistik, dan tools analitik yang lebih canggih serta algoritmanya adalah
Tokenizing	['aktor', 'yang', 'bertanggung', 'jawab', 'atas', 'analisis', 'prediktif', 'analisis', 'statistik', 'dan', 'tools', 'analitik', 'yang', 'lebih', 'canggih', 'serta', 'algoritmanya', 'adalah']
Stopword Removal	['aktor', 'bertanggung', 'analisis', 'prediktif', 'analisis', 'statistik', 'tools', 'analitik', 'canggih', 'algoritmanya']
Stemming	['aktor', 'tanggung', 'analisis', 'prediktif', 'analisis', 'statistik', 'tools', 'analitik', 'canggih', 'algoritmanya']

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 1, Des 2021, hlm. 11 - 18 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i1.1185
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

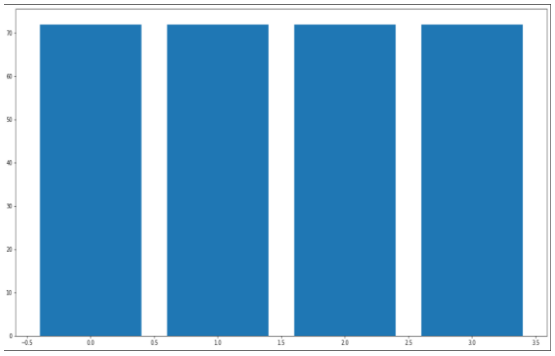
Tabel 3. TF-IDF

Preprocessing	TF-IDF
'aktor'	0.25774284640746975
'tanggung'	0.25774284640746975
'analisis'	0.44924955927307797
'prediktif'	0.25774284640746975
'statistik'	0.21396314349462084
'tools'	0.21396314349462084
'analitik'	0.22462477963653898
'canggih'	0.25774284640746975
'algoritmanya'	0.25774284640746975

Sebelum masuk ke dalam perhitungan akurasi, menyeimbangkan dataset yang tidak seimbang jumlahnya menggunakan SMOTE dimana label *Data Warehouse* berjumlah 72 soal, *Business Intelligence* berjumlah 44 soal, Data Analitik 30 soal dan terakhir pelabelan mengenai pengukuran kinerja berjumlah 15 soal, seperti yang dapat kita lihat pada Gambar 3.
Setelah dilakukan tahap SMOTE data tersebut akan menjadi seimbang seperti pada Gambar 4 yaitu dengan mengikuti kelas mayoritasnya yaitu 72 buah soal.

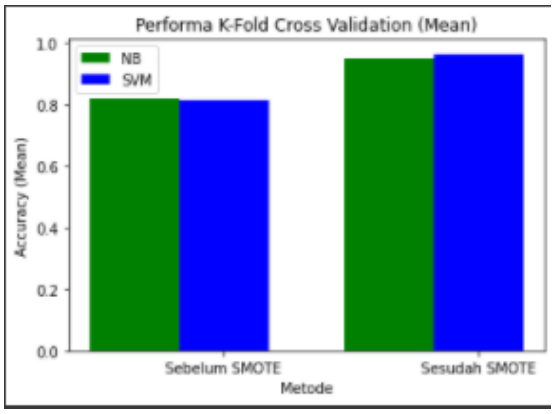


Gambar 3. Sebelum SMOTE



Gambar 4. Sesudah SMOTE

Selanjutnya, akan dilakukan perhitungan akurasi menggunakan metode *Cross Validation*. Pada tahap ini menggunakan 2 algoritma untuk mengetahui perbandingan akurasi menggunakan metode *cross-validation* ini. Gambar 5 merupakan hasil perbandingan akurasi algoritma *NaïveBayes* dan algoritma *SupportVector* menggunakan metode *cross-validation*.



Gambar 5. Perbandingan Akurasi

Dari penerapan tersebut diperoleh perbandingan rata-rata skor pada algoritma *Naïve Bayes* sebelum dilakukan SMOTE sebesar 82.02 % dan sesudah SMOTE 94.79% sedangkan pada algoritma SVM rata-rata skor sebelum SMOTE sebesar 81.39% dan sesudah SMOTE 96.52%. Berdasarkan persentase skor yang diperoleh, maka hasil skor metode *cross validation* ini masuk ke dalam klasifikasi yang baik.

JURTEKSI (Jurnal Teknologi dan Sistem Informasi) ISSN 2407-1811 (Print)
Vol. VIII No 1, Des 2021, hlm. 11 - 18 ISSN 2550-0201 (Online)
DOI: https://doi.org/10.33330/jurteksi.v8i1.1185
Available online at http://jurnal.stmikroyal.ac.id/index.php/jurteksi

SIMPULAN

Klasifikasi soal ini memberikan manfaat bagi Prodi S1 Sistem Informasi Telkom University dalam membantu pengambilan keputusan yang tepat dalam menentukan jenis soal berdasarkan kategori topik secara otomatis khususnya pada mata kuliah *Data Warehouse* dan *Business Intelligence*. Kesimpulan di dapat dari hasil analisis dimana pendekatan algoritma *NaïveBayes* dan algoritma *SupportVectorMachine* menggunakan metode SMOTE dan *Cross-Validation* melakukan kinerja yang bagus dalam klasifikasi soal berdasarkan kategori topik.

DAFTAR PUSTAKA

[1] Suharyanto and adele B. L. Mailangkay, "Penerapan E-Learning Sebagai Alat Bantu Mengajar Dalam Dunia Pendidikan," *J. Ilm. Widya*, vol. 3, pp. 17–21, 2016, doi: 10.1016/j.neubiorev.2016.02.001.

[2] G. Tika and Adiwijaya, "Klasifikasi Topik Berita Berbahasa Indonesia Menggunakan Multilayer Perceptron," *e-Proceeding Eng.*, vol. 6, no. 2, p. 2137, 2019.

[3] N. K. Wangsanegara and B. Subaeki, "IMPLEMENTASI NATURAL LANGUAGE PROCESSING DALAM PENGUKURAN KETEPATAN EJAAN YANG DISEMPURNAKAN (EYD) PADA ABSTRAK SKRIPSI MENGGUNAKAN ALGORITMA FUZZY LOGIC," *J. Tek. Inform.*, vol. 8, no. 2, 2015, doi: 10.15408/jti.v8i2.3185.

[4] 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," in *2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, 2019, pp. 112–117.

[5] 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," in *2015 International Conference on Information Technology Systems and Innovation, ICITSI 2015 - Proceedings*, 2016, doi: 10.1109/ICITSI.2015.7437696.

[6] H. SITEFANUS, "ANALISIS KINERJA METODE CROSS VALIDATION DAN K-NEAREST NEIGHBOR DALAM KLASIFIKASI DATA," pp. 7–37, 2020.

[7] A. DPanicker, A. U, and S. Venkitakrishnan, "Question Classification using Machine Learning Approaches," *Int. J. Comput. Appl.*, vol. 48, no. 13, pp. 1–4, 2017, doi: 10.5120/7405-0101.

[8] N. F. Hardifa and K. M. Lhaksmana, "Topic Classification of Islamic Question and Answer Using Naive Bayes Classifier," vol. 4, no. August, pp. 199–204, 2019, doi: 10.21108/indojc.2019.4.2.346.

[9] A. Anika, M. H. Rahman, S. Islam, A. S. Mohammad Mahdee Jameel, and C. R. Rahman, "A Comprehensive Comparison of Machine Learning Based Methods Used in Bengali Question

Classification," *2019 IEEE Int. Conf. Signal Process. Information, Commun. Syst. SPICSCON 2019*, pp. 82–85, 2019, doi: 10.1109/SPICSCON48833.2019.9065107.

[10] D. Juang, "Analisis Spam dengan Menggunakan Naïve Bayes," *J. Teknovasi*, vol. 3, no. 2, pp. 51–57, 2016.

[11] I. R. Vanani, "Text analytics of customers on twitter: Brand sentiments in customer support," *J. Inf. Technol. Manag.*, vol. 11, no. 2, pp. 43–58, 2019, doi: 10.22059/JITM.2019.291087.2410.

[12] E. Sutoyo and A. Almaarif, "Twitter sentiment analysis of the relocation of Indonesia's capital city," *Bull. Electr. Eng. Informatics*, vol. 9, no. 4, pp. 1620–1630, 2020.

[13] D. Rustiana and N. Rahayu, *Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes*, vol. 8, no. 1. 2017.

[14] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online d engan menggunakan Pembobotan TF-IDF dan Cosine Similarity," vol. 2, no. 1, pp. 306–312, 2018.

[15] S. Maldonado, J. López, and C. Vairetti, "An alternative SMOTE oversampling strategy for high-dimensional datasets," *Appl. Soft Comput. J.*, vol. 76, pp. 380–389, 2019, doi: 10.1016/j.asoc.2018.12.024.

[16] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 6, no. 3, pp. 379–385.

[17] M. M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, no. 2, pp. 88–91, Jun. 2019, doi: 10.18201//ijisae.2019252786.

[18] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliansyah, "Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi," *JUITA J. Inform.*, vol. 7, no. 2, p. 71, 2019, doi: 10.30595/juita.v7i2.4378.

[19] L. Demidova, E. Nikulchev, and Y. Sokolova, "Big Data Classification Using the SVM Classifiers with the Modified Particle Swarm Optimization and the SVM Ensembles," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 5, pp. 294–312, 2016, doi: 10.14569/ijacsa.2016.070541.

[20] O. Ghorbanzadeh, H. Rostamzadeh, T. Blaschke, K. Gholaminia, and J. Aryal, "A new GIS-based data mining technique using an adaptive neuro-fuzzy inference system (ANFIS) and k-fold cross-validation approach for land subsidence susceptibility mapping," *Nat. Hazards*, vol. 94, no. 2, pp. 497–517, 2018, doi: 10.1007/s11069-018-3449-y.