Jurnal Komputer dan Informatika
Vol. 16 No. 2, Oktober 2021: hlm 84 - 91

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

KLASIFIKASI KATEGORI WEBSITE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SVM

Widya Harum Wulandari
Program Studi Teknik Informatika, FakultasTeknologi Informasi, Universitas Tarumanagara
Jl. Letjen S.Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410
e-mail: *widya.535210100@stu.untar.ac.id*

ABSTRAK

Seiring ledakan informasi pada abad ke-21, kelimpahan informasi menjadi tantangan utama di internet. Meskipun *search engine* membantu pengguna menilai nilai situs web berdasarkan topiknya, menemukan informasi spesifik tetap menantang. Penelitian ini mengusulkan pendekatan praktis dengan membuat catatan data otomatis yang merangkum konten setiap situs web untuk tujuan kategorisasi. Tugas klasifikasi teks (TC) secara otomatis menetapkan dokumen ke kategori tertentu. Penelitian ini juga berfokus pada metode klasifikasi teks, termasuk *Naïve Bayes* dan *Support Vector Machine* (SVM), untuk mencapai akurasi yang baik. Selain dua metode tersebut, penelitian ini merinci penggunaan metode lain seperti *K-Nearest Neighbors* (KNN) dan *Random Forest* dalam konteks klasifikasi *web phishing* dan analisis sentimen. Dalam eksperimen ini, *Naïve Bayes* dan SVM dievaluasi untuk mengklasifikasikan kategori situs *web*. Data dibagi menjadi pelatihan (70%) dan pengujian (30%) dengan hasil akurasi sekitar 88% untuk *Naïve Bayes* dan 85% untuk SVM. Studi ini memberikan pemahaman lebih dalam tentang kinerja dua metode klasifikasi berbeda dalam konteks pengkategorian situs web.

Kata kunci: Klasifikasi Teks, Naïve Bayes, Support Vector Machine, Kategorisasi, Pembelajaran Mesin.

ABSTRACT

With the explosion of information in the 21st century, information overload has become a major challenge on the internet. Although search engines help users assess the value of websites based on their topics, finding specific information remains challenging. This study proposes a practical approach by creating automatic data records that summarize the content of each website for categorization purposes. Text classification (TC) tasks automatically assign documents to specific categories. This study also focuses on text classification methods, including Naïve Bayes and Support Vector Machine (SVM), to achieve good accuracy. In addition to these two methods, this study details the use of other methods such as K-Nearest Neighbors (KNN) and Random Forest in the context of web phishing classification and sentiment analysis. In this experiment, Naïve Bayes and SVM were evaluated for classifying website categories. The data was divided into training (70%) and testing (30%) with accuracy results of around 88% for Naïve Bayes and 85% for SVM. This study provides a deeper understanding of the performance of two different classification methods in the context of website categorization.

Keywords: *Text Classification, Naïve Bayes, Support Vector Machine, Categorization, Machine Learning.*

1. PENDAHULUAN

Klasifikasi situs web memainkan peran penting untuk berbagai tujuan, mulai dari analisis *Web*, *search engine*, hingga keamanan *Web*. Sejak abad ke-21, telah terjadi ledakan besar informasi, informasi yang berlebihan telah menjadi masalah paling parah di Internet. *Search engine* mengkategorikan situs *web* menurut subjek atau topik membantu pengguna untuk menilai situs *web* dan dengan demikian memperoleh informasi yang relevan dengan lebih mudah. Namun, informasi yang menarik untuk kategorisasi spesifik harus dicari di antara jumlah yang sangat besar [1]. Penelitian ini mengusulkan prosedur yang praktis untuk melakukan kategorisasi situs *web* berdasarkan pembuatan catatan data secara otomatis yang meringkas konten setiap situs web.

Text Classification (TC) adalah tugas untuk secara otomatis menetapkan dokumen ke dalam sejumlah kategori tertentu. Metode klasifikasi teks memiliki tujuan yang sama yaitu untuk menentukan label yang telah ditentukan untuk teks input yang diberikan, meskipun denominasi ini dapat merujuk ke berbagai metode khusus yang diterapkan pada domain yang berbeda [2]. Metode *Machine Learning* untuk klasifikasi teks termasuk *Naïve Bayes* sebagai metode probabilistik, mengandalkan prinsip probabilitas dan independensi kondisional untuk mengklasifikasikan teks dan *Support Vector Machine* (SVM) menggunakan pendekatan geometris dengan mencari batas keputusan optimal yang memisahkan instance dari kelas yang berbeda [3].

Pada penelitian ini digunakan 2 metode berbeda seperti yang telah disebutkan sebelumnya. Dalam konteks ini, metode klasifikasi digunakan untuk mengetahui bagaimana agar metode yang digunakan mendapatkan nilai akurasi yang baik berdasarkan dataset yang diperoleh. Untuk pelatihan dan pengujian dibagi menjadi data pelatihan (70% dari total data) dan data uji (30% dari total data).

2. TINJAUAN LITERATUR

Selain dua metode yang disebutkan, ada beberapa metode klasifikasi lainnya. Penelitian oleh [4] melakukan Analisis sentimen dan klasifikasi protes petani India menggunakan data *twitter* dengan menggunakan metode *Random Forest*. Berbagai metode dalam *data mining* dan *neural networks* telah digunakan untuk klasifikasi kategori [5]. *Naive Bayes* (NB) telah banyak digunakan untuk menganalisis dan memecahkan banyak masalah ilmiah dan teknik, seperti klasifikasi teks [6]. *Naive Bayes* merepresentasikan sebuah dokumen dengan menggunakan vektor variabel fitur biner, yang menunjukkan apakah setiap kata muncul dalam dokumen dan, dengan demikian, mengabaikan informasi frekuensi setiap kata yang muncul dalam dokumen [7]. Mempelajari NB yang optimal telah diverifikasi sebagai masalah yang khas. Oleh karena itu, untuk mempelajari jaringan *Bayesian* dengan mudah, banyak peningkatan pada NB telah diusulkan [8]. Sedangkan, untuk SVM sendiri telah dikenal sebagai salah satu metode klasifikasi yang paling sukses untuk banyak aplikasi, termasuk klasifikasi teks [9]. Mereka terlihat lebih dapat ditafsirkan daripada Deep Neural Network, dan telah banyak digunakan dalam banyak klasifikasi seperti pengenalan digit tulisan tangan, pengenalan objek, dan klasifikasi teks [10].

Kategorisasi adalah operasi memilah-milah dokumen teks ke dalam kategori teks yang telah ditentukan sebelumnya menggunakan beberapa algoritme pembelajaran mesin. Biasanya, ini mendefinisikan pendekatan yang paling penting untuk mengatur dan memanfaatkan sejumlah besar informasi yang ada dalam bentuk yang tidak terstruktur [11]. Kategorisasi teks otomatis secara konstan menjadi area penelitian dan aplikasi yang penting karena fondasi untuk makalah digital. Teks juga akan ditulis untuk kategori yang tak terhitung jumlahnya, seperti: iklan, laporan berita, artikel ilmiah, dan ulasan film [12]. pengklasifikasi terbaik dipilih setelah evaluasi menyeluruh terhadap kinerja sejumlah metode pembelajaran konvensional dan deep learning yang berbeda

Dalam penelitian oleh [13] memfokuskan pada pengaruh teknik preprocessing terhadap kinerja tiga algoritma pembelajaran mesin, yaitu *Naïve Bayesian*, DMNBtext, dan C4.5. Beberapa teknik *preprocessing* dievaluasi pada korpus teks Arab menggunakan algoritma pembelajaran mesin. Hasilnya menunjukkan bahwa algoritma pembelajaran DMNBtext memiliki kinerja yang lebih unggul dibandingkan dengan algoritma pembelajaran mesin lainnya [20]

3. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah metode eksperimen yang melibatkan manipulasi variabel-variabel tertentu untuk mengamati efeknya, dalam konteks data dan pembagian data uji-latih, termasuk variasi dalam cara data dibagi dan digunakan untuk melatih dan menguji model. Dengan pendekatan kuantitatif sebagai pendekatan penelitian, karena data diukur

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

dan dinyatakan dalam bentuk numerik. Nilai akurasi, presisi, recall, dan metrik evaluasi lainnya adalah ukuran numerik yang menggambarkan kinerja model atau eksperimen. Pengumpulan dataset dari dataset pada *platform Kaggle* dengan format *Comma Separated Value* (CSV) yang terdiri dari 1408 data.

Data CSV adalah cara standar untuk bertukar dan mengonversi data antara berbagai aplikasi terkait aplikasi. Sederhana, mudah dipahami oleh manusia dan mesin, dan selaras dengan baik dengan tipe data tabular data maksimum [14]. Banyak ahli yang mempertanyakan konsep, logika, dan penerapan CSV. Secara khusus, mereka menunjukkan bahwa CSV telah gagal menyajikan pendekatan atau kerangka kerja yang konkret untuk kepentingan atau nilai yang bertentangan, dan lain-lain [15].

3.1 Dataset

Ketika machine learning beralih ke pendekatan dengan kebutuhan data yang lebih besar dalam dekade terakhir, jenis anotasi terampil dan metodis yang diterapkan dalam praktik pengumpulan kumpulan data di era sebelumnya ditolak karena dianggap "lambat dan mahal untuk diakuisisi", dan peralihan ke pengumpulan data yang semakin banyak dan tak terkekang dari web, bersamaan dengan meningkatnya ketergantungan pada pekerja kerumunan (*crowdworkers*), dipandang sebagai keuntungan bagi pembelajaran mesin. Dataset teks berisi berbagai atribut seperti deskripsi bidang, URL, judul artikel, dan konten [16].

Dataset yang digunakan dalam penelitian ini diperoleh melalui penelusuran berbagai situs web dan telah diambil dari *platform Kaggle*. Proses pengumpulan data melibatkan ekstraksi informasi dari berbagai sumber online, dan setelahnya, teks yang dihasilkan diklasifikasikan ke dalam berbagai kategori. Metode klasifikasi ini dapat melibatkan teknik-teknik seperti *machine learning* atau pemrosesan bahasa alami untuk mengidentifikasi dan memisahkan data ke dalam kelompok-kelompok yang sesuai. Gambar 1 menunjukkan contoh dataset yang sudah diunduh.

	Unnamed: 0	website_url	cleaned_website_text	Category
0	0	https://www.booking.com/index.html?aid=1743217	official site good hotel accommodation big sav...	Travel
1	1	https://travelsites.com/expedia/	expedia hotel book sites like use vacation wor...	Travel
2	2	https://travelsites.com/tripadvisor/	tripadvisor hotel book sites like previously d...	Travel
3	3	https://www.momondo.in/?ispredir=true	cheap flights search compare flights momondo f...	Travel
4	4	https://www.ebookers.com/?AFFCID=EBOOKERS.UK.n...	bot create free account create free account si...	Travel

Gambar 1. Contoh Dataset

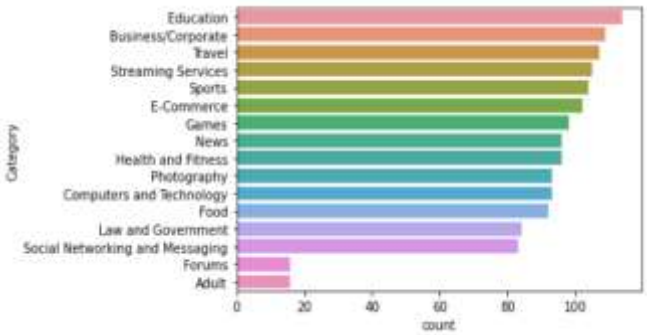
3.2 Pra-pemrosesan Data

Preprocessing merupakan tahap kritis dalam analisis data teks yang dilakukan untuk mengubah data teks yang tidak terstruktur menjadi terstruktur. Penelitian ini hanya melakukan satu proses, yaitu *Tokenizing*. Penting untuk diingat bahwa pada kolom "*cleaned_website_text*," telah dilakukan proses preprocessing sebelumnya. Proses tersebut mungkin melibatkan langkah-langkah seperti penghapusan karakter khusus, konversi teks menjadi huruf kecil, atau penghapusan *stop words*. Pada tahap *Tokenizing* karakter seperti tanda baca dihilangkan dan karakter spasi berfungsi untuk memotong kalimat menjadi gabungan kata. Hasil dari proses ini dapat dilihat contohnya seperti pada Gambar 2.

```
0    [official, site, good, hotel, accommodation, b...
1    [expedia, hotel, book, sites, like, use, vacat...
2    [tripadvisor, hotel, book, sites, like, previo...
3    [cheap, flights, search, compare, flights, mom...
4    [bot, create, free, account, create, free, acc...
Name: cleaned_website_text, dtype: object
```

Gambar 2. Contoh Hasil *Tokenizing*

Tahap berikutnya adalah pembagian kategori hasil *tokenizing*. Pada tahapan ini didapatkan hasil yang bisa dilihat pada Gambar 3 berikut ini.



Gambar 3. Kategorisasi Dataset

Dari gambar di atas dapat disimpulkan bahwa kategori yang paling banyak adalah *Education Website* lalu diikuti oleh *Corporate Website, Travel, Streaming Services,* dan sebagainya. Gambar 4 berikut ini menunjukkan jumlah kategori yang dapat terhitung berdasarkan kategorisasi dataset.

Category	website_url	cleaned_website_text
Adult	16	16
Business/Corporate	109	109
Computers and Technology	93	93
E-Commerce	102	102
Education	114	114
Food	92	92
Forums	16	16
Games	98	98
Health and Fitness	96	96
Law and Government	84	84
News	96	96
Photography	93	93
Social Networking and Messaging	83	83
Sports	104	104
Streaming Services	105	105
Travel	107	107

Gambar 4. Jumlah Kategori Terhitung

4. HASIL DAN PEMBAHASAN

Setelah semua tahapan dilakukan, peneliti mengklasifikasikan data dengan menggunakan dua metode *machine learning* yang cukup umum digunakan dalam proses klasifikasi teks, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM). Kemudian, nilai akurasi dari kedua metode tersebut akan dibandingkan. Akurasi adalah metrik evaluasi yang mengukur seberapa baik model dapat mengklasifikasi data dengan benar. Secara umum, akurasi dihitung sebagai rasio prediksi benar (*true predictions*) dibagi dengan keseluruhan jumlah data yang dievaluasi.

Namun, perlu diingat bahwa akurasi mungkin tidak memberikan gambaran yang lengkap jika data tidak seimbang (*imbalance*) atau jika terdapat kelas yang lebih dominan. Oleh karena itu, untuk tugas klasifikasi yang kompleks, seringkali disarankan untuk melihat metrik evaluasi tambahan seperti *precision, recall, dan F1-score*. Berikut merupakan hasil perhitungan nilai akurasi menggunakan dua metode berbeda.

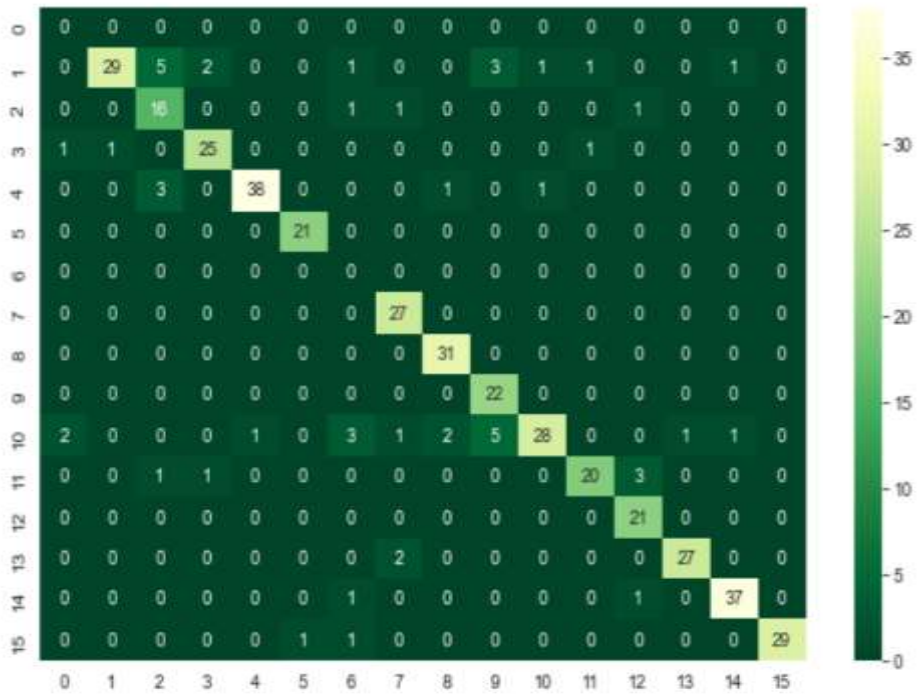
ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

Gambar 5 menyajikan nilai akurasi dari metode Naïve Bayes menggunakan ekstraksi fitur *Bag of Words* (*BoW*). Berdasarkan nilai yang tertera, metode *Naïve Bayes* memiliki nilai akurasi sekitar 88%. Hasil tersebut didapat dengan membagi dataset menjadi data latih dan data uji dalam perbandingan 7:3. Pengukuran akurasi ini memberikan gambaran tentang seberapa baik model dapat mengklasifikasikan data dengan benar. Tingkat akurasi sebesar 88% mencerminkan kemampuan model *Naïve Bayes* dalam melakukan klasifikasi dengan baik berdasarkan fitur yang diekstraksi menggunakan metode *BoW*. Pada Gambar 6 disajikan bentuk *heatmap* yang diperoleh dengan metode ini.

	precision	recall	f1-score	support
Adult	0.00	0.00	0.00	0
Business/Corporate	0.97	0.67	0.79	43
Computers and Technology	0.64	0.84	0.73	19
E-Commerce	0.89	0.89	0.89	28
Education	0.97	0.88	0.93	43
Food	0.95	1.00	0.98	21
Forums	0.00	0.00	0.00	0
Games	0.87	1.00	0.93	27
Health and Fitness	0.91	1.00	0.95	31
Law and Government	0.73	1.00	0.85	22
News	0.93	0.64	0.76	44
Photography	0.91	0.80	0.85	25
Social Networking and Messaging	0.81	1.00	0.89	21
Sports	0.96	0.93	0.95	29
Streaming Services	0.95	0.95	0.95	39
Travel	1.00	0.94	0.97	31
accuracy			0.88	423
macro avg	0.78	0.78	0.78	423
weighted avg	0.91	0.88	0.89	423

Accuracy Score: 0.8770685579196218

Gambar 5. Hasil Pengukuran Menggunakan Metode *Naive Bayes*



Gambar 6. *Heatmap* Metode *Naive Bayes*

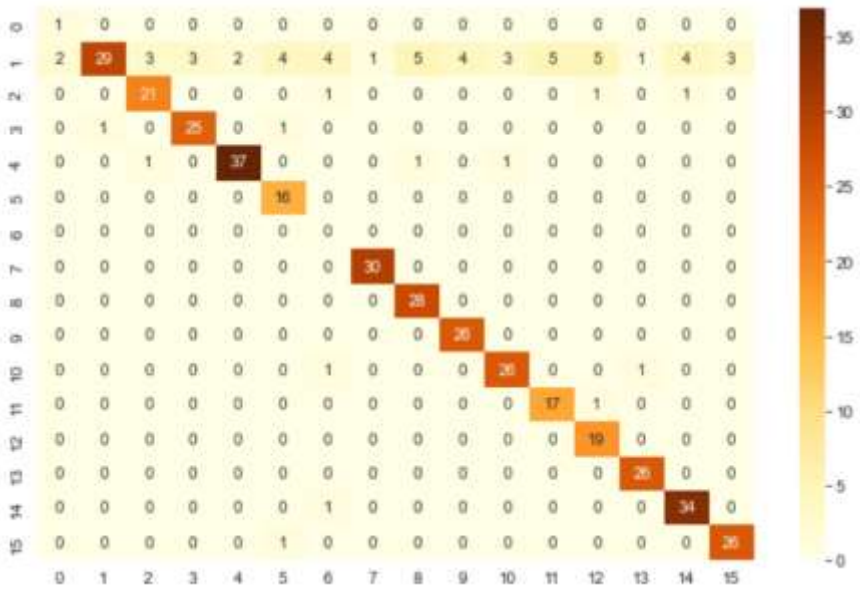
Gambar 7 menyajikan nilai akurasi dari metode SVM menggunakan ekstraksi fitur *Bag of Words* (BoW). Berdasarkan nilai yang tertera pada Gambar 7, metode SVM memiliki nilai akurasi

sekitar 85%. Hasil tersebut didapat dengan membagi dataset menjadi data latih dan data uji dalam perbandingan 7:3. Pengukuran akurasi ini memberikan gambaran tentang seberapa baik model dapat mengklasifikasikan data dengan benar. Tingkat akurasi sebesar 85% mencerminkan kemampuan model SVM dalam melakukan klasifikasi dengan baik berdasarkan fitur yang diekstraksi menggunakan metode BoW. Pada Gambar 8 disajikan bentuk *heatmap* yang diperoleh dengan metode ini.

	precision	recall	f1-score	support
Adult	0.33	1.00	0.50	1
Business/Corporate	0.97	0.37	0.54	78
Computers and Technology	0.84	0.88	0.86	24
E-Commerce	0.89	0.93	0.91	27
Education	0.95	0.93	0.94	40
Food	0.73	1.00	0.84	16
Forums	0.00	0.00	0.00	0
Games	0.97	1.00	0.98	30
Health and Fitness	0.82	1.00	0.90	28
Law and Government	0.87	1.00	0.93	26
News	0.87	0.93	0.90	28
Photography	0.77	0.94	0.85	18
Social Networking and Messaging	0.73	1.00	0.84	19
Sports	0.93	1.00	0.96	26
Streaming Services	0.87	0.97	0.92	35
Travel	0.90	0.96	0.93	27
accuracy			0.85	423
macro avg	0.78	0.87	0.80	423
weighted avg	0.89	0.85	0.84	423

Accuracy Score: 0.8534278959810875

Gambar 7. Hasil Pengukuran Menggunakan Metode SVM



Gambar 8. *Heatmap* Metode SVM

Berdasarkan hasil yang tercatat pada Gambar 5 dan Gambar 7, dapat diambil kesimpulan bahwa metode *Naïve Bayes* dengan ekstraksi fitur *Bag of Words* (*BoW*) menunjukkan kinerja yang lebih baik dibandingkan dengan metode SVM pada dataset yang digunakan. Metode *Naïve Bayes* mencapai tingkat akurasi sekitar 88%, sedangkan metode SVM mencapai tingkat akurasi sekitar 85%, saat keduanya diterapkan dengan pembagian dataset 7:3 antara data latih dan data uji.

ISSN 1410-7228 (Versi Cetak)
ISSN 2716-4233 (Versi Elektronik)

5. KESIMPULAN

Secara keseluruhan, *Text Classification* (TC) diidentifikasi sebagai tugas penting dalam dunia pemrosesan bahasa alami, di mana tujuannya adalah secara otomatis menetapkan dokumen ke dalam kategori tertentu. Metode klasifikasi teks, seperti *Naïve Bayes* dan *Support Vector Machine* (SVM), memainkan peran utama dalam mencapai tujuan ini dengan menentukan label yang telah ditentukan untuk teks input. Dalam kerangka penelitian ini, fokus ditempatkan pada dua metode klasifikasi utama, yaitu *Naïve Bayes* dan SVM. Penggunaan metode ini dalam konteks klasifikasi teks bertujuan untuk mengevaluasi kinerja keduanya berdasarkan dataset yang digunakan. Pembagian data untuk pelatihan dan pengujian dilakukan dengan membagi 70% dari total data untuk pelatihan dan 30% untuk pengujian. Dengan parameter pembagian data tersebut, *Naïve Bayes* mencapai tingkat akurasi sekitar 88%, sedangkan SVM mencapai tingkat akurasi sekitar 85%. Berdasarkan evaluasi akurasi yang diperoleh menunjukkan bahwa *Naïve Bayes* memiliki kinerja lebih baik dalam kasus ini. Hasil ini dapat memberikan panduan berharga untuk pemilihan metode klasifikasi teks yang optimal dalam skenario tertentu, serta memberikan pemahaman lebih lanjut tentang sejauh mana dua metode ini dapat diandalkan dalam tugas klasifikasi teks.

DAFTAR PUSAKA

[1] O. W. Kwon and J. H. Lee, "Text categorization based on k-nearest neighbor approach for Web site classification," Inf Process Manag, vol. 39, no. 1, pp. 25–44, Jan. 2003, doi: 10.1016/S0306-4573(02)00022-5.

[2] R. Bruni and G. Bianchi, "Robustness analysis of a Website categorization procedure based on Machine Learning," Dep. of Computer Control and Management Engineering, Sapienza University of Rome, Italy, pp. 1–25, 2018, [Online]. Available: http://www.dis.uniroma1.it/%7B~%7Dbruni/files/bruni18robustness.pdf

[3] R. Bruni and G. Bianchi, "Website categorization: A formal approach and robustness analysis in the case of e-commerce detection," Expert Syst Appl, vol. 142, Mar. 2020, doi: 10.1016/j.eswa.2019.113001.

[4] A. S. Neogi, K. A. Garg, R. K. Mishra, and Y. K. Dwivedi, "Sentiment analysis and classification of Indian farmers' protest using twitter data," International Journal of Information Management Data Insights, vol. 1, no. 2, Nov. 2021, doi: 10.1016/j.jjimei.2021.100019.

[5] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, "The influence of preprocessing on text classification using a bag-of-words representation," PLoS One, vol. 15, no. 5, 2020, doi: 10.1371/journal.pone.0232525.

[6] L. Chen, L. Jiang, and C. Li, "Modified DFS-based term weighting scheme for text classification," Expert Syst Appl, vol. 168, 2021, doi: 10.1016/j.eswa.2020.114438.

[7] S. Gan, S. Shao, L. Chen, L. Yu, and L. Jiang, "Adapting hidden naive bayes for text classification," Mathematics, vol. 9, no. 19, Oct. 2021, doi: 10.3390/math9192378.

[8] H. Zhang, L. Jiang, and L. Yu, "Attribute and instance weighted naive Bayes," Pattern Recognit, vol. 111, 2021, doi: 10.1016/j.patcog.2020.107674.

[9] H. Kim, P. Rowland, and H. Park, "Dimension reduction in text classification with support vector machines," Journal of Machine Learning Research, vol. 6, 2005.

[10] A. Wibowo Haryanto, E. Kholid Mawardi, and Muljono, "Influence of Word Normalization and Chi-Squared Feature Selection on Support Vector Machine (SVM) Text Classification," in Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, iSemantic 2018, Institute of Electrical and Electronics Engineers Inc., Nov. 2018, pp. 229–233. doi: 10.1109/ISEMANTIC.2018.8549748.

[11] A. Dhar, H. Mukherjee, N. S. Dash, and K. Roy, "Text categorization: past and present," Artif Intell Rev, vol. 54, no. 4, 2021, doi: 10.1007/s10462-020-09919-1.

[12] B. P. Yadav, S. Ghate, A. Harshavardhan, G. Jhansi, K. S. Kumar, and E. Sudarshan, "Text categorization Performance examination Using Machine Learning Algorithms," in IOP Conference Series: Materials Science and Engineering, 2020. doi: 10.1088/1757-899X/981/2/022044.

[13] R. Alshammari, "Arabic Text categorization using machine learning approaches," International Journal of Advanced Computer Science and Applications, vol. 9, no. 3, 2018, doi: 10.14569/IJACSA.2018.090332.

[14] S. M. H. Mahmud, M. A. Hossin, H. Jahan, S. R. H. Noori, and T. Bhuiyan, "CSV-ANNOTATE: Generate annotated tables from CSV file," in 2018 International Conference on Artificial Intelligence and Big Data, ICAIBD 2018, 2018. doi: 10.1109/ICAIBD.2018.8396169.
[15] J. Koo, S. Baek, and S. Kim, "The effect of personal value on CSV (Creating Shared Value)," Journal of Open Innovation: Technology, Market, and Complexity, vol. 5, no. 2, 2019, doi: 10.3390/JOITMC5020034.
[16] A. Paullada, I. D. Raji, E. M. Bender, E. Denton, and A. Hanna, "Data and its (dis)contents: A survey of dataset development and use in machine learning research," Patterns, vol. 2, no. 11. Cell Press, Nov. 12, 2021. doi: 10.1016/j.patter.2021.100336.