

Implementasi *Support Vector Machine* (SVM) Dengan *Bag-Of-Words* Untuk Klasifikasi Jamu Tradisional Madura

Rika Yunitarini
Teknik Informatika
Universitas Trunojoyo Madura
Bangkalan, Madura
rika.yunitarini@trunojoyo.ac.id

Iliya Rosida
Teknik Informatika
Universitas Trunojoyo Madura
Bangkalan, Madura
210411100051@student.trunojoyo.ac.id

Evy Maya Stefany
Pendidikan Informatika
Universitas Trunojoyo Madura
Bangkalan, Madura
Evy.Stefany@trunojoyo.ac.id

Abstract— Jamu tradisional Madura merupakan bagian dari warisan budaya Indonesia dan menawarkan beragam manfaat kesehatan yang menarik perhatian banyak orang. Namun mengingat semakin beragamnya jamu yang tersedia, sulit untuk mengelompokkan khasiat jamu Madura berdasarkan atribut yang ada. Penelitian ini bertujuan untuk mengidentifikasi jamu Madura dengan menggabungkan metode support vector machine (SVM) dan bag-of-words. Sebuah metode untuk mengklasifikasikan data teks sebagai representasi. Data yang digunakan meliputi atribut seperti khasiat jamu Madura. Fase prapemrosesan melibatkan pembersihan teks dan mengubahnya menjadi vektor kata menggunakan Bag of words. Tahap ini berfungsi sebagai masukan ke model SVM. Hasil penelitian menunjukkan bahwa metode SVM dengan pendekatan bag-of-words mencapai akurasi sebesar 85% dan berhasil mengklasifikasikan jamu tradisional Madura dengan tingkat akurasi yang baik. Penelitian ini diharapkan dapat berkontribusi pada pengembangan sistem berbasis teks yang lebih efektif dalam mengklasifikasikan jamu Madura.

Kata Kunci— *Bag-Of-Words, Jamu Madura, Klasifikasi, Support Vector Machine (SVM).*

I. PENDAHULUAN

Jamu Madura adalah warisan kesehatan tradisional yang berasal dari Pulau Madura dan harus dijaga dan dikembangkan oleh masyarakat Madura [1]. Madura selain dikenal sebagai salah satu pusat produksi garam dan tembakau, juga terkenal dengan jamu tradisionalnya. Jamu Madura dikenal karena manfaatnya secara regional, nasional, dan internasional. baik di tingkat nasional maupun internasional [2]. Masyarakat Madura masih menggunakan tanaman obat secara tradisional [3]. Jamu Madura memiliki berbagai jenis produk dengan khasiat yang terbukti dalam mengobati berbagai keluhan kesehatan. Masyarakat sering mengonsumsi jamu untuk kesehatan, seperti meredakan kelelahan, pegal linu, dan menjaga kebugaran [4]. Pada era saat ini, minat masyarakat terhadap pengobatan alami dan herbal semakin meningkat. Meskipun telah banyak obat-obatan modern, jamu tetap diminati baik di pedesaan maupun perkotaan. Salah satu pilihan yang populer adalah jamu Madura [5].

Berbagai jenis jamu dengan manfaat beragam saat ini dapat ditemukan di berbagai toko, baik dalam bentuk tradisional maupun kemasan modern. Jamu Madura memiliki keunggulan dari segi manfaatnya yang tidak hanya untuk

kesehatan, tetapi juga untuk perawatan wanita dan pasangan suami istri (pasutri)[4]. Namun, dengan banyaknya varian jamu yang tersedia, klasifikasi jamu Madura berdasarkan atributnya menjadi semakin kompleks. Dalam mendukung pelestarian dan pengembangan produk jamu Madura, penting untuk membangun sistem klasifikasi yang efektif untuk setiap jamu. Klasifikasi berdasarkan kandungan dan kategori dapat mendukung penelitian lebih lanjut mengenai efek bahan-bahan dalam konteks kesehatan tertentu, seperti kesehatan wanita atau pasutri.

Pada penelitian ini, penulis menggunakan Metode Support Vector Machine (SVM) yang telah terbukti efektif dalam berbagai tugas klasifikasi, termasuk klasifikasi teks. SVM mampu menangani data dengan dimensi yang tinggi dan dapat menghasilkan model klasifikasi yang akurat dengan menggunakan teknik representasi data teks yang tepat. Hasil ini didukung oleh penelitian sebelumnya mengenai klasifikasi teks, seperti studi yang menggunakan support vector machine (SVM) dan penelitian tentang klasifikasi teks yang tidak seimbang. Sebagai contoh, sebuah penelitian mencatat akurasi SVM sebesar 86% dalam klasifikasi teks [6].

Penelitian lain menunjukkan bahwa SVM lebih akurat dalam klasifikasi topik berita daripada metode lain, dengan tingkat keakuratan mencapai 94.24% [7]. Salah satu pendekatan pengajaran mesin yang digunakan untuk mengekstraksi fitur teks adalah pendekatan bag of word (BoW)[8]. Salah satu teknik umum dalam pengolahan teks adalah Bag-of-Words (BoW), yang mentransformasikan teks menjadi representasi numerik berdasarkan frekuensi kemunculan kata dalam dokumen. Definisi lain untuk Bag of Words (BoW) adalah model yang mempelajari kosakata dari seluruh dokumen, dan kemudian memodelkan setiap dokumen dengan menghitung jumlah kemunculan setiap kata [9]. Kombinasi Support Vector Machine (SVM) dengan BoW memberikan keunggulan dalam mengklasifikasikan teks dengan mengidentifikasi pola kata yang penting untuk menentukan kategori tertentu. Meskipun banyak penelitian telah menggunakan SVM dan BoW untuk klasifikasi teks, penerapan metode ini untuk klasifikasi jamu Madura masih minim. Tujuan penelitian ini adalah untuk mengatasi kesenjangan tersebut dengan menggunakan SVM yang dipadukan dengan BoW untuk mengklasifikasikan jamu Madura berdasarkan atribut-atribut teks yang terkait. Dengan demikian, penelitian ini tidak hanya berkontribusi pada pengembangan sistem klasifikasi yang lebih baik untuk jamu Madura tetapi juga memberikan wawasan baru dalam

penerapan metode machine learning dalam konteks pengobatan tradisional.

Dengan latar belakang ini, penelitian ini bertujuan untuk mengembangkan model klasifikasi yang dapat membantu memetakan jamu Madura secara lebih akurat dan efisien. Implementasi model ini diharapkan dapat meningkatkan kesadaran masyarakat akan manfaat jamu Madura, serta mendukung pelestarian warisan budaya Indonesia.

II. TINJAUAN PUSTAKA

Support Vector Machine adalah Metode untuk menganalisis data dan menemukan pola yang dapat digunakan untuk pengklasifikasian[10]. Dibandingkan dengan classifier lain, SVM memberikan solusi terbaik dan paling akurat. [11].SVM beroperasi dengan membangun model yang memisahkan data ke dalam berbagai kelas menggunakan hyperplane, yang berusaha memaksimalkan jarak antara kelas-kelas yang berbeda. Beberapa penelitian telah menggunakan SVM untuk berbagai penerapan, diantaranya adalah pada pengenalan citra, analisis medik, ataupun untuk melakukan prediksi. Secara khusus, Wang merangkum sejumlah penelitian tentang pengembangan SVM dan aplikasinya [12]. Dengan menggunakan kernel trick, SVM mampu menangani data yang tidak linear dan memberikan hasil yang efektif dalam berbagai aplikasi, termasuk analisis teks. Dalam pendekatan ini, teks dipecah menjadi kata-kata individu dan setiap kata dihitung frekuensinya [13]. Dalam pengolahan bahasa alami, model bag of words adalah representasi sederhana dari sebuah teks yang digunakan untuk menghitung berapa banyak katapadas yang muncul dalam sebuah dokumen. Ini juga disebut sebagai model vector space[14].Representasi Bag of Words (BoW) membantu algoritma SVM dalam memproses teks dengan menyediakan gambaran terstruktur dari konten teks [13].

Dalam penerapan metode Support Vector Machine (SVM), diperlukan dua jenis dataset, yaitu data latih dan data uji, yang akan digunakan untuk klasifikasi. Pada SVM, jumlah data latih yang lebih banyak dapat meningkatkan akurasi prediksi [15]. Rumus umum untuk SVM dengan kernel linear dapat dinyatakan sebagai berikut :

$$f(x) = \text{sign}(w \cdot x + b) \quad (1)$$

Dimana , $f(x)$ adalah fungsi yang digunakan untuk prediksi, w adalah vektor normal hyperplane, x adalah vektor input, dan b adalah bias atau intercept.

Pada penelitian ini , empat jenis kernel digunakan untuk mengatasi masalah klasifikasi, yaitu kernel linear, polinomial, Radial Basis Function (RBF), dan sigmoid[15]. Berikut adalah penjelasan mengenai masing-masing kernel beserta rumusnya:

1. Kernel Linear

Kernel ini menghitung produk titik (dot product) antara dua vektor input dalam ruang asli tanpa mengubahnya ke dalam ruang fitur yang lebih tinggi

$$K(x, y) = (x \cdot y) \quad (2)$$

2. Kernel Polinomial

Kernel polinomial mengukur hubungan polinomial antara dua vektor input dalam ruang asli.

$$K(x, y) = (x \cdot y + c)^d \quad (3)$$

3. Kernel Radial Basis Function (RBF)

Kernel RBF memanfaatkan fungsi Gaussian (juga dikenal sebagai fungsi basis radial) untuk menilai kesamaan antara dua vektor input di ruang fitur.

$$K(x, x') = \exp(-\gamma(x, y)^2) \quad (4)$$

4. Kernel Sigmoid

Kernel sigmoid digunakan untuk mengukur kesamaan antara dua vektor dan efektif dalam menangani data yang tidak dapat dipisahkan secara linear.

$$K(x, y) = \tanh(ax \cdot y + c) \quad (5)$$

III. METODE PENELITIAN

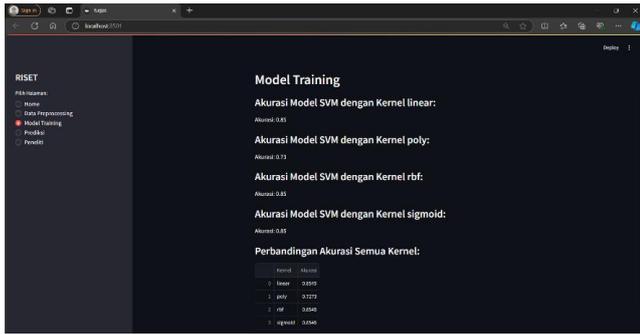
A. Tahapan Penelitian



Gambar 1. Tahapan Penelitian

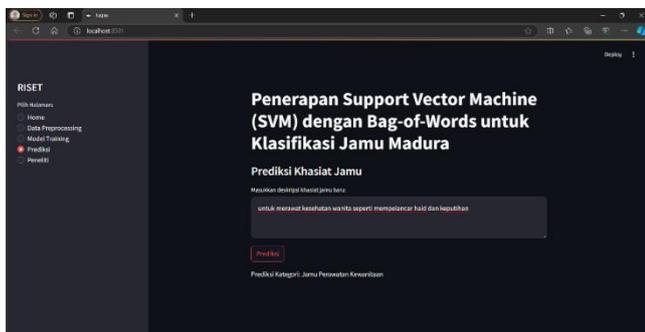
Penelitian ini terdiri dari beberapa langkah utama seperti pada gambar 1, yang diorganisir untuk memastikan setiap proses berjalan lancar dan mendalam. Tahap awal dimulai dengan Perencanaan, di mana masalah dan tujuan penelitian dirumuskan serta metodologi yang akan digunakan ditentukan. Setelah itu, dilakukan Studi Literatur, di mana berbagai referensi terkait topik penelitian dikaji untuk membangun landasan teori yang kuat. Selanjutnya, dilanjutkan dengan Observasi Lapangan dan Pengumpulan

punctuation, tokenizing, remove stopwords, dan stemming. Kemudian juga ada hasil ekstrasi fitur dari bag of words



Gambar 5. Menu Model Training

Gambar 5 merupakan tampilan hasil dari model training dengan 4 kernel yaitu kernel linear, kernel polinomial, kernel RBF dan kernel sigmoid



Gambar 6. Menu Prediksi

Gambar 6 merupakan tampilan menu prediksi dimana menu ini akan menampilkan hasil prediksi dari kasiat jamu Madura apakah khasiat yang dimasukkan masuk ke kategori jamu kesehatan, jamu pasutri atau jamu kewanitaan.

V. KESIMPULAN

Berdasarkan implementasi dan pembahasan mengenai penerapan Support Vector Machine (SVM) dengan metode Bag Of Words untuk klasifikasi jamu Madura. Dapat disimpulkan bahwa:

- Setelah melakukan uji menggunakan algoritma Support Vector Machine , sistem ini mampu mengelompokkan khasiat jamu Madura berdasarkan 3 kategori jamu berdasarkan kriteria yang ada.
- Tingkat Akurasi dari penerapan support vector machine setelah dilakukan perhitungan adalah sebesar 85 %.sehingga dapat disimpulkan bahwa algoritma SVM dapat digunakan untuk mengklasifikasi khasiat jamu Madura berdasarkan kategori yang sudah ada.

ACKNOWLEDGMENT

Penulis mengucapkan terimakasih kepada LPPM Universitas Trunojoyo Madura dan Prodi S1 Teknik Informatika Universitas Trunojoyo yang telah memberi kesempatan pada penulis untuk melakukan riset di bidang jamu Madura.

REFERENCES

- [1] [1] "Pengaruh Undang Undang Jaminan Produk Halal Terhadap Pengembangan".
- [2] [2] R. Yunitarini and E. Widiaswanti, "Analysis and Design of Indonesian Traditional Medicine (Jamu) Information System by using Prototyping Model (Case Study: Madura Island)," in *E3S Web of Conferences*, EDP Sciences, Jan. 2024. doi: 10.1051/e3sconf/202448303012.
- [3] [3] R. Yunitarini and M. A. Effindi, "Production Forecasting of Indonesian Traditional Medicine (Jamu) Based on Information System by Using Single Exponential Smoothing Method," *Management and Production Engineering Review*, vol. 15, no. 1, pp. 90–99, Mar. 2024, doi: 10.24425/mper.2024.149992.
- [4] [4] R. Yunitarini, E. Widiaswanti, and A. P. Nugroho, "SISTEM INFORMASI PERSEDIAAN JAMU MADURA MENGGUNAKAN METODE WATERFALL INFORMATION SYSTEM OF MADURA HERB STOCK USING WATERFALL METHOD," vol. 11, no. 1, 2022.
- [5] [5] A. R. Kusumo, F. Y. Wiyoga, H. P. Perdana, I. Khairunnisa, R. I. Suhandi, and S. S. Prastika, "JAMU TRADISIONAL INDONESIA: TINGKATKAN IMUNITAS TUBUH SECARA ALAMI SELAMA PANDEMI," *Jurnal Layanan Masyarakat (Journal of Public Services)*, vol. 4, no. 2, p. 465, Nov. 2020, doi: 10.20473/jlm.v4i2.2020.465-471.
- [6] [6] L. Mutawalli, M. Taufan, A. Zaen, and W. Bagye, "KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto)," 2019. [Online]. Available: <http://ejournal.stmiklombok.ac.id/index.php/jire>
- [7] [7] L. G. Irfham, A. Adiwijaya, and U. N. Wisesty, "Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 3, no. 4, p. 284, Oct. 2019, doi: 10.30865/mib.v3i4.1410.
- [8] [8] "Analisis komentar cyberbullying terhadap kata yang mengandung toksisitas dan agresi menggunakan bag of words dan tf idf dengan klasifikasi svm".
- [9] [9] J. Muara Sains, dan Ilmu Kesehatan, W. Trisari Harsanti Putri, and R. Hendrowati, "PENGALIAN TEKS DENGAN MODEL BAG OF WORDS TERHADAP DATA TWITTER," vol. 2, no. 1, pp. 129–138, 2018.
- [10] [10] C. Fitri, N. Halizah, and M. Kartikasari, "IMPLEMENTASI ALGORITMA NAÏVE BAYES DALAM PENENTUAN PEMBERIAN BEASISWA KIP KULIAH (STUDI KASUS STIKI MALANG) IMPLEMENTATION OF NAÏVE BAYES ALGORITHM IN DETERMINING KIP-KULIAH SCHOLARSHIP (CASE STUDY STIKI MALANG)," vol. 12, no. 2, 2024.
- [11] [11] J. Resti and F. Selva Jumeilah, "Terbit online pada laman web jurnal: <http://jurnal.iaii.or.id> Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," 2017. [Online]. Available: <http://jurnal.iaii.or.id>
- [12] [12] W. Jurnal, N. Indriani, E. Rainarli, K. Evita Dewi, T. dan Ilmu Komputer, and J. Dipati Ukur, "JURNAL INFOTEL Informatika-Telekomunikasi-Elektronika Peringkasan dan Support Vector Machine pada Klasifikasi Dokumen," *Jurnal Infotel*, vol. 9, no. 4, 2017, doi: 10.20895/infotel.v9i4.
- [13] [13] A. A. Firdaus, A. Id Hadiana, and A. K. Ningsih, "Klasifikasi Sentimen pada Aplikasi Shopee Menggunakan Fitur Bag of Word dan Algoritma Random Forest," *R2J*, vol. 6, no. 5, 2024, doi: 10.38035/r2j.v6i5.
- [14] [14] R. Farhan, R. Pohan, D. E. Ratnawati, and I. Arwani, "Implementasi Algoritma Support Vector Machine dan Model Bag-of-Words dalam Analisis Sentimen mengenai PILKADA 2020 pada Pengguna Twitter," 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [15] [15] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.