

PENERAPAN ALGORITMA SVM DAN SIFT DESKRIPTOR UNTUK
PENGENALAN TULISAN TANGAN PADA KARAKTER HANACARAKA
AKSARA JAWA



Disusun oleh :
Rama Tri Agung
123180053

PROGRAM STUDI INFORMATIKA
JURUSAN INFORMATIKA
FAKULTAS TEKNIK INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" YOGYAKARTA
2022

BAB I

PENDAHULUAN

1.1. Latar Belakang

Aksara Jawa "ha-na-ca-ra-ka" merupakan salah satu warisan leluhur bangsa Indonesia (Sari et al., 2018). Aksara Jawa juga bagian dari bahasa Jawa yang melekat dalam budaya Jawa. Namun pengguna bahasa Jawa berkurang jumlahnya dan hanya sedikit remaja yang mengenal aksara Jawa dengan jelas (Setiawan et al., 2019). Pengenalan tulisan tangan (*Handwriting recognition*) merupakan kemampuan komputer dalam menerima dan memproses input tulisan tangan manusia yang dapat dipahami dari sumber seperti dokumen kertas, foto, dan lain-lain. Pengenalan tulisan tangan ini berguna untuk menunjang kelestarian Aksara Jawa sebagai alat atau perangkat lunak yang memiliki kemampuan untuk mengenali tulisan tangan karakter Aksara Jawa secara otomatis (Dewa et al., 2018).

Pengenalan tulisan tangan pada karakter hanacaraka aksara jawa telah diusulkan oleh beberapa penelitian sebelumnya dengan menggunakan metode yaitu *Convolutional Neural Network* (CNN) (Dewa et al., 2018) (Rismiyati et al., 2017) (Wibowo et al., 2018), *K-Nearest Neighbor* (KNN) (Sari et al., 2018), dan *Support Vector Machine* (SVM) (Rismiyati et al., 2018). Pada metode CNN memiliki performa klasifikasi yang sangat baik dalam bidang ini dengan tingkat akurasi yang dapat mencapai 94.57% (Wibowo et al., 2018), CNN juga mahir dalam menangani inputan yang bersifat *noisy* (Rajesh et al., 2016), namun akurasi yang tinggi pada CNN membutuhkan jumlah data training yang banyak (Wibowo et al., 2018) dan dengan metode yang kompleks ini maka arsitekturnya akan cukup sulit dibangun serta dapat terjadinya *overfitting* (Rajesh et al., 2016). KNN adalah metode yang sederhana, efektif, mudah diterapkan, tidak parametrik dan memberikan tingkat kesalahan yang rendah dalam proses pelatihan (Thamilselvana & Sathiaselvan, 2015), metode ini tidak membutuhkan jumlah dataset yang banyak (Sari et al., 2018), tapi kekurangannya pada metode KNN relatif memiliki hasil performa yang kurang baik daripada metode lainnya (Naufal et al., 2021) dan sulit menemukan nilai optimal (Thamilselvana & Sathiaselvan, 2015). Kemudian metode SVM merupakan metode paling efektif dalam klasifikasi, terutama populer dalam klasifikasi teks, memiliki akurasi yang cukup tinggi (Thamilselvana & Sathiaselvan, 2015), tidak memiliki

masalah dalam *overfitting* (Rajesh et al., 2016), dan tidak membutuhkan jumlah dataset yang sangat besar (Rismiyati et al., 2018). Namun sayangnya, metode ini cukup sulit untuk mencari model parameter yang cocok maupun optimal dalam klasifikasi (Thamilselvana & Sathiaselvan, 2015).

Beberapa penelitian diatas, metode SVM dapat mengatasi masalah penggunaan dataset yang cukup besar dan memiliki performa yang cukup baik, namun penelitian sebelumnya yang menggunakan metode tersebut belum dapat menyaingi akurasi dari metode CNN yang diatas 90% (Rismiyati et al., 2018) (Sari et al., 2018). Untuk dapat meningkatkan akurasi dibutuhkan bantuan fitur ekstraksi, pada penelitian pengenalan tulisan tangan karakter lainnya (Thailand, Bangla dan Latin) telah mengusulkan metode yaitu menggunakan *Scale Invariant Feature Transform Descriptor* (SIFT Descriptor) yang berpengaruh dalam peningkatan akurasi klasifikasi secara signifikan menjadi diatas 95% dan mengungguli performa fitur ekstraksi *Histograms of Oriented Gradients* (HOG) (Surinta et al., 2015).

Sehingga pada penelitian ini pengenalan tulisan tangan pada karakter hanacaraka aksara jawa akan menerapkan metode SVM dengan bantuan SIFT dalam meningkatkan akurasi. Performa kedua metode tersebut akan dibandingkan dan menemukan metode yang paling efektif dalam hal akurasi.

1.2. Rumusan Masalah

Rumusan masalah pada penelitian ini yaitu:

1. Bagaimana menerapkan algoritma SVM dalam klasifikasi pengenalan tulisan tangan aksara jawa dengan *dataset* yang sedikit?
2. Bagaimana menerapkan algoritma SIFT sebagai ekstraksi fitur dalam membantu meningkatkan akurasi klasifikasi?
3. Bagaimana penerapan evaluasi performa akurasi dalam klasifikasi?

1.3. Batasan Masalah

Batasan masalah pada penelitian ini sebagai berikut:

1. Klasifikasi dilakukan hanya pada 20 karakter Hanacaraka Aksara Jawa.
2. Sumber data berasal dari *dataset* yang disediakan di internet oleh Kaggle.com.
3. *Dataset* yang digunakan tidak lebih dari 3000 *record*.

4. Analisis dilakukan untuk melihat performa algoritma dalam akurasi klasifikasi.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini yaitu menerapkan algoritma SVM dengan bantuan SIFT sebagai fitur ekstraksi dalam melakukan klasifikasi tulisan tangan aksara jawa dan mengidentifikasi performa akurasi algoritma yang terbaik dalam melakukan klasifikasi.

1.5. Manfaat Penelitian

Hasil penelitian dapat dimanfaatkan dalam bantuan pembelajaran siswa pada instansi pendidikan dan membangun kembali budaya bahasa khas jawa dengan mengenal lebih mudah terhadap karakter-karakternya.

1.6. Tahapan Penelitian

Tahapan-tahapan pada penelitian ini menerapkan metode proses dari CRISP-DM merujuk pada penelitian oleh Rasyidi, et al. (2021), Dewa, et al. (2018), dan Schröera, et al. (2021) yang telah dimodifikasi untuk menyesuaikan kebutuhan penelitian, yaitu sebagai berikut:

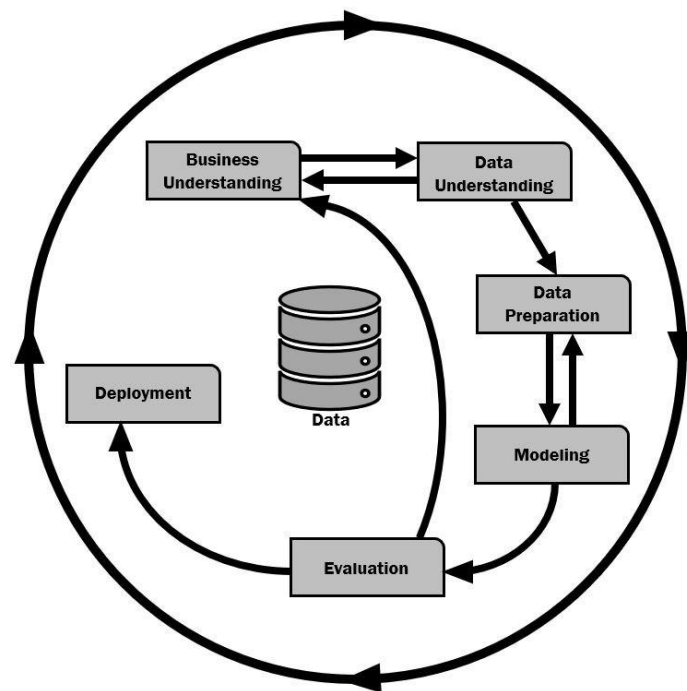
1. *Business Understanding*
Melakukan studi literatur yang berkaitan dengan penelitian ini yaitu pengenalan tulisan tangan, karakter aksara jawa, metode SVM dan SIFT.
2. *Data Understanding*
Mengumpulkan dan mengeksplorasi *dataset* yang perlu digunakan dalam penelitian ini dan memahami pola serta struktur yang penting dari data gambar tersebut.
3. *Data Preparation*
Melakukan persiapan data sebelum dilatih terhadap model. Beberapa tahap yang dilakukan disini yaitu augmentasi data gambar, *preprocessing* data gambar, fitur ekstraksi SIFT, dan normalisasi fitur serta *splitting* data menjadi *training* dan *testing*.
4. *Modelling*
Melakukan pelatihan data fitur yang telah disiapkan menggunakan model *Support Vector Machine*, serta melakukan *tuning parameter* untuk mencari parameter terbaik dari model SVM tersebut.

5. *Evaluation*

Melakukan evaluasi performa akurasi terhadap model yang telah dibangun dengan beberapa macam skenario pengujian.

6. *Deployment*

Melakukan pengembangan sistem aplikasi berbasis *website* dari hasil model yang telah selesai dilatih dan diuji dengan baik. *Deployment* ini fase terakhir yang berguna untuk menyediakan interaksi pengguna terhadap aplikasi sehingga dapat dioperasikan secara luas atau global.



Gambar xx. CRISP-DM *Data Science Process*

1.7. Sistematika Penulisan

Pada penelitian ini memiliki sistematika penulisan yang terurut yaitu pertama pada Bab I (Pendahuluan) yang menjelaskan latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, tahapan penelitian, dan sistematika penulisan. Bab I ini bertujuan untuk memudahkan pembaca memahami maksud dan tujuan penelitian ini. Kemudian pada Bab II (Tinjauan Pustaka) menjelaskan mengenai pembahasan dari penelitian terkait dengan penelitian sebelumnya yang akan digunakan sebagai referensi di penelitian ini. Selanjutnya pada Bab III (Metodologi Penelitian) yaitu menjelaskan mengenai metode-metode penyelesaian yang akan dilakukan pada penelitian ini untuk menyelesaikan masalah

yang diangkat mulai dari perancangan hingga pengujian. Setelah itu pada Bab IV (Hasil dan Pembahasan) yaitu menjelaskan mengenai analisis dan pembahasan dari hasil yang didapatkan pada sistem yang telah dibangun berdasarkan metode yang digunakan. Terakhir pada Bab V (Kesimpulan dan Saran) yaitu menjelaskan mengenai kesimpulan yang didapatkan dari hasil penelitian ini dan memberikan saran yang dapat digunakan sebagai bekal pada pengembangan penelitian ini maupun penelitian selanjutnya.

DAFTAR PUSTAKA

- Dewa, C. K., Fadhillah, A. L., & Afiahayati, A. (2018). Convolutional Neural Networks for Handwritten Javanese Character Recognition. *Indonesian Journal of Computing and Cybernetics Systems*, 12(1), 83-94.
- Naufal, M. F., Kusuma, S. F., Prayuska, Z. A., & Alexander, A. (2021). Comparative Analysis of Image Classification Algorithms for Face Mask Detection. *Journal of Information Systems Engineering and Business Intelligence*, 7(1), 56-66.
- Rajesh S., R., Beaula, A., Marikkannu, P., Sungheetha, A., & Sahana, C. (2016). Comparative study of distinctive image classification techniques. 2016 10th International Conference on Intelligent Systems and Control. Coimbatore: IEEE.
- Rasyidi, M. A., Bariyah, T., Riskajaya, Y. I., & Septyani, A. D. (2021). Classification of handwritten javanese script using random forest algorithm. *Bulletin of Electrical Engineering and Informatics*, 10 (3), 1308-1315.
- Rismiyati, Khadijah, & Adi, N. (2017). Deep learning for handwritten Javanese character recognition. 2017 1st International Conference on Informatics and Computational Sciences (pp. 59-63). Semarang: IEEE.
- Rismiyati, Khadijah, & Riyanto, D. (2018). HOG and Zone Base Features for Handwritten Javanese Character Classification. 2018 2nd International Conference on Informatics and Computational Science (pp. 131-135). Semarang: IEEE.
- Sari, C. A., Kuncoro, M. W., Setiadi, D. R. I. M., & Rachmawanto, E. H. (2018). Roundness and eccentricity feature extraction for Javanese handwritten character recognition based on K-nearest neighbor. 2018 International Seminar on Research of Information Technology and Intelligent Systems (pp. 5-10). Yogyakarta: IEEE.
- Schröera, C., Kruse, F., & Gómez, J., M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. *Procedia Computer Science*, 181, 526-534.
- Setiawan, A., Prabowo, A. S., & Puspaningrum, E. Y. (2019). Handwriting Character Recognition Javanese Letters Based on Artificial Neural Network. *International Journal of Computer, Network Security and Information System*, 1(1), 39-42.

- Surinta, O., Karaaba, M. F., Schomaker, L. R. B., & Wiering, M. A. (2015). Recognition of handwritten characters using local gradient feature descriptors. *Engineering Applications of Artificial Intelligence*, 45, 405-414.
- Thamilselvana, P., & Sathiaselvan, J. G. R. (2015). A Comparative Study of Data Mining Algorithms for Image Classification. *International Journal of Education and Management Engineering*, 5(2), 1-9.
- Wibowo, M. A., Soleh, M., Pradani, W., Hidayanto, A. N., & Arymurthy, A. M. (2017). Handwritten Javanese Character Recognition using Discriminative Deep Learning Technique. *2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering* (pp. 325-330). Yogyakarta: IEEE.