

PENERAPAN DATA MINING UNTUK PREDIKSI PENDAFTAR PPDB DI SMKN3 METRO MENGGUNAKAN MACHINE LEARNING

Mukhammad Khoirul Effendi¹, Sriyanto^{2*}, Goesderilidar³, Handoyo Widi Nugroho⁴, Joko triloka⁵

^{1,2}Magister Teknik Informatika Pascasarja IBI Darmajaya, Bandar Lampung, Indonesia

Email: 'khoiruleffendi24@gmail.com, *sriyanto@darmajaya.ac.id, goesderi.lidar@stmikindragiri.ac.id, handoyo.wn@darmajaya.ac.id, joko.triloka@darmajaya.ac.id

Abstrak

Penelitian ini bertujuan untuk menerapkan teknik data mining dalam memprediksi jumlah pendaftar Penerimaan Peserta Didik Baru (PPDB) di SMKN3 Metro menggunakan algoritma *machine learning*, khususnya *Decision Tree* (C4.5). Masalah utama yang dihadapi adalah tantangan pengelolaan data historis dan keterbatasan kapasitas sekolah dalam merencanakan penerimaan siswa secara efektif. Metode penelitian meliputi pengumpulan data historis pendaftaran, pra-pemrosesan data, penerapan algoritma machine learning, serta evaluasi kinerja model menggunakan metrik seperti *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa model *Decision Tree* (C4.5) memiliki performa terbaik dibandingkan algoritma lain, dengan nilai MSE sebesar 290,948, RMSE 17,057, MAE 11,096, dan R^2 sebesar 0,893. Akurasi prediksi yang tinggi ini menunjukkan potensi besar algoritma tersebut dalam mendukung pengelolaan PPDB secara lebih efisien. Penelitian ini diharapkan dapat menjadi solusi inovatif bagi SMKN3 Metro dalam merencanakan penerimaan siswa baru dan optimalisasi sumber daya sekolah. Selain itu, model ini dapat menjadi referensi bagi institusi pendidikan lain dalam mengadopsi teknologi serupa.

Kata Kunci: Data Mining, Prediksi Pendaftar, PPDB, Decision Tree, SMKN3 Metro

Abstract

This research focuses on implementing data mining to predict the number of registrants for new student admissions (PPDB) at SMKN3 Metro using the C4.5 machine learning algorithm. The study aims to address annual challenges in data management and school capacity limitations. By leveraging historical registration data, an accurate predictive model is developed to assist the school in planning student admissions more effectively. The methodology includes data collection and preprocessing, application of the C4.5 algorithm, and model performance evaluation based on prediction accuracy. Preliminary results indicate that the C4.5 algorithm outperforms other models, achieving a Mean Squared Error (MSE) of 290.948, Root Mean Squared Error (RMSE) of 17.057, and a coefficient of determination (R^2) of 0.893. These findings demonstrate the model's reliability in estimating the number of registrants for key competencies such as Software Engineering and Computer Network Engineering. This implementation is expected to improve the efficiency of the PPDB process and resource planning at SMKN3 Metro, while providing a practical application of data mining and machine learning in educational management.

Keywords: Data Mining, PPDB Prediction, Machine Learning, C4.5 Algorithm, SMKN3 Metro

1. PENDAHULUAN

Pendidikan memiliki peran sentral dalam pembangunan suatu negara[1], tidak hanya sebagai penyokong produktivitas tetapi juga sebagai pembentuk karakter dan keterampilan individu[2]. Di Indonesia, pendidikan dianggap sebagai fondasi utama pembangunan ekonomi, sosial, dan politik[3], menuntut adanya kebijakan yang terarah dan berkelanjutan untuk memenuhi kebutuhan zaman[4]. Salah satu aspek penting dalam sistem

pendidikan Indonesia adalah proses penerimaan peserta didik baru (PPDB)[5], yang tidak hanya menjadi akses bagi siswa ke pendidikan menengah kejuruan (SMK) tetapi juga mencerminkan tantangan dan potensi sektor pendidikan di Indonesia[6]. SMKN3 Metro, sebagai salah satu SMK terkemuka di wilayahnya, menghadapi tantangan dalam mengelola lonjakan pendaftar PPDB setiap tahunnya, menuntut solusi yang efektif melalui penggunaan teknologi informasi yang canggih.

Identifikasi masalah dalam penelitian ini mencakup keterbatasan kapasitas sekolah, kompleksitas pengelolaan data, ketidakmerataan akses teknologi informasi, persaingan dalam menarik minat calon siswa, dan tantangan responsif dan adaptif. Dalam batasan masalah, penelitian akan difokuskan pada penerapan teknik data mining menggunakan Algoritma C4.5 untuk memprediksi jumlah pendaftar PPDB di SMKN3 Metro. Perumusan masalah yang diajukan adalah bagaimana memprediksi jumlah pendaftar PPDB di SMKN3 Metro dengan akurasi tinggi, untuk membantu sekolah merencanakan kapasitas penerimaan siswa dan meningkatkan efisiensi pengelolaan proses PPDB. Tujuan utama penelitian ini meliputi pengembangan model prediksi, peningkatan efisiensi pengelolaan proses penerimaan siswa, penyediaan alat bantu untuk merencanakan kapasitas penerimaan siswa, serta peningkatan pemahaman tentang potensi data mining dan machine learning.

Adapun penelitian-penelitian relevan yang menunjukkan bagaimana metode klasifikasi seperti algoritma C4.5 dapat diterapkan secara efektif untuk memprediksi penerimaan siswa, yang sejalan dengan tujuan penelitian kami.

Penelitian pertama berjudul "Penerapan Data Mining Untuk Memprediksi Penerimaan Peserta Didik Baru Menggunakan Algoritma C4.5" dilakukan oleh Riki Winanjaya, Faisal Amir, dan Rahmad Doni di AMIK Tunas Bangsa, Pematang Siantar, dan Universitas Potensi Utama, Medan. Penelitian ini bertujuan untuk memprediksi calon siswa baru di SMK Hasanah berdasarkan data yang ada dengan menggunakan algoritma C4.5. Hasil penelitian menunjukkan bahwa variabel dengan prioritas tertinggi dalam prediksi PPDB adalah nilai tes calon siswa[7].

Penelitian kedua berjudul "Penerapan Algoritma C4.5 dalam Klasifikasi Penerimaan Peserta Didik Baru" yang menggunakan metode Decision Tree (Algoritma C4.5) untuk memprediksi penerimaan peserta didik baru di SMA Negeri 1 Bandung: Hasil analisis data dengan menggunakan metode Algoritma C4.5 menunjukkan bahwa atribut yang paling mempengaruhi dalam penerimaan peserta didik baru adalah nilai tes matematika yang memberikan kontribusi terbesar. Selanjutnya, nilai tes bahasa Inggris dan IPA juga memiliki kontribusi signifikan dalam pembentukan pohon keputusan. Atribut dengan kontribusi terkecil adalah asal sekolah (AS). Berdasarkan hasil pengujian, akurasi algoritma klasifikasi C4.5 mencapai 90,76%, dengan presisi 94,20% dan recall 92,85%. Dengan demikian, algoritma C4.5 dapat diimplementasikan dengan baik dalam penerimaan peserta didik baru.

Penelitian ketiga berjudul "Perancangan Aplikasi Prediksi Jumlah Pendaftar Siswa Baru dengan Metode Data Mining" mengimplementasikan teknik Data Mining untuk

memprediksi jumlah pendaftar siswa baru menggunakan algoritma C4.5. Variabel yang digunakan dalam penelitian ini meliputi jurusan siswa, minat siswa, dan hasil tes bakat siswa[8].

Penelitian keempat berjudul "Implementasi Teknik Data Mining untuk Prediksi Peminatan Siswa" menggunakan algoritma C4.5 untuk menentukan jurusan yang akan diambil oleh siswa berdasarkan latar belakang, minat, dan kemampuan mereka. Variabel yang digunakan dalam penelitian ini juga mencakup jurusan siswa, minat siswa, dan hasil tes bakat siswa[9].

Penelitian kelima berjudul "Penerapan Data Mining untuk Prediksi Penerimaan Peserta Didik Baru SMP" menggunakan RapidMiner untuk memprediksi penerimaan peserta didik baru di SMP. Hasil penelitian menunjukkan bahwa total nilai adalah variabel utama dalam pembuatan decision tree dengan akurasi mencapai 87,84% [10].

Penelitian-penelitian terdahulu ini memberikan landasan yang kuat untuk penerapan Data Mining dan Machine Learning dalam memprediksi pendaftar PPDB, menunjukkan bahwa algoritma seperti C4.5 dan tools seperti RapidMiner dapat digunakan untuk menghasilkan prediksi yang akurat berdasarkan berbagai variabel relevan. Temuan ini dapat digunakan sebagai referensi dalam merancang penelitian serupa di SMKN3 Metro.

2. METODOLOGI PENELITIAN

Penelitian ini bertujuan untuk memprediksi jumlah pendaftar PPDB di SMKN3 Metro menggunakan teknik data mining dan machine learning. Berikut adalah langkah-langkah metodologis yang diambil[11].

K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi data. KNN menggunakan perhitungan jarak, sedangkan SVM menggunakan perhitungan hyperplane.

Euclidean Distance sebagai metode perhitungan jarak yang umum digunakan dalam KNN, mengukur jarak antara dua titik dalam ruang fitur dengan menggunakan formula Euclidean. Misalnya, dalam ruang dua dimensi, Euclidean distance antara dua titik $P=(p_1,p_2)$ dan $Q=(q_1,q_2)$ dihitung sebagai akar dari $(p_1-q_1)^2+(p_2-q_2)^2$. Dengan menggunakan Euclidean distance, KNN menentukan tetangga terdekat berdasarkan kedekatan spasial antara data points, yang memberikan bobot lebih besar kepada data points yang secara geometris lebih mirip dengan data point yang akan diklasifikasikan. Metode K-Nearest Neighbors (KNN) memiliki keunggulan dalam kemudahan implementasi dan interpretasi, serta fleksibilitasnya dalam menangani data nonlinear atau non-parametrik.

Keunggulan SVM termasuk kemampuannya untuk mengatasi masalah overfitting karena fokus pada batas keputusan yang umum dan memaksimalkan margin, sehingga mengurangi risiko kesalahan klasifikasi pada data baru. Persamaan matematis SVM tidak dapat dijelaskan secara ringkas tanpa memperkenalkan konsep matematika lanjutan. Secara umum, untuk klasifikasi biner, persamaan hyperplane pada SVM dapat diwakili sebagai:

$$f(x) = \text{sign}(w \cdot x + b) \quad (2)$$

dimana:

$f(x)$ = fungsi keputusan

w = vektor bobot

x = fitur vektor data input

b = konstanta bias

\cdot = operasi dot product sign = fungsi tanda

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini mencakup informasi pendaftaran siswa dari tahun-tahun sebelumnya di SMKN3 Metro[20]. Pengumpulan data dilakukan dengan mengakses arsip pendaftaran dan sistem informasi sekolah. Data yang dikumpulkan meliputi variabel-variabel seperti jumlah pendaftar per kompetensi, hasil seleksi, dan informasi relevan lainnya.

2.2 Pra-Pemrosesan Data

Langkah ini penting untuk meningkatkan kualitas data dan menghasilkan model yang lebih akurat[7]. Data yang dikumpulkan kemudian diproses untuk memastikan kebersihan dan kesiapannya. Langkah ini mencakup pembersihan data untuk mengatasi nilai yang hilang atau tidak konsisten, normalisasi data untuk memastikan skala yang konsisten, dan transformasi variabel untuk mempersiapkan data bagi analisis lebih lanjut.

2.3 Pembagian Data

Data dibagi menjadi dua set utama: set pelatihan dan set pengujian. Set pelatihan digunakan untuk membangun dan melatih model prediksi, sementara set pengujian digunakan untuk mengevaluasi kinerja model. Pembagian dilakukan secara acak dengan rasio 80:20 antara set pelatihan dan set pengujian.

2.4 Pemilihan dan Implementasi Algoritma

Beberapa algoritma *machine learning* diterapkan untuk membangun model prediksi[12], yaitu:

- Regresi Linear*, untuk memahami hubungan linier antara variabel[13].
- Decision Tree (C4.5)*, untuk mengembangkan model berbasis pohon keputusan yang dapat menangani data dengan struktur yang kompleks[14].
- Random Forest*, untuk meningkatkan akurasi dengan menggabungkan beberapa pohon keputusan[15].

- k-Nearest Neighbors (kNN)*, untuk mengklasifikasikan data berdasarkan kedekatannya dengan data lainnya[16].
- Support Vector Machine (SVM)*, untuk memisahkan data dalam ruang fitur yang lebih tinggi.
- Neural Network*, untuk menangkap pola yang lebih kompleks melalui jaringan saraf tiruan.

2.5 Evaluasi Model

Kinerja setiap model dievaluasi menggunakan beberapa metrik, yaitu *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, *Mean Absolute Error (MAE)*, dan koefisien determinasi (R^2). *MAPE (Mean Absolute Percentage Error)* juga digunakan untuk menilai akurasi prediksi dalam bentuk persentase[11].

2.6 Analisis Hasil

Hasil dari setiap model dibandingkan untuk menentukan model dengan performa terbaik. Model dipilih sebagai yang paling efektif berdasarkan nilai MSE terendah, RMSE terendah, MAE terendah, dan R^2 tertinggi. Analisis lebih lanjut dilakukan untuk memahami kesesuaian prediksi model dengan data real jumlah pendaftar, terutama untuk kompetensi utama.

2.7 Kerangka Teori

Metodologi ini didasarkan pada teori data mining dan machine learning, yang menekankan penggunaan teknik statistik dan algoritma pembelajaran mesin untuk mengekstrak informasi dan pola dari data besar[17]. *Decision Tree (C4.5)* dipilih karena kemampuannya dalam menangani data yang kompleks dan menghasilkan model yang interpretable.

Dengan pendekatan ini, penelitian diharapkan dapat memberikan prediksi yang akurat mengenai jumlah pendaftar PPDB di SMKN3 Metro dan memberikan dasar yang kuat untuk perencanaan dan pengelolaan proses penerimaan siswa.

3. HASIL DAN PEMBAHASAN

Penelitian ini mengevaluasi performa berbagai algoritma machine learning dalam memprediksi jumlah pendaftar PPDB di SMKN3 Metro, yaitu *Decision Tree (C4.5)*, *Random Forest*, *k-Nearest Neighbors (kNN)*, *Support Vector Machine (SVM)*, dan *Neural Network*. Hasil analisis menunjukkan bahwa model *Decision Tree (C4.5)* adalah yang paling efektif dan akurat dibandingkan dengan model lainnya.

3.1 Model Decision Tree (C4.5)

Model ini menunjukkan performa terbaik dengan nilai *Mean Squared Error (MSE)* terendah sebesar 290.948, *Root Mean Squared Error (RMSE)*

terendah sebesar 17.057, *Mean Absolute Error* (MAE) terendah sebesar 11.096, dan koefisien determinasi (R^2) tertinggi sebesar 0.893. Nilai *Mean Absolute Percentage Error* (MAPE) sebesar 0.260 juga menunjukkan akurasi prediksi yang sangat baik. *Decision Tree* (C4.5) mampu menghasilkan prediksi yang mendekati data aktual untuk kompetensi utama seperti Rekayasa Perangkat Lunak dan Teknik Komputer & Jaringan, menjadikannya model yang sangat andal untuk aplikasi ini.

3.2 Model Random Forest

Meskipun model ini menunjukkan performa yang cukup baik, dengan MSE sebesar 371.392 dan R^2 sebesar 0.864, hasilnya masih sedikit di bawah *Decision Tree*. *Random Forest* memberikan akurasi yang baik namun tidak sebaik *Decision Tree* dalam hal prediksi.

3.3 Model k-Nearest Neighbors (kNN)

Model ini menunjukkan hasil yang kurang memuaskan, dengan MSE sebesar 1112.515 dan R^2 sebesar 0.593. Performa yang lebih rendah ini mengindikasikan bahwa kNN kurang efektif dalam menangani data yang kompleks dalam konteks PPDB.

3.4 Model Support Vector Machine (SVM)

Model ini memiliki performa yang buruk dengan MSE sebesar 2680.549 dan R^2 sebesar 0.018. Hasil ini menunjukkan bahwa SVM kurang cocok untuk memprediksi jumlah pendaftar dalam kasus ini, kemungkinan karena ketidakmampuan model untuk menangani variabilitas data secara efektif.

3.5 Model Neural Network

Menunjukkan hasil yang paling mengecewakan dengan MSE sebesar 6791.022 dan R^2 negatif sebesar -1.487. Hasil ini menunjukkan bahwa *Neural Network* tidak mampu menangkap pola dalam data dengan baik, kemungkinan akibat dari *overfitting* atau kurangnya *parameter tuning* yang tepat.

3.6 Analisis Hasil

Secara keseluruhan, *Decision Tree* (C4.5) terbukti menjadi model yang paling efektif dalam memprediksi jumlah pendaftar PPDB. Performa buruk dari model SVM dan *Neural Network* dapat dihubungkan dengan kurangnya parameter tuning yang optimal dan ketidakmampuan model-model ini untuk menangani kompleksitas data dengan baik. Perbedaan yang signifikan antara prediksi dan data real untuk beberapa kompetensi mengindikasikan perlunya evaluasi lebih lanjut untuk mengidentifikasi faktor-faktor yang mempengaruhi akurasi prediksi.

Dalam konteks implementasi praktis, *Decision Tree* (C4.5) harus dipertimbangkan

sebagai model utama untuk sistem prediksi PPDB di SMKN3 Metro. Sementara model lain memerlukan optimasi lebih lanjut atau mungkin dipertimbangkan untuk aplikasi yang berbeda. Evaluasi berkala dan pengembangan berkelanjutan dari model prediksi akan sangat penting untuk memastikan akurasi dan relevansi hasil seiring dengan perubahan tren dan data pendaftar.

Model	MSE	RMSE	MAE	MAPE	R2
Tree	290.948	17.057	11.096	0.260	0.893
Random Forest	371.392	19.272	14.133	0.448	0.864
kNN	1112.515	33.354	26.804	0.822	0.593
SVM	2680.549	51.774	37.278	0.928	0.018
Neural Network	6791.022	82.408	66.714	0.774	-1.487

Gambar 1. Kesimpulan Berdasarkan Hasil Analisis

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma *Decision Tree* (C4.5) adalah model prediksi yang paling efektif untuk memperkirakan jumlah pendaftar PPDB di SMKN3 Metro. Dengan nilai *Mean Squared Error* (MSE) terendah, *Root Mean Squared Error* (RMSE) terendah, *Mean Absolute Error* (MAE) terendah, serta koefisien determinasi (R^2) tertinggi, model ini memberikan akurasi yang sangat baik dibandingkan dengan model lainnya, termasuk *Random Forest*, *k-Nearest Neighbors* (kNN), *Support Vector Machine* (SVM), dan *Neural Network*. *Decision Tree* juga mampu mendekati data real untuk kompetensi utama seperti Rekayasa Perangkat Lunak dan Teknik Komputer & Jaringan, menunjukkan kemampuannya dalam memprediksi jumlah pendaftar dengan baik.

Implikasi dari penelitian ini adalah bahwa SMKN3 Metro dapat memanfaatkan model *Decision Tree* untuk merencanakan kapasitas penerimaan siswa dan meningkatkan efisiensi pengelolaan PPDB. Dengan mengintegrasikan model ini ke dalam sistem manajemen sekolah, proses perencanaan dan alokasi sumber daya dapat dilakukan dengan lebih akurat dan strategis. Selain itu, penelitian ini menyoroti perlunya evaluasi berkala dan pengembangan berkelanjutan dari model prediksi untuk memastikan keakuratan hasil sesuai dengan dinamika dan perubahan tren pendaftar.

Namun, penelitian ini juga memiliki keterbatasan. Yaitu adanya variabel tambahan yang belum sepenuhnya terakomodasi dalam model. Keterbatasan ini mengindikasikan perlunya penelitian lebih lanjut untuk mengidentifikasi dan memasukkan faktor-faktor yang mempengaruhi akurasi prediksi secara lebih komprehensif. Selain itu, meskipun *Decision Tree* terbukti unggul dalam konteks ini, pengembangan dan evaluasi model

lainnya tetap penting untuk memastikan fleksibilitas dan adaptabilitas dalam berbagai situasi dan jenis data.

5. REFERENCES

- [1] A. Widodo, "Pentingnya Pendidikan Karakter dalam Dunia Pendidikan," *SHEs Conf. Ser. 4 2077–2081*, vol. 4, no. 5, pp. 2077–2081, 2021, [Online]. Available: [http://repo.iain-tulungagung.ac.id/5510/5/BAB 2.pdf](http://repo.iain-tulungagung.ac.id/5510/5/BAB%202.pdf)
- [2] R. Frederich, Nurhayati, and S. F. Purba, "Peranan Pendidikan Dalam Meningkatkan Pertumbuhan Ekonomi Di Indonesia," *J. Ilm. Ekon. Bisnis*, vol. 28, no. 1, pp. 123–136, 2023, doi: 10.35760/eb.2023.v28i1.7227.
- [3] S. D. Yufarika, "Tantangan Dunia Pendidikan Dan Implikasinya Terhadap Perubahan Kebijakan," *Ar-Rosikhun J. Manaj. Pendidik. Islam*, vol. 2, no. 2, pp. 156–161, 2023, doi: 10.18860/rosikhun.v2i2.21812.
- [4] Kemendikbud, "Permendikbud Nomor 1 Tahun 2021 tentang Penerimaan Peserta Didik Baru Jenjang TK, SD, SMP, SMA dan SMK," *Permendikbud*, pp. 1–25, 2021.
- [5] P. R. Indonesia, "UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 20 TAHUN 2003 TENTANG SISTEM PENDIDIKAN NASIONAL," vol. 4, pp. 147–173, 2003.
- [6] R. Winanjaya, F. Amir, and R. Doni, "Penerapan Data Mining Untuk Memprediksi Penerimaan Peserta Didik Baru Menggunakan Algoritma C4.5," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 1, 2019, doi: 10.30645/senaris.v1i0.1.
- [7] N. R. Lase and F. Riandari, "Perancangan Aplikasi Prediksi Jumlah Pendaftar Siswa Baru Dengan Metode Regresi Linier (Studi Kasus: SMA RK Deli Murni Bandar Baru)," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 3, pp. 330–334, 2020, doi: 10.32672/jnkti.v3i3.2520.
- [8] N. Manullang, R. W. Sembiring, I. Gunawan, I. Parlina, and Irawan, "Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5," *J. ILMU Komput. DAN Teknol.*, vol. 2, no. 2, pp. 1–5, 2021, [Online]. Available: <http://creativecommons.org/licenses/by/4.0/%0Ahttp://ejournal.uhb.ac.id/index.php/IKOMTI>
- [9] K. A. AGUSTINE, "PENERAPAN DATA MINING UNTUK PREDIKSI PENERIMAAN PESERTA DIDIK BARU SMP/MTS JALUR PMPA DI SMA NEGERI 8 PALEMBANG," 2020. [Online]. Available: <https://talenta.usu.ac.id/politeia/article/view/3955>
- [10] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.
- [11] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*. in Springer Texts in Statistics. Springer New York, 2013. [Online]. Available: https://books.google.co.id/books?id=qcI_AAAAQBAJ
- [12] D. A. Freedman, *Statistical Models: Theory and Practice*. Cambridge University Press, 2009. [Online]. Available: https://books.google.co.id/books?id=fW_9BV5Wpf8C
- [13] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1007/bf00116251.
- [14] L. Breiman, "Random forests. Random Forests," *Mach. Learn.*, vol. 45, no. 45, pp. 5–32, 2001, [Online]. Available: <https://link.springer.com/article/10.1023/A:1010933404324>
- [15] MacQueen, James and others, "Some methods for classification and analysis of multivariate observations," *Proc. fifth Berkeley Symp. Math. Stat. Probab.*, vol. 1, no. 14, pp. 281–297, 1967, [Online]. Available: http://books.google.de/books?hl=de&lr=&id=IC4Ku_7dBFUC&oi=fnd&pg=PA281&dq=MacQueen+some+methods+for+classification&ots=nNTcK1IdoQ&sig=fHzdVcbvmYJ-lTNHu1HncmOFokM#v=onepage&q=MacQueen+some+methods+for+classification&f=false
- [16] Y. E. Yana And N. Nafi'iyah, "Classification Of Banana Types Based On Color, Texture, Image Shape Features Using Svm And Knn," *Research: Journal Of Computer, Information System & Technology Management*, Vol. 4, No. 1, Pp. 28–36, 2021