Perbandingan SVM dan Random Forest dalam Memprediksi Kelulusan Santri Menuju Timur Tengah

MZ Amirudin[1], Kusrini[2*]
[1,2]Universitas Amikom Yogyakarta, Indonesia
E-mail: zunuz4@gmail.com, kusrini@amikom.ac.id

Article Info
Article History
Received: 2025-08-05
Revised: 2025-09-12
Published: 2025-10-06

Keywords:
Support Vector Machine;
Random Forest;
Graduation Prediction;
Islamic Boarding School;
Machine Learning;
SMOTE.

Abstract
This study was conducted to compare the performance of the Support Vector Machine (SVM) and Random Forest (RF) algorithms in predicting the graduation of students from Pondok Pesantren Imam Bukhari to pursue higher education in Middle Eastern universities. The dataset includes academic scores, Arabic language proficiency, and Qur'an memorization, with a total of 1,168 student records. The preprocessing stage involved data conversion, normalization, and handling class imbalance using the Synthetic Minority Oversampling Technique (SMOTE). The models were evaluated using metrics such as accuracy, precision, recall, and F1-score. Based on the evaluation results, the Random Forest algorithm achieved a higher accuracy of 61.1%, while SVM performed better in identifying accepted students, with a recall of 39% and an F1-score of 0.32. These findings suggest that SVM is more suitable when the focus is on identifying high-potential students. This study serves as a foundation for developing a machine learning–based decision support system in Islamic boarding school environments.

Artikel Info
Sejarah Artikel
Diterima: 2025-08-05
Direvisi: 2025-09-12
Dipublikasi: 2025-10-06

Kata kunci:
Support Vector Machine;
Random Forest;
Prediksi Kelulusan;
Pesantren;
Machine Learning;
SMOTE.

Abstrak
Penelitian ini dilakukan guna membandingkan performa algoritma Support Vector Machine (SVM) dan Random Forest (RF) dalam memprediksi kelulusan santri Pondok Pesantren Imam Bukhari untuk melanjutkan studi ke universitas di Timur Tengah. Data yang digunakan meliputi nilai akademik, kemampuan bahasa Arab, dan jumlah hafalan Al-Qur'an, dengan total 1.168 data santri. Proses pra-pemrosesan mencakup konversi data, normalisasi, serta penanganan ketidakseimbangan kelas memanfaatkan metode Synthetic Minority Oversampling Technique (SMOTE). Evaluasi model dilakukan dengan menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Berdasarkan hasil evaluasi, algoritma Random Forest menunjukkan tingkat akurasi yang lebih tinggi sebesar 61,1%, sedangkan SVM menunjukkan kinerja lebih baik dalam mengenali santri yang diterima, dengan recall sebesar 39% dan F1-score 0,32. Temuan ini menunjukkan bahwa SVM lebih cocok digunakan ketika fokus prediksi diarahkan pada identifikasi santri berpotensi. Penelitian ini beperan sebagai dasar pengembangan sistem pendukung keputusan berbasis machine learning di lingkungan pesantren.

I. PENDAHULUAN

Kemajuan teknologi kini mendorong lembaga pendidikan menerapkan metode-machine learning dalam proses penyaringan dan ramalan kinerja akademik [1]. Dalam skenario ini, praktik data mining umumnya dilihat dari dua lensa, komersial dan akademis, walau kedua dunia sering tumpang tindih [2]. Sejumlah algoritma, termasuk support vector machine dan random forest, sudah lazim dipakai untuk memperkirakan beraneka fenomena [3]. *Support Vector Machine* (SVM) jadi favorit klasifikasi karena mampu menuntaskan model dalam bentuk dual lewat teknik quadratic programming [4]. Random Forest, sebaliknya, lahir dari struktur Decision Tree dengan membangun ribuan pohon hingga terbentuk hutan hasil yang lebih stabil [5]. Kompetensi Lulusan (SKL) adalah standar minimal yang harus dicapai oleh peserta didik agar diakui sebagai lulusan dari suatu jenjang pendidikan [6]. SKL menilai kombinasi pengetahuan, keterampilan, dan sikap, sehingga cara utama lembaga menetapkan siapa yang boleh lulus atau tidak [7].

Satu kajian menunjukkan bahawa kaedah *Support Vector Machine* (SVM) dapat meramalkan kelulusan pelajar dalam pendaftaran baharu dengan ketepatan mencecah 94,4 % [8]. Dalam penyelidikan yang lain, Random Forest dipilih bagi menilai pencapaian akhir pelajar, dan menghasilkan ketepatan sebesar 92,4 % [9]. Pengukuran tambahan bagi SVM terdapat dalam laporan yang melaporkan nilai F1-score setinggi 0,92 [10]. Jurnal Khatulistiwa Informatika pula

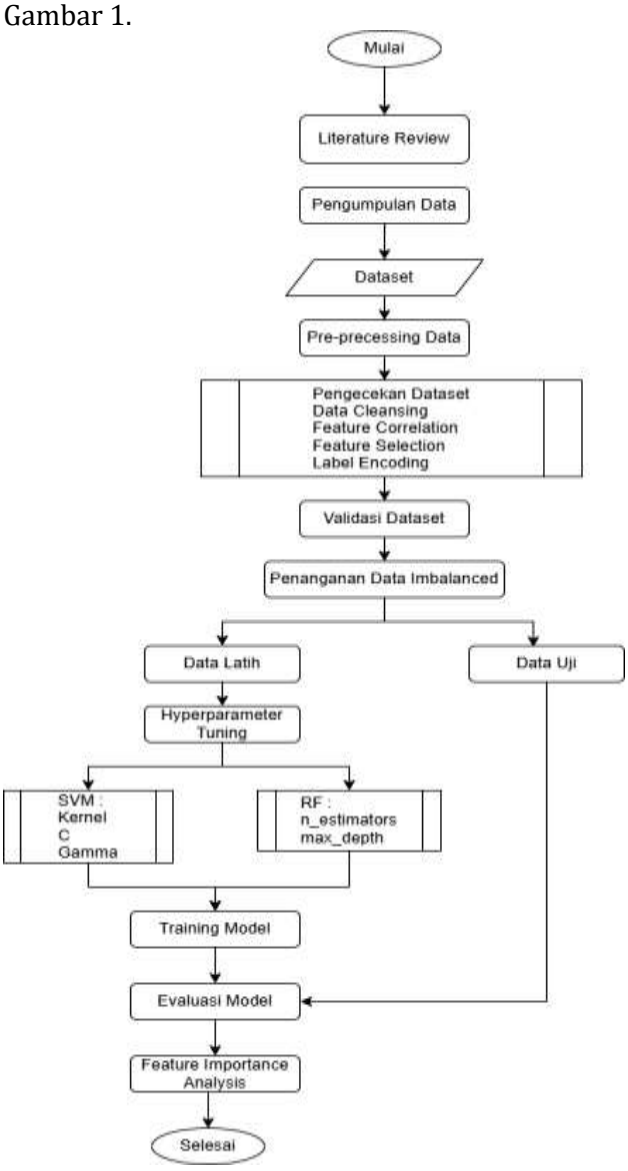
mendapati bahawa SVM boleh menggunakan rekod pembelajaran dalam talian untuk menjangkakan kelulusan, lalu mencapai ketepatan 85 % [11]. Walau bagaimanapun, pendekatan pembelajaran mesin itu masih kurang diterokai dalam konteks pengajian di pesantren. Kebanyakan proses pemilihan pelajar di institusi berkenaan bergantung kepada pemerhatian guru, hafalan, dan pencapaian umum, sementara penerimaan universiti di Timur Tengah sering kali mengutamakan penilaian akademik, kemahiran Bahasa Arab, serta hafalan Al-Qurān. Keadaan itu menekankan kewajipan untuk membangunkan sistem sokongan keputusan yang memberikan anggaran objektif berdasarkan data, sekali gus mengurangkan bias subjektif yang sering berlaku.

Sejumlah penelitian terbaru menunjukkan algoritma pembelajaran mesin mampu meningkatkan mutu pendidikan di tanah air, termasuk dalam lingkungan pesantren. Dalam salah satu studi yang dipublikasikan di Jurnal Informatika dan Teknik Elektro Terapan, penerapan *Random Forest* pada santri Tahfidz di Pondok Pesantren Al Kautsar melahirkan tingkat akurasi sebesar 99,6 persen, precision 100 persen, dan recall 98,8 persen [12]. Studi lain yang tercatat dalam Jurnal Ilmiah Penelitian dan Pembelajaran Informatika membandingkan performa *Support Vector Machine* (SVM) dan *Random Forest* dalam meramalkan capaian belajar mahasiswa, dan menemukan bahwa Random Forest menduduki posisi lebih tinggi dengan akurasi 97 persen, sementara SVM hanya 85 persen [13]. Di bidang seleksi fitur, penggunaan Information Gain Ratio sebelum klasifikasi dengan Random Forest pada pengelompokan pembayaran kuliah menghasilkan akurasi 78,5 persen [14].

Pada penelitian yang dilakukan sebelumnya menunjukkan bahwa penggabungan algoritma *Support Vector Machine* dan teknik pemilihan fitur Relief mampu meningkatkan akurasi prediksi kelulusan siswa dari 69% menjadi 82% [15]. Studi yang lain melaporkan ketepatan 93,3% setelah Random Forest digunakan untuk meramalkan nilai akhir semester [16]. Hasil-hasil itu mengindikasikan kedua model, SVM dan RF, berjalan baik bila dihadapkan pada dataset pendidikan yang kaya variabel. Berdasar bukti tersebut, penelitian ini membandingkan performa kedua algoritma dalam memperkirakan peluang santri Pondok Pesantren Imam Bukhari untuk melanjutkan studi ke universitas di Timur Tengah. Agar analisis holistik, dataset mengombinasikan nilai akademik, kemampuan berbahasa Arab, serta jumlah hafalan Al-Qur-an. Angka akurasi, presisi, recall, dan F1-score akan disusun sebagai dasar untuk memberikan rekomendasi dalam menentukan algoritma apa yang paling efektif untuk keputusan berbasis data di lingkungan pesantren.

II. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan menggunakan pendekatan *machine learning*, metode ini adalah salah satu jenis metode yang spesifikasinya sistematis, terencana dan terstruktur dengan jelas dari awal hingga pembuatan desain penelitiannya [17]. Bahasa pemrograman yang digunakan adalah python dengan memanfaatkan *tools google colabs*. Adapun alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

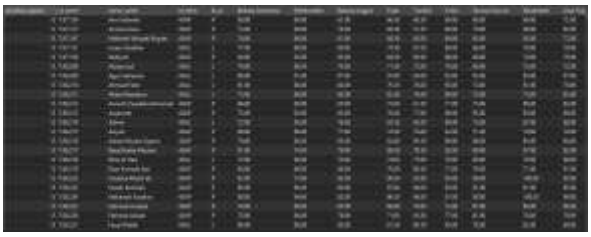
A. Studi Literatur

Penelitian ini mengeksplorasi penggunaan algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) untuk meramal kelulusan siswa, dengan meneliti jurnal nasional yang berkaitan. Artikel-artikel yang dibaca mencakup analisis nilai akademis, pola belajar, serta hasil hafalan dan kecakapan bahasa, sehingga profil belajar siswa bisa digambarkan utuh sebelum kedua metode dipakai. Pada penelitian yang ditelaah, SVM kerap dianggap unggul pada dataset berdimensi tinggi, karena dapat menarik garis pemisah dengan jarak maksimum antara kelas [18]. Di lain pihak, RF lebih cenderung menjaga model tetap sederhana dan stabil, sebab ia mengkombinasikan hasil ratusan pohon keputusan yang dibangun dari sampel acak [19].

B. Pengumpulan *Data*

Sumber data penelitian ini diambil dari sistem informasi akademik Pondok Pesantren Imam Bukhari. Yang dicatat antara lain nilai akademik santri, skor tiap mata pelajaran agama, kemampuan bahasa Arab, dan jumlah hafalan Al-Qur-an. Di samping itu, diinventarisir pula status akhir, yaitu apakah santri melanjutkan studi ke universitas di Timur Tengah atau tidak.

Pengumpulan dilakukan dengan menelusuri arsip internal pesantren dan mentransfernya ke dalam berkas Excel. Seluruh dataset terdiri atas 1.168 entri santri dan 22 variabel numerik, yang meliputi nilai pelajaran umum maupun agama. Untuk analisis awal, label diterima atau tidak diterima dibuat berdasarkan pembagian 70:30, sehingga 30 persen santri dikategorikan diterima dan 70 persen tidak diterima. Sampel dataset ditunjukkan pada Gambar 2.



Gambar 2. Sampel Dataset

C. *Pra-Pemrosesan Data*

Tahap pra-pemrosesan data meliputi pemeriksaan awal dataset, pembersihan informasi yang tidak akurat, dan akhirnya validasi kualitas seluruh koleksi. Langkah-langkah ini ditujukan agar data yang disiapkan untuk melatih model memiliki mutu tinggi dan dapat dipahami dengan tepat oleh algoritma pembelajaran mesin [20]. Langkah pertama adalah penghapusan kolom-kolom yang tidak relevan atau bersifat identitas, seperti nama santri, ID kelas, dan jenis kelamin. Selanjutnya, dilakukan konversi kolom "Jumlah Hafalan" yang semula berupa format teks seperti "2 Juz 10 Lembar" menjadi data numerik berdasarkan asumsi 1 juz = 20 lembar. Dengan demikian, kolom tersebut diubah menjadi format numerik tunggal sebagai representasi total hafalan. Setelah itu, seluruh kolom yang berisi nilai dikonversi ke tipe data numerik dan baris dengan nilai kosong (missing value) dihapus untuk menghindari error dalam proses pelatihan model. Untuk memperkuat proses klasifikasi, label target "kelulusan" ditambahkan dalam bentuk variabel dummy yaitu label_diterima, dengan asumsi distribusi 70% santri tidak diterima dan 30% diterima sebagai data awal simulasi. Data kemudian dinormalisasi menggunakan metode *standard scaling* agar skala antar variabel menjadi sebanding. Penting untuk melakukan normalisasi sebelum latihan, karena algoritma seperti SVM sangat peka terhadap variasi skala fitur. Setelah praproses selesai, data dibagi menjadi dua bagian, yaitu 80 persen untuk pelatihan dan 20 persen untuk pengujian, dengan stratified sampling agar distribusi kelas tetap seimbang.

.

D. *Penanganan Data Imbalanced*

Class imbalance adalah fenomena umum dalam analisis data, di mana jumlah pengamatan dalam satu kelas, biasanya yang lebih besar, jauh melebihi pengamatan dalam kelas lain yang lebih kecil [21]. Untuk mengatasi distribusi tidak seimbang pada label-target, penerapan teknik *Synthetic Minority Oversampling Technique,* atau SMOTE, menjadi pilihan yang populer. Teknik ini menciptakan salinan sintetis untuk anggota kelas minoritas, dalam hal ini santri yang diterima, dengan meramu sampel baru berdasarkan titik-titik tetangga terdekat di dalam ruang fitur. Tujuannya adalah mengurangi bias model yang biasanya mengarah ke kelas mayoritas dan sekaligus meningkatkan sensitivitas terhadap kelas minoritas. Agar evaluasi tetap menggambar-kan keadaan di dunia nyata, SMOTE hanya diterapkan pada set-latihan, bukan pada set-

uji. Setelah langkah tersebut, model dilatih dengan data latih yang sudah diseimbangkan, sementara evaluasi tetap dilakukan pada set-uji asli yang mempertahankan distribusi yang sebenarnya.

E. *Perancangan Model*

Perancangan model dalam penelitian ini mencakup pemilihan algoritma klasifikasi, penentuan parameter utama (*hyperparameter*), serta persiapan data latih untuk proses pelatihan. Adapun algoritma yang digunakan adalah *Support Vector Machine* (SVM) dan *Random Forest* (RF), yang masing-masing memiliki keunggulan dalam klasifikasi non-linear dan pengelolaan data berfitur banyak. Pada tahap awal perancangan, dilakukan penyesuaian parameter penting untuk masing-masing algoritma. Untuk algoritma SVM, parameter yang ditentukan meliputi jenis kernel (dalam penelitian ini menggunakan kernel RBF), nilai parameter regulasi, dan nilai gamma yang memengaruhi bentuk kurva pemisah antar kelas. Sedangkan pada algoritma *Random Forest*, parameter yang diatur antara lain jumlah pohon (*n_estimators*) dan kedalaman maksimum pohon (*max_depth*). Penentuan nilai parameter dilakukan melalui proses pencarian menggunakan metode *grid search* dan validasi silang pada data latih. Setelah parameter ditentukan, model SVM dan RF dibangun menggunakan data latih yang sebelumnya telah melalui proses *balancing* dengan teknik SMOTE. Seluruh proses perancangan dilakukan menggunakan bahasa pemrograman *Python* dengan pustaka *Scikit-Learn*, dan dirancang agar dapat menangani ketidakseimbangan data dan kompleksitas variabel input seperti nilai akademik, kemampuan bahasa Arab, serta jumlah hafalan santri. Tahap ini menjadi dasar dalam membentuk model yang siap dilatih dan dievaluasi, sekaligus menentukan konfigurasi optimal untuk mencapai performa klasifikasi terbaik pada data uji.

F. *Evaluasi Model*

Evaluasi model dilakukan dengan menggunakan dataset uji yang telah disisihkan sebelumnya, tanpa melakukan *oversampling*, agar hasilnya benar-benar mencerminkan kinerja model yang sebenarnya. Metrik evaluasi yang digunakan dalam penelitian ini meliputi akurasi, *precision, recall*, dan *F1-score*. Akurasi mengukur proporsi keseluruhan prediksi yang benar, sedangkan presisi dan recall menilai sejauh mana model secara tepat dan sensitif mengidentifikasi kelas santri yang diterima. *F1-score* digunakan sebagai metrik gabungan untuk mengatasi ketidakseimbangan data. Selain itu, ditampilkan juga *confusion matrix* untuk melihat distribusi hasil prediksi secara lebih rinci.

G. *Feature Importance Analyst*

Untuk algoritma *Random Forest*, dilakukan analisa *feature importance* untuk mengidentifikasi variabel-variabel yang berpengaruh dalam proses klasifikasi kelulusan santri. Analisis ini dilakukan dengan melihat kontribusi masing-masing fitur terhadap pengurangan impuritas dalam pembentukan pohon keputusan. Variabel dengan skor importance tertinggi dianggap memiliki peran besar dalam menentukan output prediksi model. Dalam konteks penelitian ini, fitur-fitur seperti nilai Bahasa Arab, jumlah hafalan Al-Qur'an, dan mata pelajaran agama seperti Fiqih dan Aqidah menunjukkan skor kontribusi yang paling dominan. Temuan ini sesuai dengan karakteristik seleksi universitas Timur Tengah yang memang menekankan aspek kemampuan bahasa dan hafalan. Selain itu, hasil ini juga dapat digunakan sebagai masukan strategis bagi pihak pesantren dalam merancang kurikulum atau memberikan bimbingan khusus kepada santri dengan potensi tinggi. Seluruh proses evaluasi dan analisis dilakukan menggunakan bahasa pemrograman Python dengan bantuan pustaka *Scikit-Learn*, serta divisualisasikan menggunakan *matplotlib* untuk memudahkan interpretasi hasil feature importance secara visual.

III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data santri dari Pondok Pesantren Imam Bukhari yang telah dikumpulkan dalam bentuk nilai akademik, kemampuan bahasa Arab, dan jumlah hafalan Al-Qur'an. Total terdapat 1.168 data santri dengan 22 variabel numerik yang digunakan sebagai fitur. Untuk keperluan analisis, label diterima atau tidak diterima dibuat dengan distribusi 70% tidak diterima dan 30% diterima. Setelah proses pra-pemrosesan dan konversi data selesai, data dibagi menjadi dua bagian yaitu 80% sebagai data latih dan 20% sebagai data uji.

```
Train size: (934, 22)
Test size : (234, 22)
```

Gambar 3. Split Dataset

Distribusi kelas yang tidak seimbang pada data latih (653 tidak diterima vs 281 diterima) ditangani menggunakan teknik yang dikenal dengan nama *Synthetic Minority Oversampling Technique* (SMOTE). Setelah SMOTE diterapkan, maka data latih akan menjadi seimbang dengan jumlah masing-masing kelas sebanyak 653. Langkah ini penting untuk dilakukan guna memastikan bahwa model tidak bias terhadap kelas mayoritas.

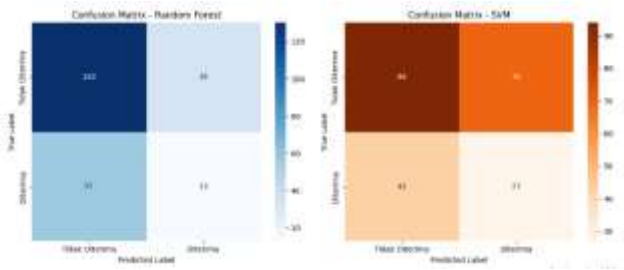
```
Distribusi sebelum SMOTE: Counter({0: 653, 1: 281})
Distribusi sesudah SMOTE : Counter({0: 653, 1: 653})
```

Gambar 4. Penerapan Teknik SMOTE

Selanjutnya dilakukan pelatihan model menggunakan kedua algoritma, yaitu *Support Vector Machine* (SVM) dan *Random Forest* (RF). Parameter yang digunakan pada SVM mencakup kernel RBF, sedangkan pada RF digunakan jumlah pohon default (100) tanpa pembatasan kedalaman pohon secara eksplisit. Setelah pelatihan, kedua model diuji dengan menggunakan data uji yang tidak mengalami *oversampling* agar hasil evaluasi mencerminkan performa sebenarnya. Hasil evaluasi menunjukkan bahwa model *Random Forest* memiliki tingkat akurasi sebesar 61,1%, dengan *precision* untuk kelas diterima sebesar 0,28, *recall* sebesar 0,19, dan *F1-score* sebesar 0,22. Sementara itu, model SVM menghasilkan akurasi sebesar 51,7%, dengan *precision* untuk kelas diterima sebesar 0,28, *recall* sebesar 0,39, dan *F1-score* sebesar 0,32. Dari hasil pelatihan ini dapat dilihat bahwa meskipun *Random Forest* memiliki akurasi keseluruhan yang lebih tinggi, SVM justru lebih baik dalam mengenali santri yang benar-benar diterima (kelas minoritas), yang terlihat dari nilai *recall* dan *F1-score* yang lebih tinggi.

Gambar 5. Hasil Evaluasi Model Dengan Metode *Random Forest* dan *Support Vector Machine*

Confusion matrix ternyata memperkuat temuan ini. Pada model *Random Forest*, dari 70 santri yang sebenarnya diterima, hanya 13 yang berhasil dikenali, sementara sisanya diklasifikasikan sebagai tidak diterima. Sebaliknya, pada model SVM, sebanyak 27 dari 70 santri yang diterima berhasil dikenali, walaupun masih terdapat kekeliruan dalam klasifikasi santri yang sebenarnya tidak diterima.



Gambar 5. Confusion Matrix

Perbandingan ini menunjukkan *trade-off* antara akurasi dan sensitivitas terhadap kelas minoritas. Dalam konteks prediksi kelulusan santri ke universitas Timur Tengah, kemampuan model untuk mengenali santri yang berpotensi diterima (kelas 1) menjadi sangat penting. Oleh karena itu, meskipun akurasi SVM lebih rendah, model ini bisa dianggap lebih unggul dalam konteks seleksi santri yang bersifat preskriptif. Selain evaluasi kuantitatif, dilakukan juga analisis feature importance menggunakan *Random Forest*. Hasilnya menunjukkan bahwa fitur-fitur seperti nilai Bahasa Arab, Hafalan Qur'an, dan mata pelajaran inti seperti Fiqih dan Tauhid termasuk dalam lima besar kontributor utama terhadap prediksi. Ini mendukung asumsi bahwa kemampuan bahasa Arab dan hafalan menjadi indikator kuat dalam seleksi keberangkatan santri ke Timur Tengah. Namun, penelitian ini memiliki keterbatasan karena label yang digunakan masih merupakan data dummy. Model belum dilatih pada data riil yang menunjukkan status diterima atau tidak berdasarkan hasil seleksi sebenarnya. Oleh karena itu, akurasi dan metrik lain belum dapat dijadikan ukuran akhir. Penelitian lanjutan sangat disarankan untuk menggunakan data historis yang valid serta mempertimbangkan penambahan variabel non-akademik seperti wawancara, rekomendasi ustadz, dan motivasi pribadi santri.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma pembelajaran mesin seperti SVM dan RF dapat digunakan untuk mendukung pengambilan keputusan dalam seleksi santri berbasis data. SVM menunjukkan keunggulan dalam mengenali potensi santri yang diterima, sedangkan RF unggul dalam akurasi umum. Pemilihan

algoritma terbaik dapat disesuaikan dengan fokus institusi, apakah menginginkan akurasi tinggi secara keseluruhan atau sensitivitas tinggi terhadap santri berprestasi.

IV. SIMPULAN DAN SARAN

A. Simpulan

Penelitian ini membandingkan algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) dalam memprediksi kelulusan santri Pondok Pesantren Imam Bukhari menuju universitas di Timur Tengah. Hasil menunjukkan bahwa *Random Forest* memiliki akurasi lebih tinggi sebesar 61,1%, sedangkan SVM unggul dalam mengenali santri yang diterima dengan *recall* 39% dan *F1-score* 0,32. Hal ini menunjukkan bahwa SVM lebih sensitif terhadap kelas minoritas dan dapat menjadi pilihan ketika fokus prediksi diarahkan pada identifikasi santri berpotensi. Penelitian lanjutan disarankan mempertimbangkan variabel tambahan non-akademik untuk meningkatkan akurasi dan relevansi model.

B. Saran

Pembahasan terkait penelitian ini masih sangat terbatas dan membutuhkan banyak masukan, saran untuk penulis selanjutnya adalah mengkaji lebih dalam dan secara komprehensif tentang Perbandingan SVM dan Random Forest dalam Memprediksi Kelulusan Santri Menuju Timur Tengah.

DAFTAR RUJUKAN

D. Ikasari, Y. Fajarwati and W. , "Analisis Sentimen dan Klasifikasi Tweets Berbahasa Indonesia Terhadap Transportasi Umum MRT Jakarta Menggunakan Naive Beyes Classifier," *Jurnal Ilmiah Informatika Komputer,* vol. 25, no. 1, pp. 64-75, 2020.

E. Haryatmi and S. H. Hervianti, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *Jurnal Rekayasa Sistem dan Teknologi Informasi (RESTI),* vol. 5, no. 2, pp. 386-392, 2021.

E. Novianto, S. and D. Prasetyo, "Perbandingan Metode Klsaifikasi Random Forest dan Support Vector Machine Dalam Memprediksi Capaian Mahasiswa," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika),* vol. 9, no. 4, pp. 1821-1833, 2024.

F. N. Rahman, T. A. Y. Siswa and R. , "Penerapan Metode GA-RU Pada Algoritma Random Forest Untuk Mengatasi Class Imbalance Data Beasiswa KIP-Kuliah," *Building of Informatics, Technology and Science (BITS),* vol. 6, no. 4, p. 2345−2357, 2025.

H. Andrianof, A. P. Gusman and O. A. Putra, "Implementasi Algoritma Random Forest untuk Prediksi Kelulusan Mahasiswa Berdasarkan Data Akademik: Studi Kasus di Perguruan Tinggi Indonesia," *Jurnal Sains Informatika Terapan (JSIT),* vol. 4, no. 1, pp. 24-28, 2025.

L. M. Huizen, M. Basyier and M. Idris, "Meningkatkan kinerja SVM: Dampak berbagai teknik seleksi fitur pada akurasi prediksi," *AITI (Jurnal Teknologi Informasi),* vol. 22, no. 1, pp. 1-14, 2025.

M. and A. Yudhistira, "Analisis Sentimen Petani Milenial Pada Media SosialX Menggunakan Algortitma Support Vector Machine (SVM)," *Jurnal Pendidikan dan Teknologi Indonesia (JPTI),* vol. 5, no. 3, pp. 845-857, 2025.

N. Maftucha, S. Salma, N. Rahmayuna and N. Wakhidah, "Perbandingan Algoritma Machine Learning Dalam Memprediksi Kelulusan Siswa," *Jurnal TEKNO KOMPAK,* vol. 19, no. 2, pp. 116-128, 2025.

P. H. Gunawan and I. V. Paputungan, "Deteksi Tingkat Potensi Kelulusan Calon Mahasiswa menggunakan Algoritma Random Forest," *Sistemasi: Jurnal Sistem Informasi,* vol. 14, no. 5, pp. 2055-2065, 2025.

R. A. Haristu and P. H. P. Rosa, "Penerapan Metode Random Forest untuk Prediksi Win Ratio Pemain Player Unknown Battleground," *MEANS (Media Informasi Analisa dan Sistem),* vol. 4, no. 2, pp. 120-128, 2019.

R. A. Permana and S. Sahara, "Metode Support Vector Machine Sebagai Penentu Kelulusan Mahasiswa pada Pembelajaran Elektronik," *Junrla Khatulistiwa Informatika,* vol. VII, no. 1, pp. 50-58, 2019.

R. Aulia, "Upaya Peningkatan Standar Kompetensi Kelulusan," *Journal of Education,* vol. 2, no. 1, pp. 122-132, 2022.

R. Munawarah, O. Soesanto and M. R. Faisal, "Penerapan Metode SUpport Vector Machine Pada Diagnosa Hepatitis," *Kumpulan jurnaL Ilmu Komputer (KLIK),* vol. 04, no. 01, pp. 103-113, 2016.

R. Rismaya, D. Yuniarto and D. Setiadi, "Penerapan Algoritma Machine Learning dalam Prediksi Prestasi Akademik Mahasiswa," *Router: Jurnal Teknik Informatika dan Terapan,* vol. 3, no. 1, pp. 15-23, 2025.

S. Data Mining Untuk Klasifikasi dan Klasterisasi Data, Bandung: Informatika, 2019.

S. Kusworo, N. A. Santoso and R. D. Kurniawan, "Prediksi Nilai Akhir Semester Siswa Menggunakan Algoritma Random Forest," *Jurnal Sains dan Ilmu Terapan,* vol. 7, no. 2, pp. 246-256, 2024.

S. Linawati, S. Nurdiani, K. Handayani and L. , "Prediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma Random Forest dan C4.5," *Khatulistiwa Informatika,* vol. VIII, no. 1, pp. 47-52, 2020.

S. Sobari, A. I. Purnamasari, A. Bahtiar and K. , "Meningkatkan Model Prediksi Kelulusan Santri Tahfidz di Pondok Pesantren Al-Kautsar Menggunakan Algoritma Random Forest," *JITET (Jurnal Informatika dan Teknik Elektro Terapan),* vol. 13, no. 1, pp. 732-738, 2025.

S. Susanti, D. N. Manurung, L. . J. C. Ginting, M. U. Nazha and R. Siregar, "Kualifikasi Penentuan Kelulusan dan Analisis Penilaian Pendidikan Melalui Kemampuan Peserta Didik," *Semantik: Jurnal Riset Ilmu Pendidikan, Bahasa dan Budaya,* vol. 2, no. 3, pp. 43-50, 2024.

T. A. Y. Siswa and W. J. Pranoto, "Implementasi Seleksi Fitur Information Gain Ratio Pada Algoritma Random Forest Untuk Model Data Klasifikasi Pembayaran Kuliah," *Dinamika Informatika,* vol. 15, no. 1, pp. 41-49, 2023.

W. and S. , "Kontribusi Keluarga Dalam Prediksi Mahasiswa Lulus Tepat Waktu Menggunakan Model Support Vector Machine," *Duta.com (Jurnal Ilmiah Teknologi Informasi dan Komunikasi),* vol. 17, no. 1, pp. 25-36, 2024.