

Penerapan Algoritma *Machine Learning* Untuk Penjurusan Siswa Sekolah Menengah Kejuruan Berdasarkan Nilai Raport dan *Psikotest*

Gian Maulana*, Afiyati

Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Jakarta
41519010028@student.mercubuana.ac.id

Penelitian ini bertujuan untuk mengevaluasi keefektifan sistem penjurusan siswa di Sekolah Menengah Kejuruan (SMK) dengan menggunakan empat algoritma machine learning, yaitu K-Nearest Neighbor, Naive Bayes, Support Vector Machine, dan Random Forest. Hasil penelitian menunjukkan bahwa algoritma Random Forest memiliki akurasi terbesar dibandingkan algoritma lainnya. Saat ini, proses penjurusan siswa di SMK tempat penelitian ini dilakukan, masih diproses secara manual melalui perhitungan nilai raport, nilai tes mandiri, dan nilai psikotes. Proses penjurusan secara manual tersebut memakan waktu yang cukup lama. Implementasi algoritma Random Forest dapat menjadi solusi untuk mempercepat proses penjurusan siswa di SMK tersebut. Algoritma Random Forest memiliki akurasi terbaik di antara algoritma lain, yaitu 37% hampir mencapai 38%.

Article History:

Received: Jan 9, 2023

Revised: Apr 9, 2023

Accepted: Apr 16, 2023

Published: Apr 24, 2023

Kata Kunci— Algoritma, Akurasi, Machine Learning, Perbandingan, Penjurusan

DOI: 10.22441/jitkom.2023.v7i1.008

I. PENDAHULUAN

Sekolah Menengah Kejuruan (SMK) adalah sekolah menengah yang memberikan pendidikan kejuruan bagi siswa setelah menyelesaikan pendidikan dasar. SMK biasanya memberikan pelatihan kepada siswa agar siap terjun ke dunia kerja atau melanjutkan ke perguruan tinggi. SMK memiliki program keahlian yang terfokus pada bidang tertentu, seperti teknik, industri, pariwisata, atau administrasi. Siswa di SMK dapat memilih program keahlian sesuai dengan minat dan bakat mereka. Cara penjurusan siswa di SMK bervariasi tergantung pada sekolah masing-masing. Biasanya, penjurusan siswa di SMK dilakukan dengan menggunakan beberapa nilai yaitu: nilai rapor, nilai tes mandiri dan nilai psikotes.

Penelitian ini dilakukan dengan menggunakan data yang diambil dari Sekolah Menengah Kejuruan Telkom Jakarta. Sekolah Menengah Kejuruan Telkom Jakarta adalah salah satu sekolah yang berada dalam naungan Yayasan Sandhykara Putra Telkom. Sekolah ini didirikan pada tahun 1992 dengan nama Sekolah Teknik Menengah (STM) Telkom Sandhy Putra Jakarta. Saat ini SMK Telkom Jakarta memiliki empat bidang studi keahlian, yaitu Teknik Transmisi, Teknik Jaringan Akses, Teknik Komputer Jaringan, dan Rekayasa Perangkat Lunak.

Proses penjurusan SMK Telkom Jakarta sendiri adalah dengan menggunakan nilai raport mereka selama lima semester ketika di SMP yang mereka tempuh, lalu proses selanjutnya adalah melalui kegiatan psikotes. Berdasarkan hasil psikotes tersebut, para siswa dan siswi pendaftar yang telah lolos akan mendapatkan jurusan mereka untuk mengikuti kegiatan pembelajaran di SMK Telkom Jakarta.

Penjurusan siswa sendiri merupakan suatu proses penempatan siswa ke dalam penjurusan tertentu, sehingga siswa-siswi mampu mengikuti kegiatan pembelajaran dan menyerapnya secara optimal. Namun karena sulitnya dalam melakukan pengklasifikasian jurusan siswa-siswi menjadi kendala bagi pihak sekolah dalam mempertimbangkan kriteria siswa-siswi yang diinginkan, sehingga menjadikan waktu yang tidak efisien karena harus menghitung nilai siswa-siswi satu per satu melalui hasil psikotes.

Dalam upaya menyelesaikan masalah yang dihadapi oleh sekolah, penelitian ini bertujuan untuk mengembangkan sistem pengklasifikasian jurusan bagi siswa yang telah menjalani tes psikologis di SMK Telkom Jakarta. Tujuan utama dari sistem ini adalah untuk memudahkan sekolah dalam menentukan jurusan yang sesuai bagi setiap siswa, serta untuk meningkatkan efisiensi waktu dalam proses pengambilan keputusan. Dengan demikian, metode machine learning seperti K-Nearest Neighbor, Naive Bayes, Support Vector Machine, dan Random Forest akan diterapkan untuk mengelompokkan siswa ke dalam jurusan yang sesuai dengan kemampuan dan minat mereka [1].

Salah satu solusi yang dicadangkan adalah menggunakan metode algoritma Naive Bayes untuk mengklasifikasikan pemilihan jurusan di SMK Telkom Jakarta. Metode ini dapat digunakan untuk mengelompokkan data yang belum terklasifikasi ke dalam kelompok yang sesuai berdasarkan kesamaan yang dimilikinya. Proses ini dilakukan dengan cara meminimalisir nilai dari fungsi objektif yang digunakan [2]. Dalam klasifikasi Naive Bayes, asumsi yang digunakan adalah bahwa keberadaan atau tidaknya suatu ciri dari sebuah kelas tidak memiliki hubungan dengan ciri dari kelas lainnya [3].

Berdasarkan hasil analisis yang telah dilakukan sebelumnya menggunakan algoritma Naïve Bayes, nilai akurasi model tersebut dapat diketahui sebesar 93,75%, dengan nilai precision sebesar 83,33% dan nilai recall sebesar 100% [4].

Metode K-Nearest Neighbor merupakan salah satu metode klasifikasi yang mengandalkan kemiripan dengan data latih yang telah diberikan. Setiap sampel data akan diuji validitasnya untuk menghindari terjadinya data outliers yang dapat mempengaruhi hasil akhir. Dengan demikian, pembobotan dari setiap data dapat memberikan hasil yang bisa diandalkan [5]. Hasil evaluasi menunjukkan bahwa algoritma Nearest Neighbor sangat akurat untuk digunakan dalam proses pemilihan jurusan melalui data mining. Sehingga, algoritma K-Nearest Neighbor dapat memberikan solusi bagi masalah penentuan kelayakan siswa untuk memasuki suatu jurusan [6]. Algoritma klasifikasi ini termasuk dalam kelompok algoritma instance-based classifier, yaitu algoritma yang hanya melakukan proses klasifikasi saat ada objek yang ingin diketahui kelasnya. Sehingga, algoritma ini tidak membangun suatu model dari data pelatihan, melainkan data pelatihan itu sendiri yang menjadi representasi dari knowledge model klasifikasi tersebut [7].

II. LITERATURE REVIEW

Penelitian terdahulu telah menggunakan algoritme machine learning antara lain Naïve Bayes, Support Vector Machine, dan K-Nearest Neighbor. Namun, penelitian terdahulu masih terdapat kekurangan yang menjadi potensi untuk diteliti selanjutnya, seperti terlihat pada Tabel 1. Penelitian ini menggunakan ketiga algoritme tersebut dan menambahkan algoritme Random Forest yang belum digunakan oleh penelitian sebelumnya.

Tabel 1. Penelitian terdahulu dan potensi penelitian selanjutnya

Referensi	Metode	Hasil	Potensi
[1]	Naïve Bayes (NB)	Akurasi: 33,34% Presisi: 21,8% Recall: 33,77%	Nilai akurasi rendah. Algoritme NB dengan teknik Unsupervised Discretization atau metode Fuzzy menjadi potensi penelitian selanjutnya.
[4]	Naïve Bayes (NB)	Akurasi: 93,5% Presisi: 83,33% Recall: 100%	Algoritme NB digunakan untuk penentuan keputusan peminatan siswa, namun menggunakan jenis nilai berbeda.
[6]	K-Nearest Neighbor (KNN)	Akurasi: 67,95%	Hasil menunjukkan bahwa sistem rekomendasi masih kurang optimal dalam memberikan rekomendasi penjurusan.
[8]	Support Vector Machine (SVM) & Naïve Bayes (NB)	Akurasi: 90% Presisi: 91,33% Recall: 91,77%	Penelitian selanjutnya disarankan menggunakan dataset, atribut, dan jumlah data yang lebih variatif serta melakukan uji

			statistik untuk menguji validitas hasil.
[7]	K-Nearest Neighbor (KNN)	Akurasi: 49,29% Presisi: 44,48% Recall: 83,89%	Algoritma KNN dengan metode Euclidean Distance dapat meningkatkan prediksi jurusan siswa.

III. METODOLOGI PENELITIAN

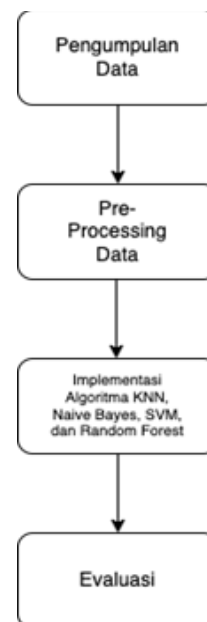
Penelitian ini menggunakan metode machine learning, metode machine learning sendiri dapat digunakan untuk memprediksi kelas suatu objek dengan menggunakan data yang telah dikumpulkan sebelumnya. Machine learning akan mempelajari pola-pola dari data tersebut dan menggunakannya untuk membuat model prediksi. Kemudian, model tersebut dapat digunakan untuk memprediksi kelas dari objek baru yang tidak diketahui kelasnya. Machine learning juga dapat mengoptimalkan model prediksi tersebut secara terus-menerus seiring dengan bertambahnya data baru yang tersedia. Dengan demikian, metode machine learning dapat memberikan kemampuan untuk memprediksi kelas suatu objek dengan akurasi yang tinggi.

A. Penelitian Kuantitatif

Untuk menyelesaikan masalah yang ada, jenis penelitian yang dipilih adalah penelitian kuantitatif. Alasannya adalah karena dalam penelitian ini menggunakan data yang disajikan dalam bentuk tabel, sehingga dapat dilakukan perhitungan statistik.

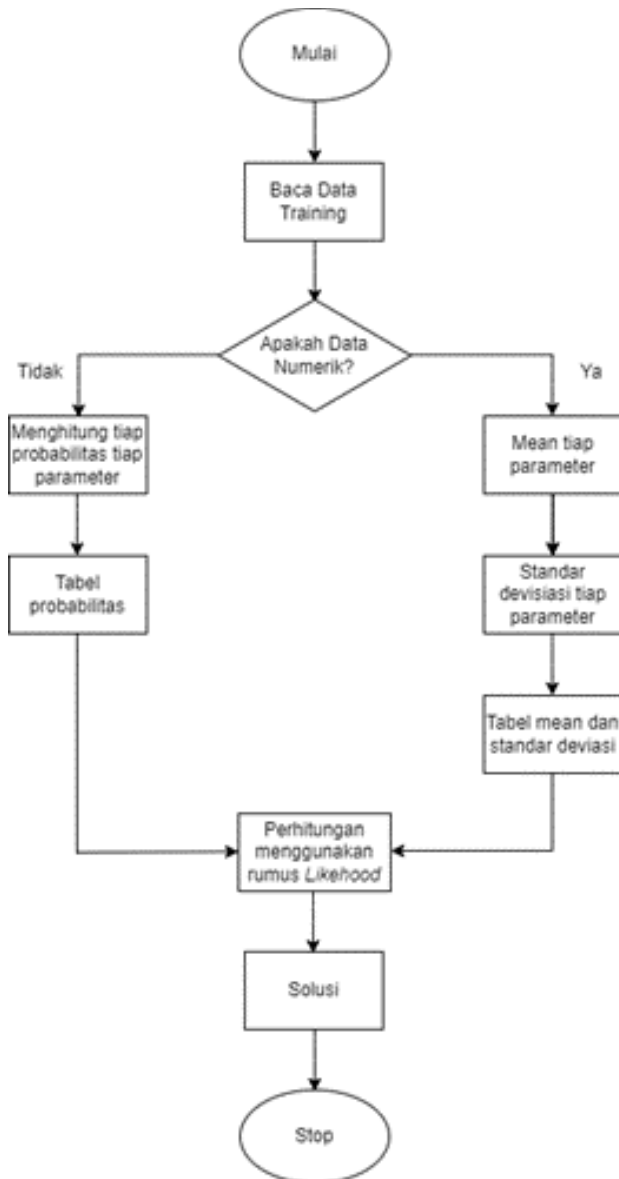
B. Tahapan Penelitian dan Implementasi Algoritma

Pada tahapan penelitian ini, tahapan penelitian yang dilakukan adalah pengumpulan data, pre-processing, implementasi algoritma dan evaluasi seperti terlihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Dalam tahap implementasi algoritma, penulis menggunakan algoritma Naive Bayes untuk memproses data siswa yang lulus dari Sekolah Menengah Pertama (SMP) pada tahun angkatan 2019-2020. Untuk menguji keefektifan algoritma tersebut, kami menggunakan tools Google Collab. Hasil dari pengujian ini kemudian akan digunakan untuk menentukan jurusan yang sesuai bagi siswa pendaftar. Flowchart implementasi algoritma Naive Bayes dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Naive Bayes

Berdasarkan Gambar 2, dapat dilihat bahwa proses pengklasifikasian naïve bayes dimulai dengan membaca data training, jika data tersebut merupakan data numerik maka akan dilanjutkan ke tahap perhitungan nilai mean dan standar deviasi dari tiap parameter, sedangkan apabila data training tersebut bukan data numerik maka akan masuk ke tahap perhitungan probabilitas dari tiap parameter. Setelah itu akan dilanjutkan ke tahap perhitungan nilai likelihood dimana dari hasil perhitungan tersebut akan diperoleh solusi.

Gambar 3 ditampilkan flowchart atau alur Algoritma K-Nearest Neighbor (KNN) untuk memberikan hasil pada sistem penjurusan melalui hasil tes mandiri dan psikotes. Gambar 3 menampilkan penggunaan rumus Euclidean Distance untuk menghitung jarak antara dua titik dalam ruang Euclidean, dengan formula sebagai berikut :

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + \dots + (n_2 - n_1)^2} \quad (1)$$

Di mana :

- d adalah jarak Euclidean
- x1, y1, ..., n1 adalah koordinat dari titik pertama
- x2, y2, ..., n2 adalah koordinat dari titik kedua

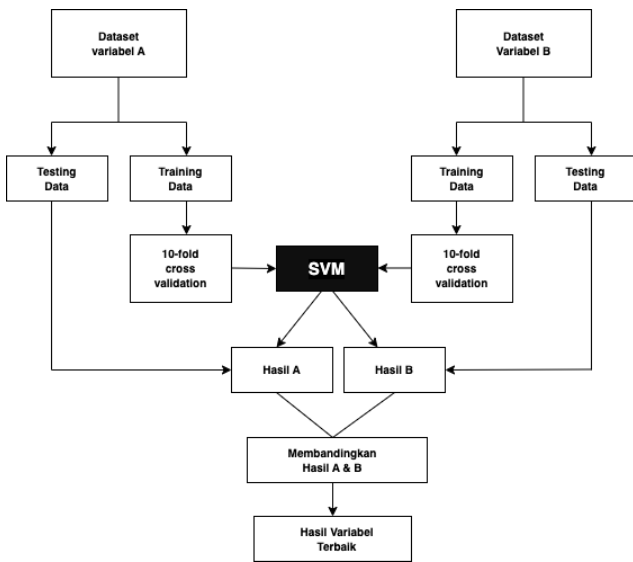
Jika hanya menggunakan dua dimensi (misalnya, x dan y), rumus tersebut dapat ditulis sebagai berikut:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2)$$



Gambar 3. Flowchart KNN

Berikut dibawah ini *flowchart* implementasi algoritma SVM seperti yang terlihat pada Gambar 4.



Gambar 4. Flowchart SVM

Gambar 4 menunjukkan alur input dataset variabel A merupakan dataset yang terdiri dari atribut *open*, *high*, dan *low* dengan label *close*, sedangkan dataset variabel B dataset yang terdiri dari atribut *open*, *high*, *low* dan *factory news* dengan label *close*. Kedua dataset dengan variabel beda tersebut di proses dengan model algoritma SVM dengan 10-fold *cross validation* sehingga didapat 2 hasil, proses selanjutnya membandingkan dua hasil tersebut mana yang terbaik.

Berikut ini adalah flowchart implementasi algoritma Random Forest seperti yang terlihat pada Gambar 5.

Pada Gambar 5 dapat dilihat bahwa proses algoritma *Random Forest* dalam menentukan hasil yang baik. Proses selalu diawali dengan input data lalu melakukan *pre-processing* data, lalu melakukan pembagian data dan men-training data sekaligus menguji data dengan melakukan testing data sehingga dapat menghasilkan akurasi yang maksimal dari pengujian yang akan dilakukan.

IV. HASIL DAN ANALISA

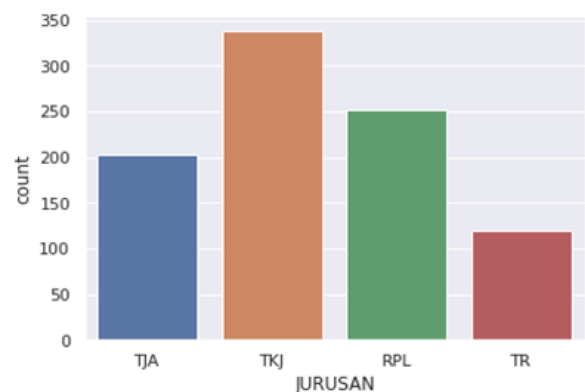
Tahapan ini diawali dengan pengumpulan data. Pengumpulan data dilakukan dengan mengumpulkan data siswa yang berstatus lulus dari Sekolah Menengah Pertama (SMP) tahun angkatan 2019 sampai 2020. Dari data yang sudah ada, terdapat atribut-atribut sebagai berikut:

1. Tahun Angkatan
2. Nama
3. Jenis Kelamin
4. Asal Sekolah
5. Matematika 1-5
6. Bahasa Indonesia 1-5
7. Bahasa Inggris 1-5
8. Kepribadian
9. IQ
10. Psikotes



Gambar 5. Flowchart Random Forest

Pada gambar berikut, dapat kita lihat jumlah perbandingan dari siswa tetap dalam penjurusan siswa SMK Telkom Jakarta. Bahwa jurusan TKJ dan RPL sangat mendominasi dalam terhitung dua tahun terakhir. Sedangkan jurusan TJA dan TR memiliki jumlah siswa yang lebih sedikit seperti yang terlihat pada Gambar 6 berikut.



Gambar 6. Target Jurusan

Tabel Berdasarkan pengujian yang telah dilakukan, pada bagian ini penulis akan menjabarkan hasil dari serangkaian pengujian menggunakan keempat algoritma berdasarkan tiap jurusan yang telah berhasil menampilkan akurasi.

A. Hasil K-Nearest Neighbor

Hasil Akurasi, Precision, Recall dan F1-Score untuk jurusan RPL menggunakan metode split validation pada data train serta data test menggunakan algoritma KNN, seperti terlihat pada Tabel 2.

Tabel 2. Hasil Split Validation pada Algoritma KNN jurusan RPL

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	41%	37%	28%	32%
Test	37%	31%	24%	27%

Kesimpulan pada tabel 2 adalah bahwa pada jurusan RPL memiliki tingkat akurasi prediksi yang cukup rendah, yaitu sekitar 37-41% pada train dan test report. Selain itu, algoritma tersebut juga memiliki tingkat precision, recall, dan F1-score yang relatif rendah, yaitu sekitar 31-37%, 24-28%, dan 27-32%.

Tabel 3. Hasil Split Validation Algoritma KNN jurusan TJA

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	41%	34%	13%	19%
Test	37%	23%	10%	14%

Seperti yang dilihat pada Tabel 3 dimiliki kesimpulan bahwa pada jurusan TJA memiliki tingkat akurasi prediksi yang cukup rendah, yaitu sekitar 37-41% pada train dan test report. Selain itu, algoritma tersebut juga memiliki tingkat precision, recall, dan F1-score yang relatif lebih rendah dari jurusan RPL, yaitu sekitar 23-34%, 10-13%, dan 14-19%.

Tabel 4. Hasil Split Validation Algoritma KNN jurusan TKJ

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	41%	34%	13%	19%
Test	37%	23%	10%	14%

Dapat dilihat pada Tabel 4 bahwa jurusan TKJ memiliki tingkat akurasi prediksi yang juga cukup rendah dari kedua jurusan sebelumnya, yaitu 37-41% pada train dan test report. Namun, algoritma tersebut memiliki tingkat precision, recall, dan F1-score yang sedikit lebih tinggi dari jurusan RPL dan TJA yaitu sekitar 39-41%, 73-81%, dan 51-57%.

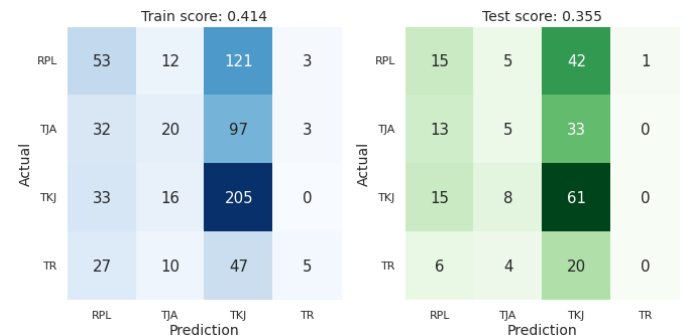
Tabel 5. Hasil Split Validation Algoritma KNN jurusan TKJ

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	41%	34%	13%	19%
Test	37%	23%	10%	14%

Dapat dilihat pada Tabel 5 bahwa jurusan TR memiliki tingkat akurasi prediksi yang paling rendah dari ketiga jurusan sebelumnya, sekitar 39-41%, 73-81%, dan 51-57% pada tingkat precision, recall, dan F1-score.

Confusion Matrix dengan menggunakan split validation terlihat pada Gambar 7 dibawah ini. Dapat disimpulkan hasil

dari Confusion Matrix yang telah diuji adalah bahwa train score pada algoritma KNN adalah sebesar 41% dan test score adalah 35%. Dapat dikatakan bahwa tingkat akurasi model tersebut relatif rendah. Angka-angka ini menunjukkan bahwa model tersebut tidak terlalu efektif dalam memprediksi label suatu data, baik pada data training maupun pada data testing.



Gambar 7. Hasil Algoritma KNN

B. Naïve Bayes

Hasil Akurasi, Precision, Recall dan F1-Score menggunakan metode split validation pada data train serta data test menggunakan Naïve Bayes dapat dilihat pada Tabel 5.

Tabel 6. Hasil Split Validation Algoritma Naive Bayes jurusan RPL

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	38%	37%	27%	31%
Test	27%	19%	13%	15%

Kesimpulan pada Tabel 6 adalah bahwa pada jurusan RPL memiliki tingkat akurasi prediksi yang cukup rendah, yaitu sekitar 27 -38% pada train dan test report. Selain itu, algoritma tersebut juga memiliki tingkat precision, recall, dan F1-score yang relatif rendah, yaitu sekitar 19-37%, 13-27%, dan 15-31%.

Tabel 7. Hasil Split Validation Algoritma Naive Bayes jurusan TJA

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	38%	26%	43%	33%
Test	27%	17%	25%	20%

Pada Tabel 7 memiliki kesimpulan bahwa pada jurusan TJA memiliki tingkat akurasi prediksi yang cukup rendah, yaitu sekitar 27-38% pada train dan test report. Selain itu, algoritma tersebut juga memiliki tingkat precision, recall, dan F1-score yang relatif lebih rendah dari jurusan RPL, yaitu sekitar 17-26%, 25-43%, dan 20-33%.

Tabel 8. Hasil Split Validation Algoritma Naive Bayes jurusan TKJ

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	38%	48%	56%	51%
Test	27%	37%	49%	42%

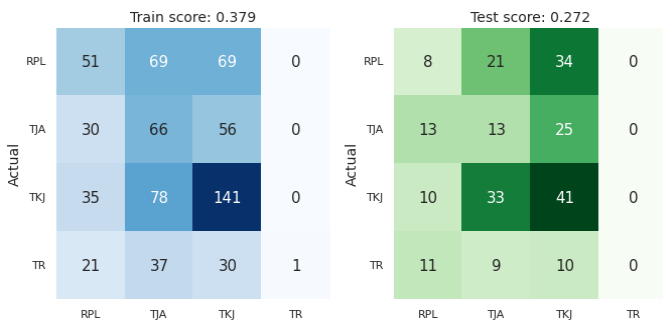
Dapat dilihat pada Tabel 8 bahwa pada jurusan TKJ memiliki tingkat akurasi prediksi yang juga cukup rendah dari

kedua jurusan sebelumnya, yaitu 27-38% pada *train* dan *test report*. Namun, algoritma tersebut memiliki tingkat *precision*, *recall*, dan *F1-score* yang sedikit lebih tinggi dari jurusan RPL dan TJA yaitu sekitar 37-48%, 49-56%, dan 42-51%.

Tabel 9. Hasil Split Validation Algoritma Naive Bayes jurusan TR

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	38%	48%	56%	51%
Test	27%	37%	49%	42%

Seperti yang terlihat pada Tabel 9 bahwa jurusan TR memiliki tingkat akurasi prediksi yang paling buruk dari ketiga jurusan sebelumnya, sekitar 0-100%, 0-1%, dan 0-2% pada tingkat *precision*, *recall*, dan *F1-score*. Confusion Matrix dengan menggunakan split validation dapat dilihat pada Gambar 8.



Gambar 8. Hasil Algoritma Naive Bayes

C. Support Vector Machine

Hasil Akurasi, Precision, Recall dan F1-Score menggunakan metode split validation pada data train serta data test menggunakan Support Vector Machine pada jurusan RPL.

Tabel 10. Hasil Split Validation Algoritma SVM jurusan RPL

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	41%	38%	22%	28%
Test	35%	24%	16%	19%

Kesimpulan pada Tabel 10 adalah bahwa pada jurusan RPL memiliki tingkat akurasi prediksi yang rendah, yaitu sekitar 35-41% pada *train* dan *test report*. Selain itu, algoritma tersebut juga memiliki tingkat *precision*, *recall*, dan *F1-score* yang relatif rendah, yaitu sekitar 24-38%, 22-16%, dan 19-28%.

Tabel 11. Hasil Split Validation Algoritma SVM jurusan TJA

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	41%	0%	0%	0%
Test	35%	0%	0%	0%

Kesimpulan pada Tabel 11 memiliki kesimpulan bahwa pada jurusan TJA memiliki tingkat akurasi prediksi yang sangat buruk, yaitu memiliki 0% pada tingkat *precision*, *recall*, dan *F1-Score*.

Tabel 12. Hasil Split Validation Algoritma SVM jurusan TKJ

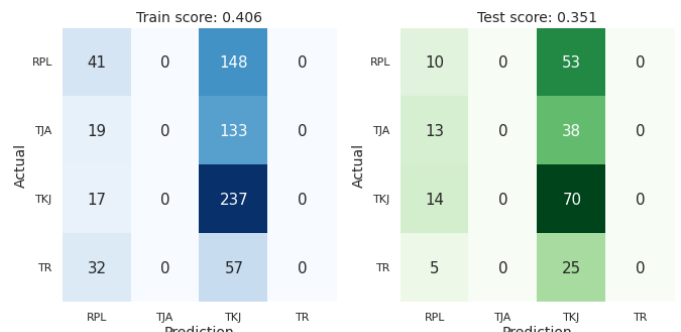
Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	41%	41%	93%	57%
Test	35%	38%	83%	52%

Dapat dilihat bahwa pada Tabel 12 jurusan TKJ memiliki tingkat akurasi prediksi yang cukup baik dari kedua jurusan sebelumnya, pada *precision*, *recall*, dan *F1-score* yang sedikit lebih tinggi dari jurusan RPL dan TJA yaitu sekitar 38-41%, 83-93%, dan 52-57%.

Tabel 13. Hasil Split Validation Algoritma SVM jurusan TR

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	41%	0%	0%	0%
Test	35%	0%	0%	0%

Pada Tabel 13 dapat ditarik kesimpulan bahwa jurusan TR memiliki tingkat akurasi prediksi yang paling buruk sama seperti jurusan TJA, yakni 0% pada tingkat *precision*, *recall*, dan *F1-score*. Confusion Matrix menggunakan split validation dapat dilihat pada Gambar 9 berikut.



Gambar 9. Hasil Algoritma SVM

Hasil evaluasi model algoritma *Support Vector Machine* pada Gambar 9 menunjukkan bahwa tingkat akurasi model pada data training adalah sebesar 46% dan pada data testing adalah 35%. Tingkat akurasi tersebut dapat dikatakan juga masih sangat rendah dan memiliki tingkat akurasi yang hampir mirip dengan algoritma KNN, yang menunjukkan bahwa model tersebut tidak efektif dalam memprediksi label suatu data.

D. Random Forest

Hasil Akurasi, Precision, Recall dan F1-Score menggunakan metode split validation pada data train serta data test menggunakan *Random Forest* pada jurusan RPL dapat dilihat pada Tabel 14.

Tabel 14. Hasil Split Validation Algoritma Random Forest jurusan RPL

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	87%	92%	90%	91%
Test	37%	31%	35%	33%

Kesimpulan pada Tabel 13 adalah bahwa pada jurusan RPL memiliki tingkat akurasi prediksi yang cukup baik, yaitu sekitar

37-87% pada *train* dan *test report*. Selain itu, algoritma tersebut juga memiliki tingkat *precision*, *recall*, dan *F1-score* yang relatif stabil, yaitu sekitar 31-92%, 31-92%, dan 33-91%.

Tabel 15. Hasil Split Validation Algoritma Random Forest jurusan TJA

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	87%	96%	81%	88%
Test	37%	47%	18%	26%

Tabel 15 memiliki kesimpulan bahwa pada jurusan TJA memiliki tingkat akurasi prediksi yang cukup rendah, yaitu sekitar 37-87% pada *train* dan *test report*. Selain itu, algoritma tersebut juga memiliki tingkat *precision*, *recall*, dan *F1-score* yang relatif sedikit lebih rendah dari jurusan RPL, yaitu sekitar 47-96%, 18-81%, dan 26-88%.

Tabel 16. Hasil Split Validation Algoritma Random Forest jurusan TKJ

Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	87%	79%	100%	88%
Test	37%	40%	63%	49%

Dapat dilihat pada Tabel 16 bahwa pada jurusan TKJ memiliki tingkat akurasi prediksi yang cukup baik dari kedua jurusan sebelumnya, pada *precision*, *recall*, dan *F1-score* yang sedikit lebih tinggi dari jurusan RPL dan TJA yaitu sekitar 40-70%, 63-100%, dan 49-88%.

Tabel 17. Hasil Split Validation Algoritma Random Forest jurusan TR

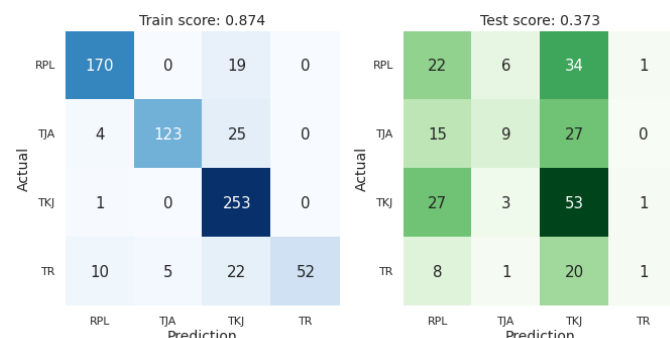
Eksperimen	Akurasi	Precision	Recall	F1-Score
Train	87%	100%	58%	74%
Test	37%	33%	3%	6%

Pada Tabel 17 dapat dilihat bahwa jurusan TR memiliki tingkat akurasi prediksi yang paling rendah jika dibandingkan dengan jurusan lainnya, yakni 33-100% pada tingkat *precision*, 3-58% pada *recall*, dan 6-74% pada *F1-score*.

Confusion Matrix menggunakan split validation dapat dilihat pada Gambar 11. Berdasarkan hasil evaluasi dari model algoritma *Random Forest*, dapat ditarik kesimpulan bahwa tingkat akurasi model pada data training adalah sebesar 87% dan pada data testing adalah 37%. Tingkat akurasi tersebut dapat dikatakan paling tinggi dari algoritma pembandingan lainnya. Namun meski demikian, tingkat akurasi tersebut masih dapat dikatakan rendah dan menunjukkan bahwa model tersebut belum efektif untuk memprediksi suatu data.

Empat algoritma yang diuji tidak menunjukkan hasil yang positif, dengan akurasi di bawah 50%. Namun, algoritma *Random Forest* memiliki akurasi terbaik di antara ketiga algoritma lainnya (*K-Nearest Neighbor*, *Naive Bayes*, dan *Support Vector Machine*). Algoritma *Random Forest* menunjukkan akurasi sebesar 37%, hampir mendekati 38%. Algoritma lainnya memiliki hasil yang lebih rendah, dengan *K-*

Nearest Neighbor memiliki akurasi sebesar 35%, *Naive Bayes* sebesar 27%, dan *Support Vector Machine* sebesar 35,5%.



Gambar 10. Hasil Algoritma Random Forest

Penulis menyimpulkan bahwa akurasi machine learning yang rendah mungkin disebabkan oleh kualitas data yang buruk. Data yang tidak akurat atau tidak relevan dapat mempengaruhi kemampuan model machine learning untuk memprediksi dengan benar. Jika data yang digunakan untuk melatih model tidak mewakili data yang akan diuji, maka model tersebut mungkin tidak akan mampu menangani data baru dengan baik dan menghasilkan akurasi yang rendah. Oleh karena itu, penting untuk memastikan bahwa data yang digunakan untuk melatih model machine learning berkualitas tinggi dan relevan dengan tujuan yang ingin dicapai.

V. KESIMPULAN

Penerapan algoritma machine learning untuk penjurusan siswa SMK Telkom Jakarta berdasarkan nilai raport dan psikotest belum memberikan hasil yang akurat. Meskipun empat algoritma telah digunakan, hasilnya masih belum cukup karena kualitas data yang belum baik dan tidak relevan. Algoritma *Random Forest* memiliki akurasi terbaik di antara algoritma lain, yaitu 37% hampir mencapai 38%. Algoritma lain memiliki akurasi yang lebih rendah. Dapat disimpulkan bahwa kualitas data sangat mempengaruhi hasil prediksi yang diinginkan, sehingga algoritma apapun tidak akan memberikan hasil maksimal jika tidak menggunakan data berkualitas.

Untuk meningkatkan akurasi prediksi penjurusan, saran penulis adalah menambah fitur yang berpengaruh dan memperhatikan kualitas data untuk relevansi dengan penelitian. Cara lain untuk meningkatkan akurasi machine learning adalah dengan memperbaiki data yang digunakan, seperti menghilangkan data yang tidak akurat, menghilangkan missing value, menyeleksi fitur yang relevan, dan menggunakan data yang lebih banyak. Dengan demikian, diharapkan dapat meningkatkan akurasi model machine learning yang rendah karena kualitas data yang belum memadai untuk memprediksi penjurusan siswa.

DAFTAR PUSTAKA

- [1] A. Z. Mafakhir and A. Solichin, "Penerapan Metode Naïve Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta," *Fountain Informatics J.*, vol. 5, no. 1, p. 21, 2020, doi: 10.21111/fij.v5i1.4007.

- [2] F. Ekawati, "Algoritma Naïve Bayes Untuk Penentuan Jurusan Pada Siswa Madrasah Aliyah," *Technol. J. Ilm.*, vol. 9, no. 1, p. 42, 2018, doi: 10.31602/tji.v9i1.1101.
- [3] F. E. Prabowo and A. Kodar, "Analisis Prediksi Masa Studi Mahasiswa Menggunakan Algoritma Naïve Bayes," *J. Ilmu Tek. dan Komput.*, vol. 3, no. 2, p. 147, 2019, doi: 10.22441/jitkom.2020.v3.i2.008.
- [4] D. Putra and A. Wibowo, "Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, pp. 84–92, 2020. J. W. Cooper, "The singleton pattern," in *Introduction to Design Patterns in C#*. Boston, MA: Addison Wesley, 2002, pp. 145 – 148.
- [5] M. Mughniy, R. C. Wihandika, and B. H. Prasetyo, "Sistem Rekomendasi Psikotes untuk Penjurusan Siswa SMA menggunakan Metode Modified K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 282–287, 2018.
- [6] S. Saepudin, M. Muslih, P. Studi Sistem Informasi, U. Nusa Putra, J. Raya Cibolang Kaler No, and K. Sukabumi, "Pemilihan Jurusan Dengan Metode K-Nearest Neighbor Untuk Calon Siswa Baru," *J. Rekayasa Teknol. Nusa Putra*, vol. 5, no. 2, pp. 15–19, 2019.
- [7] M. Bayu Agung Rahmadi, "Sistem Penjurusan IPA / IPS Menggunakan Algoritma K-Nearest Neighbor Pada SMA Muhammadiyah 13 Jakarta," *Semin. Nas. Sains Teknol. Inf.*, pp. 300–305, 2019.
- [8] O. Arifin and T. B. Sasongko, "Analisa perbandingan tingkat performansi metode support vector machine dan naïve bayes classifier," *Semin. Nas. Teknol. Inf. dan Multimed.* 2018, vol. 6, no. 1, pp. 67–72, 2018.