SISTEM PEMETAAN JALUR KARIR LULUSAN SMK MENGGUNAKAN METODE RANDOM FOREST DAN K-NEAREST NEIGHBORS

TESIS

Oleh:
NIA KURNIAWATI IMAMI
NIM. 220605220011

PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

SISTEM PEMETAAN JALUR KARIR LULUSAN SMK MENGGUNAKAN METODE RANDOM FOREST DAN K-NEAREST NEIGHBORS

TESIS

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Magister Komputer (M.Kom)

Oleh:
NIA KURNIAWATI IMAMI
NIM. 220605220011

PROGRAM STUDI MAGISTER INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2025

SISTEM PEMETAAN JALUR KARIR LULUSAN SMK MENGGUNAKAN METODE RANDOM FOREST DAN K-NEAREST NEIGHBORS

TESIS

Oleh:
NIA KURNIAWATI IMAMI
NIM. 220605220011

Telah diperiksa dan disetujui untuk di uji :
Tanggal 28 Oktober 2025

Pembimbing I	Pembimbing II
Dr. Mokhamad Amin Hariyadi, M.T	Dr. Ir. Yunifa Miftachul Arif, S.ST., M.T
NIP. 19670118 200501 1 001	NIP 19830616 201101 1 004

Mengetahui
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Prof. Dr. Ir. Muhammad Faisal, S.Kom., MT
NIP. 19740510 200501 1 007

SISTEM PEMETAAN JALUR KARIR LULUSAN SMK MENGGUNAKAN METODE RANDOM FOREST DAN K-NEAREST NEIGHBORS

TESIS

Oleh:
NIA KURNIAWATI IMAMI
NIM. 220605220011

Telah Dipertahankan di Depan Dewan Penguji Tesis
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Magister Komputer (M.Kom)
Tanggal: 12 November 2025

Susunan Dewan Penguji			**Tanda Tangan**
Penguji I	:	Dr. M. Ainul Yaqin, M.Kom NIP. 19761013 200604 1 004	( )
Penguji II	:	Dr. Agung Teguh Wibowo Almais, M.T NIP. 19860301 202321 1 016	( )
Pembimbing I	:	Dr. Mokhamad Amin Hariyadi, M.T NIP. 19670118 200501 1 001	( )
Pembimbing II	:	Dr. Ir. Yunifa Miftachul Arif, S.ST., M.T NIP 19830616 201101 1 004	( )

Mengetahui dan Mengesahkan
Ketua Program Studi Magister Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Prof. Dr. Ir. Muhammad Faisal, S.Kom., MT
NIP. 19740510 200501 1 007

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini :

Nama	: Nia Kurniawati Imami
NIM	: 220605220011
Program Studi	: Magister Informatika
Fakultas	: Sains dan Teknologi

Menyatakan dengan sebenar-benarnya bahwa Tesis yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupan pengambilalihan data, tulisan, atau pikiran orang lain yang saya akui sebagai tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada Daftar Pustaka

Apabila di kemudian hari terbukti atau dapat dibuktikan Tesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut

Malang, 24 Desember 2025

Yang menyatakan

Nia Kurniawati Imami

NIM. 220605220011

MOTTO

“Jangan tunggu sampai kamu merasa siap, beranilah melangkah dan percayalah bahwa Allah SWT akan menunjukkan kekuatanNya dan membuka jalan terbaik di depanmu. “

HALAMAN PERSEMBAHAN

Alhamdulillah dengan penuh rasa syukur kehadirat Allah SWT, Tesis ini dipersembahkan sebagai wujud ikhtiar dan pengabdian dalam menuntut ilmu.

Tesis ini penulis persembahkan dengan penuh cinta dan hormat kepada kedua orang tua – *Imam Subroto-Soetjiati* dan mertua *M. Harmoko-Mistikaningsih* tercinta yang senantiasa memberikan doa, kasih sayang, dukungan moral, dan pengorbanan tanpa batas.

Dengan penuh rasa syukur, penulis juga mempersembahkan tesis ini kepada suami tercinta *Wahyu Wuriadi* yang senantiasa memberikan dukungan, pengertian, kesabaran, dan doa dalam setiap langkah penulis. Tak lupa juga, *My Little Boy Atthar Mauza Satria*, yang menjadi sumber inspirasi, kebahagiaan, dan motivasi terbesar bagi penulis. Semoga karya ini kelak menjadi teladan bahwa menuntut ilmu adalah bagian dari ibadah dan perjuangan hidup.

Penulis juga mempersembahkan karya ini kepada keluarga besar lain yang tidak dapat disebutkan satu persatu, yang selalu memberikan motivasi, perhatian, dan dukungan dalam setiap langkah perjuangan penulis.

Ucapan terima kasih dan penghargaan setinggi-tingginya penulis sampaikan kepada para dosen pembimbing (Dr. M. Amin Hariyadi, M.T - Dr. Ir. Yunifa Miftachul Arif, S.ST., M.T) dan dosen penguji (Dr. M. Ainul Yaqin, M. Kom - Dr. Agung Teguh Wibowo Almais, M.T) serta seluruh dosen yang telah memberikan ilmu, bimbingan, serta inspirasi selama proses pendidikan dan penyusunan tesis ini.

Tidak lupa, tesis ini penulis persembahkan kepada rekan-rekan seperjuangan yang telah menjadi sahabat diskusi, pemberi semangat, dan sumber motivasi selama masa studi. Semoga kebersamaan dan kerja keras yang terjalin menjadi kenangan dan manfaat yang berharga.

Akhir kata, semoga tesis ini dapat memberikan manfaat bagi dunia pendidikan, khususnya pendidikan kejuruan, serta menjadi kontribusi nyata dalam pengembangan ilmu pengetahuan dan teknologi.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tesis yang berjudul **"Sistem Pemetaan Jalur Karir Lulusan SMK Menggunakan Metode Random Forest Dan K-Nearest Neighbors"** sebagai salah satu syarat untuk memperoleh gelar Magister.

Tesis ini disusun sebagai upaya untuk mendukung peningkatan kualitas penyaluran lulusan SMK ke dunia kerja melalui pemanfaatan metode klasifikasi, serta memperkuat peran Bursa Kerja Khusus (BKK) dalam menyediakan layanan pemetaan jalur karir yang lebih tepat dan berbasis data.

Penulis menyampaikan terima kasih kepada dosen pembimbing dan seluruh pihak yang telah memberikan bimbingan, dukungan, serta doa selama proses penyusunan tesis ini. Ucapan terima kasih juga penulis sampaikan kepada keluarga atas dukungan dan motivasi yang senantiasa mengiringi.

Penulis menyadari bahwa tesis ini masih jauh dari kesempurnaan. Oleh sebab itu, kritik dan saran yang membangun sangat diharapkan. Semoga tesis ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan dan dunia pendidikan kejuruan.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

ABSTRAK

Imami, Nia Kurniawati, 2025, **Sistem Pemetaan Jalur Karir Lulusan SMK dengan Menggunakan Metode Random Forest dan K-Nearst Neighbors.** Program Magister Informatika Universitas Islam Negeri Maulana Malik Ibrahim, Pembimbing: (1) Dr. M. Amin Hariyadi, M.T (2) Dr. Ir. Yunifa Miftachul Arif, S.ST., M.T

Kata Kunci : Sistem Pemetaan, Jalur Karir Lulusan SMK, Random Forest, K-Nearest Neighbors.

Sekolah Menengah Kejuruan merupakan sekolah yang dipersiapkan untuk siswa yang siap kerja setelah lulus sekolah. Oleh karenanya selama proses pembelajaran tidak hanya teori-teori saja yang diberikan melainkan juga praktik bahkan ada program Praktek Kerja Lapangan (PKL) atau magang dimana siswa akan mendapatkan pengalaman kerja nyata di lapangan. Dengan pembekalan yang sedemikian rupa diharapkan siswa lulusan SMK mampu bersaing di dunia kerja sesuai dengan bidang keahliannya. Namun menurut data Statistic Indonesia tahun 2022 mengenai jumlah pengangguran mencapai 8.42% tertinggi dibanding lulusan sekolah lainnya dan tercatat 60% siswa lulusan SMK bekerja di luar bidang keahliannya. Oleh karena itu dibutuhkan suatu algoritma yang dapat membantu memprediksi jalur karir siswa lulusan SMK, pada penelitian ini menggunakan algoritma random forest dan K-nearst neighbors. Data yang digunakan pada penelitian ini berasal dari Bursa Kerja Khusus (BKK) yang memuat 8 kompetensi pokok siswa SMK. Pada pengujian hyperparameter tunning dilakukan dengan kombinasi data training sebesar 60% sampai 90% dan data testing antara 40% sampai 10%. Dari pengujian tersebut didapat hasil terbaik pada kombinasi data training sebesar 80% dan data testing sebesar 20%. Lalu kombinasi tersebut diterapkan pada data testing pada masing-masing algoritma yaitu random forest dan K-nearst neighbors. Hasil dari pengujian testing menunjukkan bahwa kedua algoritma memberikan kinerja klasifikasi yang baik, dengan K-nearst neighbors sedikit lebih unggul dibandingkan Random Forest pada data uji. Dengan demikian, algoritma KNN dapat dipertimbangkan sebagai pilihan utama untuk implementasi awal sistem, sementara Random Forest dapat menjadi model alternatif yang lebih stabil ketika ukuran data diperbesar.

ABSTRACT

Imami, Nia Kurniawati, 2025, **Career Path Mapping System for Vocational High School Graduates Using the Random Forest and K-Nearest Neighbors Methods.** Master of Informatics Program, Maulana Malik Ibrahim State Islamic University, Supervisors: (1) Dr. M. Amin Hariyadi, M.T. (2) Dr. Ir. Yunifa Miftachul Arif, S.ST., M.T.

Keywords: Mapping System, Career Path for Vocational High School Graduates, Random Forest, K-Nearest Neighbors.

Vocational High Schools (VHSs) are schools that prepare students for employment after graduation. Therefore, during the learning process, they are not only taught theory but also practice. There is even a Field Work Practice (PKL) or internship program where students gain real-world work experience. With such training, it is hoped that vocational high school graduates will be able to compete in the workforce according to their field of expertise. However, according to Statistics Indonesia data from 2022, unemployment reached 8.42%, the highest among graduates of other schools, and 60% of vocational high school graduates worked outside their field of expertise. Therefore, an algorithm is needed that can help predict the career path of vocational high school graduates. This study uses the random forest and K-nearst neighbors algorithms. The data used in this study comes from the Special Job Exchange (BKK) which contains 8 core competencies of vocational high school students. In the hyperparameter tuning test, a combination of 60% to 90% of training data and 40% to 10% of testing data was carried out. From this test, the best results were obtained with a combination of 80% of training data and 20% of testing data. Then, this combination was applied to the testing data for each algorithm, namely random forest and K-nearst neighbors. The results of the testing show that both algorithms provide good classification performance, with K-nearst neighbors slightly superior to Random Forest on the test data. Thus, the KNN algorithm can be considered as the main choice for initial system implementation, while Random Forest can be a more stable alternative model when the data size is enlarged.

مستخلص البحث

إمامي، نيا كورنياواتي، 2025، نظام رسم خرائط مسار الوظائف لخريجي المدارس الثانوية المهنية باستخدام خوارزمية الغابة العشوائية وأقرب الجيران. رسالة الماجستير. قسم المعلومات، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج، المشرف الأول: د. محمد أمين هاريادي، الماجستير؛ المشرف الثاني: د. يونيفا مفتاح العارف، الماجستير.

الكلمات الرئيسية: نظام رسم خرائط، مسار وظائف لخريجي مدارس ثانوية مهنية، غابة عشوائية، حوارزمية أقرب جيران.

المدرسة الثانوية المهنية هي مدرسة معدة للطلاب المستعدين للعمل بعد التخرج. لذلك، خلال عملية التعلم، لا يتم تقديم النظريات فقط، بل أيضًا التدريب العملي، وهناك حتى برنامج التدريب العملي الميداني حيث يحصل الطلاب على خبرة عملية حقيقية في الميدان. مع هذا الإعداد، يُتوقع أن يتمكن خريجو المدارس الثانوية المهنية من المنافسة في سوق العمل وفقًا لمجال تخصصهم. ومع ذلك، وفقًا لبيانات إحصاءات إندونيسيا لعام 2022، بلغت نسبة البطالة 8.42٪، وهي الأعلى مقارنة بخريجي المدارس الأخرى، وقد تم تسجيل أن 60٪ من خريجي المدارس الثانوية المهنية يعملون خارج مجال تخصصهم. لذلك، هناك حاجة إلى خوارزمية يمكن أن تساعد في التنبؤ بمسار مهنة خريجي المدارس الثانوية المهنية، في هذا البحث يتم استخدام خوارزمية الغابة العشوائية وأقرب الجيران . البيانات المستخدمة في هذه الرسالة مستمدة من بورصة العمل الخاصة التي تضم 8 كفاءات أساسية لطلاب المدارس الثانوية المهنية .في اختبار ضبط المعلمات الفائقة تم استخدام مزيج من بيانات التدريب بنسبة تتراوح بين 60% إلى 90% وبيانات الاختبار بين 40% إلى 10%. من هذا الاختبار تم الحصول على أفضل النتائج عند مزيج بيانات التدريب بنسبة 80% وبيانات الاختبار بنسبة 20%. ثم تم تطبيق هذا المزيج على بيانات الاختبار لكل من خوارزمية الغابة العشوائية (*Random Forest*)وأقرب الجيران (*K-nearest neighbors*). أظهرت نتائج الاختبار أن كلا الخوارزميتين تقدمان أداء تصنيفي جيدًا، حيث كانت خوارزمية *K-nearest neighbors* متفوقة قليلاً مقارنة بالغابة العشوائية على بيانات الاختبار. وبالتالي، يمكن اعتبار خوارزمية KNN الخيار الرئيسي لتطبيق النظام في البداية، بينما يمكن أن تكون الغابة العشوائية نموذجًا بديلًا أكثر استقرارًا عند زيادة حجم البيانات.

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Setiap lulusan Sekolah Menengah Kejuruan (SMK) dipersiapkan untuk menjadi tenaga kerja yang handal karena pada hakikatnya pada saat mereka lulus menyelesaikan pendidikannya, lulusan SMK diharapkan mampu mengimplementasikan ilmu yang telah didapatkan selama mengikuti pembelajaran di sekolah. Lulusan SMK tidak sekadar berkualitas dan kompeten, tapi kompetensi yang dimiliki harus sesuai dengan yang dibutuhkan oleh perusahaan. Sehingga banyak tenaga kerja lulusan SMK yang terserap di sektor industri. Delapan kompetensi pokok yang harus dimiliki lulusan SMK, yakni (1) communication skills; (2) *critical and creative thinking;* (3) *inquiry/reasoning skills;* (4) *interpersonal skills;* (5) *multicultural/multilingual literacy;* (6) *problem solving;* (7) *information/digital literacy;* dan (8) *technological skills* (Sutianah, 2020). Berdasarkan delapan kompetensi lulusan tersebut, dapat dikategorikan poin 1-6 adalah aspek *soft skill* dan 7-8 merupakan aspek *hard skills*.

Pemerintah berupaya dalam pemenuhan ketrampilan siswa SMK dengan menggandeng industri dalam hal penggunaan kurikulum, pelaksanaan pembelajaran dan praktek kerja (Peraturan Pemerintah RI, 2020). SMK dirancang sebagai institusi pendidikan yang menyiapkan peserta didik untuk langsung masuk ke dunia kerja sesuai dengan jurusan keahliannya. Namun, realitas di lapangan menunjukkan bahwa banyak lulusan SMK justru bekerja di sektor informal atau bidang yang tidak sesuai dengan kompetensi yang telah mereka

pelajari. Berdasarkan data (Anton Wirawan, 2023)**,** sekitar 60% lulusan SMK di Indonesia bekerja di luar bidang keahliannya, dan hanya sebagian kecil yang berhasil terserap di industri yang sesuai dengan jurusan saat sekolah. Hal ini menjadi salah satu permasalahan yang kerap terjadi di dunia pendidikan kejuruan di Indonesia dan turut menyumbang pada angka pengangguran terbuka yang tinggi di kalangan lulusan SMK, yang menurut (Statistik & Indonesia, 2022) mencapai 8.42%**,** tertinggi dibandingkan jenjang pendidikan lainnya.

Bursa Kerja Khusus yang selanjutnya disingkat BKK adalah unit pelayanan pada satuan pendidikan menengah, satuan pendidikan tinggi, dan lembaga pelatihan kerja, yang memberikan fasilitas penempatan tenaga kerja kepada alumninya (Menteri Tenaga Kerja, 2016). Berdasarkan pengertian di atas maka BKK merupakan lembaga yang dibentuk di SMK atau perguruan tinggi atau lembaga pelatihan yang merupakan mitra dari Dinas ketenagakerjaan di kabupaten masing masing. Tujuan dari dibentuknya BKK di SMK adalah untuk memberikan pelayanan dan informasi lowongan kerja, melaksanakan pemasaran, penyaluran, dan penempatan tenaga kerja, serta sebagai wadah yang menanamkan jiwa kewirausahaan bagi alumni melalui kegiatan pelatihan (RI, 2024). Maka BKK berfungsi untuk mempertemukan alumni dari SMK dengan industri sebagai pencari kerja. Mekanisme dapat disesuaikan dengan kesepakatan bersama antara SMK dengan DU/DI. Kemudian BKK meliputi seksi administrasi, informasi penempatan kerja (IPK), dan bagian interview awal sebelum disalurkan kepada pengguna. BKK juga menyediakan layanan latihan untuk mempersiapkan alumni masuk ke dunia kerja seperti pelatihan menghadapi *interview*, trik dan tips lolos seleksi serta latihan kepribadian. Maka dari itu BKK merupakan salah satu

komponen penting dalam mengukur keberhasilan pendidikan di SMK, Semakin banyak kerjasama antara SMK dan DU/DI yang dijembatani oleh BKK maka akan semakin mudah SMK dalam menyalurkan alumninya ke dunia kerja karena BKK menjadi lembaga yang berperan mengoptimalkan penyaluran tamatan SMK dan sumber informasi untuk pencari kerja.

Dalam Islam, bekerja dan mencari rezeki yang halal adalah suatu kewajiban. Hal ini ditegaskan dalam **QS. Al-Jumu'ah (62:10):**

Artinya :

"Apabila telah ditunaikan shalat, maka bertebaranlah kamu di muka bumi dan carilah karunia Allah, dan ingatlah Allah banyak-banyak agar kamu beruntung."

Menurut Tafsir Ibnu Katsir (2011), ayat ini menjelaskan bahwa setelah menunaikan kewajiban ibadah, umat Islam *diperintahkan untuk berusaha, bekerja, dan mencari rezeki yang halal* di muka bumi. Perintah "*bertebaranlah kamu di muka bumi"* tidak hanya bermakna fisik, tetapi juga mengandung makna *aktif, produktif, dan berikhtiar* dalam meningkatkan kesejahteraan hidup. Ibnu Katsir menegaskan bahwa Islam tidak melarang bekerja dan berdagang setelah ibadah, bahkan mendorong umatnya untuk memanfaatkan waktu dan potensi yang dimiliki secara optimal.

Korelasi ayat ini terletak pada konsep *ikhtiar terencana dalam mencari karunia Allah*. Pemetaan jalur karir merupakan bentuk ikhtiar modern dan sistematis untuk membantu lulusan SMK menemukan pekerjaan yang sesuai

dengan kompetensi, minat, dan kemampuan yang dimiliki. Hal ini sejalan dengan perintah Al-Qur'an agar manusia tidak pasif setelah menunaikan ibadah, tetapi berusaha secara sungguh-sungguh untuk memperoleh penghidupan yang layak.

Lebih lanjut, Tafsir Ibnu Katsir (2011) menekankan bahwa mencari karunia Allah harus tetap diiringi dengan *kesadaran spiritual dan etika*, sebagaimana perintah *"dan ingatlah Allah banyak-banyak"*. Dalam konteks penelitian ini, penggunaan metode klasifikasi seperti Random Forest dan K-Nearest Neighbors tidak hanya bertujuan untuk meningkatkan akurasi penempatan kerja lulusan SMK, tetapi juga untuk menghadirkan sistem yang adil, objektif, dan bertanggung jawab. Dengan demikian, pemanfaatan teknologi informasi dalam pemetaan karir menjadi sarana yang mendukung nilai-nilai Islam dalam bekerja secara profesional dan bermartabat.

Dengan demikian, QS. Al-Jumu'ah ayat 10 menurut Tafsir Ibnu Katsir (2011) memberikan landasan teologis bahwa *usaha mencari pekerjaan dan pengembangan karir merupakan bagian dari perintah agama*. Penelitian ini menjadi wujud implementasi nilai tersebut dalam konteks pendidikan kejuruan dan dunia kerja modern, khususnya BKK dalam membantu lulusan SMK agar mampu berikhtiar secara tepat, terarah, dan produktif untuk mencapai keberhasilan hidup di dunia dan akhirat.

Dengan memanfaatkan data siswa di BKK maka dapat memberikan rekomendasi yang lebih akurat dan relevan. Pemilihan metode klasifikasi dilakukan dengan membandingkan beberapa algoritma didalam metode klasifikasi dengan mengacu pada performa akurasi, *presisi, recall* dan *f1-score* yang memiliki nilai tertinggi.

Selain memberikan manfaat bagi BKK, sistem pemetaan jalur karir berbasis klasifikasi ini juga bermanfaat bagi berbagai pihak. Bagi siswa, sistem ini dapat membantu mereka dalam mengambil keputusan karir yang lebih baik dan lebih sesuai dengan kemampuan serta minat mereka. Bagi sekolah, sistem ini dapat menjadi alat bantu dalam memberikan bimbingan karir yang lebih efektif dan efisien. Sementara bagi industri, sistem ini dapat membantu dalam memperoleh tenaga kerja yang lebih sesuai dengan kebutuhan, sehingga dapat meningkatkan produktivitas dan efisiensi kerja (RI, 2024). Metode klasifikasi merupakan teknik untuk mengelompokkan data berdasarkan karakteristik tertentu ke dalam kelas atau kategori yang sudah ditentukan sebelumnya (Misaria Tarigan, Putu, and Lestari, 2023). Terdapat beberapa jenis metode klasifikasi yang digunakan dalam dalam pemetaan jalur karir lulusan siswa SMK misalnya pada penelitian yang dilakukan oleh Gokarn et al., (2024) yang menggunakan empat jenis metode klasifikasi yaitu *K-Nearest Neighborn, Support Vector Machine*, *Decision Tree, dan Random Forest.* Hasil dari penelitian tersebut metode *K-NN* menghasilkan nilai akurasi tertinggi. Selain itu (Pandey & L S, 2022) meneliti tentang memprediksi jalur karir siswa menggunakan metode *Naïve Bayes, Decision Tree,* dan *K-Nearest Neighbors (K-NN).* Hasil evaluasi menunjukkan bahwa *K-NN* memberikan akurasi tertinggi sebesar 91%. Penelitian lain yang dilakukan oleh Sinha et al., (2023) menggunakan metode *yaitu Decision Tree, Naïve Bayes, Support Vector Machine (SVM), dan Neural Network.* Hasil evaluasi menunjukkan bahwa metode *Decision Tree* memiliki akurasi tertinggi sebesar 96.6%.

Penelitian lain yang dilakukan oleh Selain itu penelitian yang dilakukan oleh

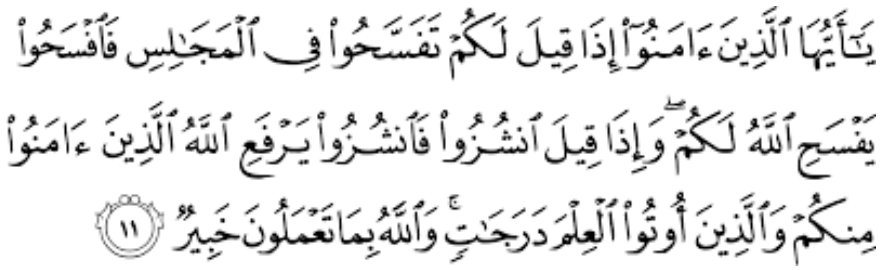
Betrand et al., (2025) membandingkan hasil klasifikasi dari beberapa metode tentang lulusan karir siswa SMK. Metode yang dibandingkan yaitu *Decision Tree, Random Forest, dan Naïve Bayes*. Dari penelitian tersebut didapatkan hasil evaluasi bahwa algoritma Random Forest memberikan akurasi tertinggi sebesar 93%. Penelitian yang dilakukan oleh Agustiningsih et al., (2023) tentang klasifikasi lulusan siswa SMK di dunia industri. Pada penelitian ini membandingkan 3 metode yaitu *extreme gradient boosting (xgboost), random forest*, dan *logistic regression*. Hasil yang didapat pada penelitian ini adalah skor *train* 97.36%, skor *test* 68.71% dan skor akurasi 67% oleh *Random Forest*. Penelitan lain yang dilakukan oleh Mahmud Nawawi et al., (2024) tentang prediksi penempatan karir dengan metode klasifikasi machine learning yaitu *Random Forest, Decision Tree, Naïve Bayes, KNN, dan SVM*. Hasil yang didapat pada penelitian ini adalah akurasi tertinggi diperoleh oleh metode *Random Forest* dengan skor akurasi sebesar 87%.

Selain metode di atas, metode *neural network* juga pernah digunakan pada pengelompokan karir lulusan pasca sarjana oleh Haque et al., (2025) .Studi ini melibatkan beberapa model ML termasuk Jaringan Syaraf Tiruan (JST), *CatBoost*, dan pengklasifikasi BERT. Pada penelitian ini menghasilkan akurasi optimal 88% diperoleh dengan menerapkan seleksi fitur sebelum dan sesudah penyisipan, dengan model *BERT-Boruta*. Penelitian lain menggunakan metode *naive bayes* dan *decision tree* dalam mengklasifikasi kinerja staf universitas (Farhana, 2021). Dari penelitian ini didapatkan hasil 96.15% untuk *naive bayes* dan 94.23% untuk *decision tree.*

Dengan latar belakang tersebut, penelitian ini bertujuan untuk

mengembangkan dan mengimplementasikan sistem pemetaan jalur karir bagi lulusan SMK dengan menggunakan metode klasifikasi, yang diintegrasikan dengan fungsi BKK. Penelitian ini akan mengeksplorasi berbagai algoritma klasifikasi untuk menentukan algoritma yang paling efektif dalam memetakan jalur karir yang sesuai. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam bidang bimbingan karir di SMK, serta dapat diimplementasikan secara luas untuk meningkatkan kesesuaian antara pendidikan dan pekerjaan bagi lulusan SMK.

Hal ini sejalan dengan firman Allah dalam QS. Al-Mujadilah (58:11):



Artinya :

"Allah akan meninggikan derajat orang-orang yang beriman di antaramu dan orang-orang yang diberi ilmu beberapa derajat."

Dalam hadis yang diriwayatkan oleh HR Muslim No. 2699 juga disebutkan keutamaan tingginya kedudukan orang berilmu.

قَالَ رَسُولُ اللَّهِ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ:

"مَنْ سَلَكَ طَرِيقًا يَلْتَمِسُ فِيهِ عِلْمًا سَهَّلَ اللهُ لَهُ بِهِ طَرِيقًا إِلَى الْجَنَّةِ".

Artinya :

"Barang siapa menempuh jalan untuk mencari ilmu, maka Allah akan mudahkan baginya jalan menuju surga."

Ayat dan hadist ini menunjukkan bahwa ilmu memiliki peran besar dalam

meningkatkan taraf hidup seseorang, yang sesuai dengan tujuan sistem pemetaan jalur karir ini.

QS. Al-Mujādilah ayat 11 menjelaskan bahwa Allah SWT meninggikan derajat orang-orang yang beriman dan orang-orang yang berilmu. Berdasarkan Tafsir Ibnu Katsir (2011), ayat ini menegaskan bahwa ilmu pengetahuan memiliki kedudukan yang tinggi karena menjadi sarana bagi manusia untuk memperoleh kemuliaan, kehormatan, dan keutamaan dalam kehidupan. Orang yang memiliki ilmu dan mengamalkannya dengan benar akan memperoleh peningkatan derajat, baik secara spiritual maupun sosial, sebagai bentuk penghargaan dari Allah SWT atas usaha dan pemanfaatan ilmu tersebut.

Korelasi keterkaitan ayat ini terletak pada pemanfaatan ilmu pengetahuan dan teknologi sebagai sarana peningkatan kualitas sumber daya manusia. Penerapan metode klasifikasi dalam sistem pemetaan jalur karir bertujuan untuk membantu lulusan SMK memperoleh rekomendasi karir yang sesuai dengan kompetensi dan potensi yang dimiliki. Dengan demikian, penelitian ini merupakan implementasi nilai QS. Al-Mujādilah ayat 11, yaitu penggunaan ilmu secara aplikatif dan bertanggung jawab untuk meningkatkan derajat dan masa depan lulusan SMK secara lebih terarah dan objektif.

1.2 Pernyataan Masalah

Seiring perkembangan teknologi dan informasi serta ditemukannya lulusan siswa SMK yang bekerja tidak sesuai dengan kompetensi bidang keahlian yang dipelajari saat di sekolah maka pada penelitian ini dilakukan pemetaan jalur karir

lulusan Siswa SMK. Adapun pernyataan masalah pada penelitian ini adalah sebagai berikut:

a. Bagaimana memprediksi karir dengan menggunakan algoritma *Random Forest* dan *K-Nearest Neighbors* ?
b. Bagaimana mengukur *confusion matrix* hasil prediksi karir dengan menggunakan algoritma *Random Forest* dan *K-Nearest Neighbors* ?

1.3 Tujuan Penelitian

Tujuan penelitian pada penelitian ini adalah sebagai berikut :

a. Untuk memprediksi karir dengan menggunakan algoritma *Random Forest* dan *K-Nearest Neighbors*.
b. Untuk mengukur *confusion matrix* hasil prediksi karir dengan menggunakan algoritma *Random Forest* dan *K-Nearest Neighbors*

1.4 Manfaat Panelitian

Adapun manfaat dari penelitian ini adalah:

a. Membantu siswa dalam menentukan jalur karir yang sesuai
b. Sebagai alat bantu dalam bimbingan karir di SMK
c. Industri dapat memperoleh tenaga kerja yang lebih sesuai dengan kebutuhan dan kompetensi yang diharapkan.
d. Sebagai referensi bagi peneliti lain untuk mengembangkan sistem ini.

1.5 Batasan Masalah

a. Data yang digunakan pada penelitian ini berasal dari dari lulusan SMKN 1 Wonorejo bagian BKK.

b. Penelitian ini berfokus pada pembuatan model menggunakan metode *Random Forest* dan *K-Nearst Neighbors*.
c. Penelitian ini hanya mencari metode terbak berdasarkan perbandingan dua metode tersebut.

BAB II

STUDY PUSTAKA

2.1 Pemetaan Jalur Karir Lulusan SMK

Ada beberapa penelitian terdahulu yang terkait dengan Sistem pemetaan jalur karir diantaranya penelitian (Betrand et al., 2025) membahas pembangunan sistem bimbingan karir berbasis machine learning dengan membandingkan tiga algoritma klasifikasi, yaitu *Decision Tree, Random Forest,* dan *Naïve Bayes*. Sistem dirancang untuk memberikan rekomendasi karir yang sesuai bagi siswa berdasarkan atribut seperti nilai akademik, preferensi minat karir, kepribadian, dan aktivitas ekstrakurikuler. Ketiga algoritma diuji untuk menilai efektivitas klasifikasi pilihan karir terhadap profil siswa. Hasil evaluasi menunjukkan bahwa algoritma *Random Forest* memberikan akurasi tertinggi sebesar 93%, diikuti oleh *Decision Tree* dengan akurasi 90%, sedangkan *Naïve Bayes* mencatat akurasi sebesar 87%. Studi ini menyimpulkan bahwa *Random Forest* adalah algoritma yang paling optimal dalam konteks sistem bimbingan karir berbasis data, karena mampu menangani variabel kompleks dengan hasil klasifikasi yang lebih akurat dan stabil.

Penelitian oleh Gokarn et al., (2024) membahas pengembangan sistem bimbingan karir cerdas berbasis machine learning untuk membantu siswa memilih jalur karir yang sesuai dengan profil akademik dan kepribadian mereka. Sistem ini dirancang untuk memberikan rekomendasi karir berdasarkan analisis data siswa menggunakan algoritma klasifikasi. Atribut yang digunakan dalam penelitian ini

meliputi nilai akademik, hasil kuisioner minat dan bakat, serta hasil kuis bidang tertentu. Output pada penelitian ini adalah untuk menghasilkan rekomendasi jalur karir dengan 3 pilihan teratas. Pada pemrosesan data dilakukan beberapa perlakuan antara lain : penggabungan dua dataset (nilai dan minat), normalisasi nilai numerik menjadi kategorikal serta pembagian data 70:30. Penelitian ini membandingkan performa empat algoritma *yaitu K-Nearest Neighbor (K- NN)*, *Support Vector Machine (SVM)*, *Decision Tree*, dan *Random Forest*. Hasil eksperimen menunjukkan bahwa *K-NN* memiliki akurasi tertinggi sebesar 94%, *Random Forest* dengan akurasi 81%, sedangkan *SVM* 86% dan *Decision Tree* 89%. Studi ini menyimpulkan bahwa pendekatan *machine learning*, khususnya algoritma *K-NN*, sangat potensial untuk diimplementasikan dalam sistem rekomendasi karir yang adaptif dan personal di lingkungan pendidikan menengah.

Penelitian berjudul "*Application of Data Mining Techniques in Assessing the Performance of Vocational High School Students*" yang dilakukan oleh Adi Putra et al., (2024) membahas penerapan algoritma klasifikasi untuk menilai kinerja siswa SMK berdasarkan data historis pendidikan. Penelitian ini menggunakan beberapa atribut penting, yaitu nilai akademik siswa, tingkat kehadiran, dan hasil observasi guru terhadap perilaku serta keterampilan siswa. Tiga algoritma klasifikasi yang diuji dalam penelitian ini adalah *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *K-Nearest Neighbor (K-NN)*. Berdasarkan hasil evaluasi, algoritma SVM menunjukkan performa terbaik dengan akurasi sebesar 93.2%, *precision* 93.4%, *recall* 93.2%, dan *F1-score* 93.1%, mengungguli algoritma lainnya. Temuan ini menunjukkan bahwa metode klasifikasi berbasis *SVM* sangat efektif untuk memprediksi performa siswa di lingkungan pendidikan vokasi dan

dapat digunakan sebagai dasar dalam pengambilan keputusan akademik serta pembinaan karir siswa (Adi Putra et al., 2024) .

Pembahasan penerapan berbagai algoritma *machine learning* untuk membangun sistem prediksi karir yang membantu siswa memilih jalur karir yang sesuai berdasarkan data akademik dan kepribadian dilakukan oleh Sinha et al., (2023). Atribut yang digunakan dalam penelitian ini mencakup nilai akademik, minat siswa, bakat, kepribadian, dan keterampilan non- akademik. Beberapa algoritma yang dibandingkan meliputi *Decision Tree, Naïve Bayes, Support Vector Machine (SVM),* dan *Neural Network*. Hasil evaluasi menunjukkan bahwa algoritma *Decision Tree* memiliki akurasi tertinggi sebesar 96.6%, disusul oleh *Neural Network* dengan akurasi 94.3%, sementara *SVM* dan *Naïve Bayes* masing-masing mencatatkan akurasi sebesar 92.1% dan 90.2%. Penelitian ini menyimpulkan bahwa metode klasifikasi berbasis *machine learning* efektif untuk memberikan rekomendasi karir berbasis data, dengan *Decision Tree* menjadi pilihan paling akurat dalam konteks dataset yang digunakan.

Penelitian yang dilakukan oleh Idakwo et al., (2022) mengembangkan sistem rekomendasi jalur karir bagi mahasiswa jurusan Ilmu Komputer dengan pendekatan hibrida berbasis teknik *ensemble machine learning*. Sistem ini dirancang untuk membantu mahasiswa memilih jalur karir yang sesuai dengan minat, kemampuan, dan rekam jejak akademik mereka. Atribut yang digunakan mencakup nilai akademik, minat pribadi, keterampilan interpersonal, dan bakat teknis. Empat algoritma pembelajaran mesin yang digunakan sebagai model dasar adalah *Decision Tree (C4.5), Naïve Bayes, K-Nearest Neighbor (K-NN),* dan *Support Vector Machine (SVM)*. Hasil evaluasi menunjukkan bahwa *Decision*

Tree memiliki akurasi tertinggi di antara model dasar dengan nilai 82.01%. Namun, dengan menerapkan teknik *ensemble Bagging*, akurasi meningkat signifikan menjadi 90.65%, dengan *precision, recall,* dan *F1-score* masing-masing sebesar 90.7%, 90.6%, dan 90.6%. Temuan ini menegaskan bahwa pendekatan hibrida menggunakan *ensemble learning* dapat secara efektif meningkatkan akurasi sistem rekomendasi karir, sehingga memberikan panduan yang lebih tepat bagi mahasiswa dalam merencanakan jalur karir mereka di bidang Teknologi Informasi.

Penelitian (Pandey & L S, 2022) membahas pemanfaatan algoritma klasifikasi untuk memprediksi jalur karir siswa berdasarkan prestasi akademik dan keterampilan pribadi. Tujuan utamanya adalah mengembangkan sistem prediktif yang membantu siswa menentukan pilihan karir yang sesuai. Atribut yang digunakan meliputi nilai akademik, keterampilan teknis, soft skills, dan keterlibatan dalam kegiatan ekstrakurikuler. Tiga algoritma yang dibandingkan dalam penelitian ini adalah *Naïve Bayes, Decision Tree,* dan *K-Nearest Neighbors (K-NN).* Hasil evaluasi menunjukkan bahwa KNN memberikan akurasi tertinggi sebesar 91%, diikuti oleh *Decision Tree* dengan akurasi 88%, dan *Naïve Bayes* sebesar 85%. Studi ini menyimpulkan bahwa *K-NN* merupakan algoritma yang paling efektif untuk klasifikasi karir berbasis data siswa, dan pendekatan ini sangat potensial untuk diterapkan dalam sistem bimbingan karir di sekolah menengah. Jurnal lain yang dilakukan oleh Purnomo & Sururi, (2022) bertujuan untuk memprediksi kemampuan siswa dalam menghadapi dunia kerja dengan membandingkan kinerja dua algoritma klasifikasi, yaitu *Naïve Baye*s dan *K-Nearest Neighbor (K-NN).* Studi ini menggunakan data siswa dari SMK yang

mencakup atribut-atribut seperti nilai akademik, tingkat kehadiran, jenis kelamin, jurusan, dan hasil evaluasi praktik kerja industri (PKL). Kedua algoritma diuji untuk mengklasifikasikan apakah seorang siswa memiliki potensi besar untuk bersaing di dunia kerja. Hasil evaluasi menunjukkan bahwa *K-NN* memberikan akurasi tertinggi sebesar 98.22%, dengan *precision* 99.38% dan *recall* 98.77%, sedangkan *Naïve Bayes* mencatat akurasi 97.66%, dengan *precision* sempurna sebesar 100% dan *recall* 97.59%. Kesimpulannya, kedua algoritma memiliki performa tinggi, namun *K-NN* sedikit lebih unggul dalam akurasi keseluruhan, menjadikannya lebih direkomendasikan untuk prediksi kesiapan kerja siswa SMK.

Penelitian yang dilakukan oleh Vignesh et al., (2021) mengembangkan sistem bimbingan karir cerdas berbasis web yang memanfaatkan algoritma machine learning untuk membantu siswa memilih jalur karir yang sesuai dengan minat dan kemampuan mereka. Sistem ini menggunakan atribut seperti nilai akademik, minat karir, hasil tes kepribadian, dan keterampilan teknis sebagai input untuk merekomendasikan bidang studi atau karir yang paling cocok bagi pengguna. Beberapa algoritma klasifikasi yang diterapkan dalam sistem ini meliputi *K-Nearest Neighbor (K-NN), Support Vector Machine (SVM),* dan *Naïve Bayes.* Hasil evaluasi menunjukkan bahwa algoritma *K-NN* mencapai akurasi tertinggi, melebihi 90%, dalam mengklasifikasikan jalur karir yang sesuai. Hal ini menunjukkan bahwa pendekatan berbasis machine learning dapat meningkatkan efektivitas sistem bimbingan karir dengan memberikan rekomendasi yang lebih personal dan *data-driven*.

Penelitian oleh Al-Dossari et al., (2020) (mengusulkan sistem rekomendasi karir bernama *CareerRec* yang dirancang khusus untuk membantu lulusan bidang Teknologi Informasi (TI) dalam menentukan jalur karir yang paling sesuai. Sistem ini menggunakan pendekatan machine learning untuk mengidentifikasi kecocokan antara profil lulusan dan jenis pekerjaan. Atribut yang digunakan mencakup IPK, jenis kelamin, universitas asal, kemampuan teknis *(programming*, *networking*, *database*), dan hasil pelatihan industri. Beberapa algoritma klasifikasi yang dibandingkan dalam penelitian ini meliputi *Naïve Bayes, Random Forest,* dan *Support Vector Machine (SVM)*. Hasil evaluasi menunjukkan bahwa *Random Forest* memberikan akurasi terbaik sebesar 96%, disusul oleh *SVM* dengan 94%, dan *Naïve Bayes* sebesar 91%. Studi ini menyimpulkan bahwa *CareerRec* berbasis *Random Forest* efektif untuk membantu lulusan TI memilih karir yang relevan dengan kompetensi mereka, dan dapat diintegrasikan sebagai sistem pendukung keputusan dalam layanan bimbingan karir pendidikan tinggi.

Penelitian yang dilakukan oleh Wibisono et al., (2024) membahas klasifikasi jenis pekerjaan alumni berdasarkan nilai mata kuliah selama masa studi menggunakan dua algoritma machine learning, yaitu *Naïve Bayes* dan *K-Nearest Neighbor (K-NN)*. Tujuan utama dari studi ini adalah membangun sistem prediktif untuk mengelompokkan lulusan ke dalam bidang pekerjaan yang sesuai dengan latar belakang akademiknya. Atribut yang digunakan dalam klasifikasi adalah nilai-nilai dari sejumlah mata kuliah inti di program studi Pendidikan Teknologi Informasi. Hasil evaluasi menunjukkan bahwa kedua algoritma menghasilkan tingkat akurasi yang sama, yaitu sebesar 66.66%, namun dengan karakteristik yang berbeda dalam pemrosesan dan kompleksitas. Penelitian ini menyimpulkan

bahwa meskipun akurasi belum optimal, klasifikasi berbasis nilai akademik memiliki potensi untuk dikembangkan lebih lanjut sebagai alat bantu dalam pemetaan karir lulusan.

Selain metode di atas, metode neural network juga pernah digunakan pada pengelompokan karir lulusan pasca sarjana oleh Haque et al., (2025) .Studi ini melibatkan beberapa model *ML* termasuk *Jaringan Syaraf Tiruan (JST), CatBoost,* dan pengklasifikasi *BERT*. Model dasar (tanpa seleksi fitur dan penyisipan) mencapai akurasi tertinggi dengan model *ANN* (79%). Selanjutnya, penerapan *ETC* untuk seleksi fitur meningkatkan akurasi, dengan *CatBoost* mencapai 83%. Transformasi lebih lanjut dengan penyisipan berbasis *BERT* meningkatkan akurasi menjadi 85% menggunakan pengklasifikasi *BERT*. Lalu akurasi optimal 88% diperoleh dengan menerapkan seleksi fitur sebelum dan sesudah penyisipan, dengan model *BERT-Boruta*. Temuan dari studi ini menunjukkan bahwa penggunaan pendekatan seleksi fitur dua tahap yang dikombinasikan dengan penyisipan *BERT* secara signifikan meningkatkan akurasi klasifikasi. . Penelitian lain menggunakan metode *naive bayes* dan *decision tree* dalam mengklasifikasi kinerja staf universitas (Farhana, 2021). Dalam klasifikasi, kinerja peneliti dan staf akademik diamati oleh penanggung jawab yang terkait dengan penelitian dan mencatat data. Setelah proses ini, model dihasilkan menggunakan set data pelatihan, kemudian model diuji dengan set data pengujian tanpa kelas atribut. Hasil dari penelitian ini, klasifikasi *Naïve Bayes* dapat mengklasifikasikan kinerja akademik dengan aktivitas penelitian lebih baik daripada *decision tree* dan N*aïve Bay*es, masing-masing sebesar 96.15% dan 94.23%.

2.2 Dasar Teori

2.2.1 Random Forest

Random Forest merupakan salah satu teknik dalam machine learning yang bekerja dengan menggabungkan hasil dari banyak pohon keputusan untuk meningkatkan tingkat akurasi prediksi (Betrand et al., 2025) .Metode ini merupakan pengembangan dari pendekatan *Classification and Regression Tree (CART)* dan memiliki sejumlah kelebihan dibandingkan pendekatan tersebut. Karena membangun banyak pohon berdasarkan kombinasi data dan atribut yang dipilih secara acak, algoritma ini sering disebut sebagai "hutan" pohon keputusan. Setiap pohon dalam *Random Forest* digunakan untuk melakukan evaluasi terhadap data, dan keputusan akhir diambil berdasarkan mayoritas prediksi yang dihasilkan oleh seluruh pohon. Proses pembentukan pohon dimulai dengan pemilihan atribut akar yang ditentukan berdasarkan nilai gain yang diperoleh dari perhitungan entropi, dan proses ini diulang hingga semua data tergolong dalam kelas yang sama.

Implementasi metode *random forest* dapat digunakan untuk beberapa kasus seperti yang dilakukan oleh Betrand et al., (2025) membahas pembangunan sistem bimbingan karir berbasis machine learning dengan membandingkan tiga algoritma klasifikasi, yaitu *Decision Tree, Random Forest, dan Naïve Bayes*. Dari penelitian tersebut didapatkan hasil evaluasi bahwa algoritma *Random Forest* memberikan akurasi tertinggi sebesar 93%, diikuti oleh *Decision Tree* dengan akurasi 90%, sedangkan *Naïve Bayes* mencatat akurasi sebesar 87%. Penelitian yang dilakukan oleh Agustiningsih et al., (2023) tentang klasifikasi lulusan siswa SMK di dunia industri. Pada penelitian ini membandingkan 3 metode yaitu *extreme gradient*

boosting (xgboost), *random forest*, dan *logistic regression*. Hasil dari ketiga metode tersebut mendapatkan skor *train* 91.70%, skor *test* 66.88% dan skor akurasi 67% yang dihasilkan oleh algoritma *XGBoost*, skor *train* 97.36%, skor *test* 68.71% dan skor akurasi 67% oleh *Random Forest*, dan skor *train* 51.14%, skor *test* 50.43% dan skor akurasi 50% oleh *Logistic Regression*. Penelitian oleh Al-Dossari et al., (2020) (mengusulkan sistem rekomendasi karir bernama *CareerRec* yang dirancang khusus untuk membantu lulusan bidang Teknologi Informasi (TI) dalam menentukan jalur karir yang paling sesuai. Sistem ini menggunakan pendekatan machine learning untuk mengidentifikasi kecocokan antara profil lulusan dan jenis pekerjaan. Atribut yang digunakan mencakup IPK, jenis kelamin, universitas asal, kemampuan teknis (*programming*, *networking*, *database*), dan hasil pelatihan industri. Beberapa algoritma klasifikasi yang dibandingkan dalam penelitian ini meliputi *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine (SVM)*. Hasil evaluasi menunjukkan bahwa *Random Forest* memberikan akurasi terbaik sebesar 96%, disusul oleh *SVM* dengan 94%, dan *Naïve Bayes* sebesar 91%.

2.2.2 K-Nearest Neighbors (K-NN)

K-Nearest Neighbors (K-NN) adalah salah satu algoritma klasifikasi yang termasuk dalam kelompok *lazy learner* atau *instance-based learning*, di mana proses pembelajaran tidak dilakukan secara eksplisit saat pelatihan, tetapi dilakukan saat proses prediksi. *K-NN* bekerja berdasarkan prinsip bahwa objek yang serupa cenderung berada dalam kelompok yang sama. Oleh karena itu, untuk memprediksi kelas dari suatu data baru, *K-NN* akan mencari K tetangga terdekat

dari data tersebut, kemudian menentukan kelas berdasarkan mayoritas kelas dari tetangga tersebut (rachmadani et al., 2020).

Beberapa penelitian yang menggunakan metode *KNN* dalam menentukan jalur karir seperti yang dilakukan oleh (Purnomo & Sururi, 2022) bertujuan untuk memprediksi kemampuan siswa dalam menghadapi dunia kerja dengan membandingkan kinerja dua algoritma klasifikasi, yaitu *Naïve Bayes* dan *K-Nearest Neighbors (K-NN).* Studi ini menggunakan data siswa dari SMK yang mencakup atribut-atribut seperti nilai akademik, tingkat kehadiran, jenis kelamin, jurusan, dan hasil evaluasi praktik kerja industri (PKL). Kedua algoritma diuji untuk mengklasifikasikan apakah seorang siswa memiliki potensi besar untuk bersaing di dunia kerja. Hasil evaluasi menunjukkan bahwa *K-NN* memberikan akurasi tertinggi sebesar 98.22%, dengan *precision* 99.38% dan *recall* 98.77%, sedangkan *Naïve Bayes* mencatat akurasi 97.66%, dengan *precision* sempurna sebesar 100% dan *recall* 97.59%. Penelitian yang dilakukan oleh Vignesh et al., (2021) mengembangkan sistem bimbingan karir cerdas berbasis web yang memanfaatkan algoritma machine learning untuk membantu siswa memilih jalur karir yang sesuai dengan minat dan kemampuan mereka. Sistem ini menggunakan atribut seperti nilai akademik, minat karir, hasil tes kepribadian, dan keterampilan teknis sebagai input untuk merekomendasikan bidang studi atau karir yang paling cocok bagi pengguna. Beberapa algoritma klasifikasi yang diterapkan dalam sistem ini meliputi *K-Nearest Neighbors (K-NN), Support Vector Machine (SVM),* dan *Naïve Bayes*. Hasil evaluasi menunjukkan bahwa algoritma *K-NN* mencapai akurasi tertinggi, melebihi 90%, dalam mengklasifikasikan jalur karir yang sesuai

2.2.3 Kompetensi Lulusan Siswa Sekolah Menengah Kejuruan (SMK)

Lulusan siswa SMK diharapkan dapat memberikan nilai tambah dibandingkan dengan siswa lulusan SMA dimana keterampilan dalam dunia kerja atau bisnis hanya didapat di SMK. Menurut penelitian yang dilakukan oleh Sutianah, (2020), tedapat 8 kompetensi yang harus dimililki oleh lulusan SMK, antara lain :

1. *Communication Skills* (Kemampuan Komunikasi)

   Kemampuan yang digunakan dalam berkomunikasi kepada orang lain secara efektif baik secara lisan maupun tulisan dengan dalam berbagai konteks secara langsung , digital, maupun dalam lintas budaya.

2. *Critical and Creative Thinking* (Berpikir Kritis dan Kreatif)

   Kemampuan untuk menganalisa, mengevaluasi, mengidentifikasi suatu kondisi dan dapat memunculkan ide yang kreatir dan terbaru serta dapat memberikan solusi yang tepat dalam pengambilan keputusan yang komplek dan inovatif.

3. *Inquiry/Reasoning Skills* (Keterampilan Penalaran dan Investigasi)

   Kemampuan untuk berfikir logis, menyusun hipotesis, menganalisa sebuah data dan menyimpulkan berdasarkan bukti-bukti yang ada

4. *Interpersonal Skills* (Keterampilan Interpersonal)

   Kemampuan yang digunakan dalam tim kerja dimana membutuhkan jiwa kepemimpinan dan social dalam menjalin hubungan antar tim.

5. *Multicultural/Multilingual Literacy* (Literasi Multikultural dan Multibahasa)

   Kemampuan dalam menghormati, memahami, menghargai dalam berinteraksi dengan berbagai individu dari berbagai latar belakang.

6. *Problem Solving* (Pemecahan Masalah)

Kemampuan memberikan solusi terbaik dengan cara mengidentifikasi masalah, mengembangkan dan mengevaluasi solusi alternative berdasarkan data yang ada.

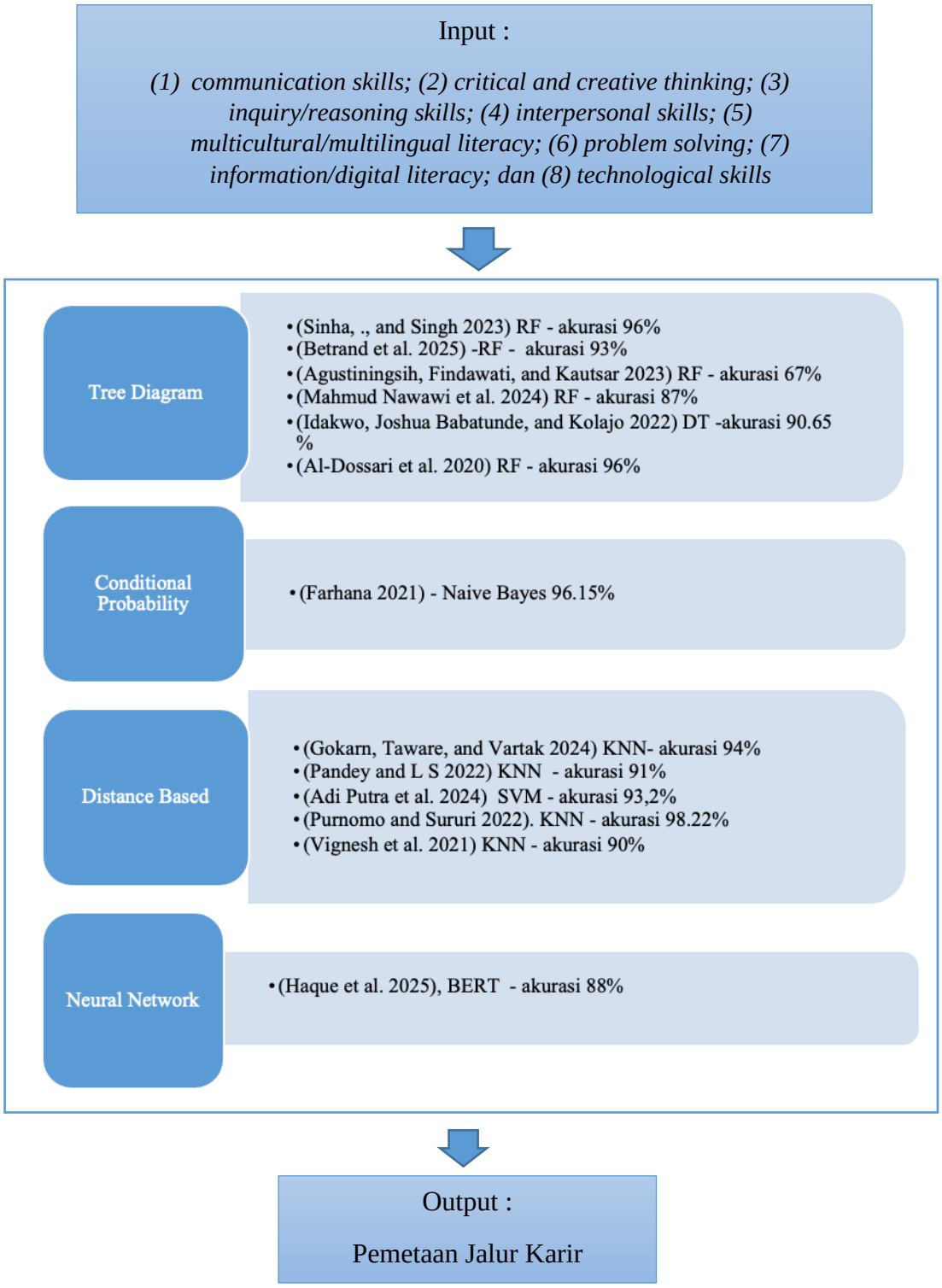
7. *Information/Digital Literacy* (Literasi Informasi dan Digital)

Kemampuan dalam menggunakan, menemukan, dan menciptakan informasi yang berbasis teknologi digital yang memiliki etika dalam berkomunikasi .

8. *Technological Skills* (Keterampilan Teknologi)

Kemampuan yang digunakan dalam menciptakan solusi dengan cara menggunakan, dan mengembangkan teknologi untuk menyelesaikan tugas serta peguasaan perangkat lunak dan ketrampilan pemrograman dasar.

2.3 Kerangka Teori



Gambar 2. 1 Kerangka Teori

Berdasarkan kerangka teori yang disajikan Gambar 2.1 maka dapat ditentukan langkah-langkah yang akan dilakukan dalam menyusun penelitian ini, langkah-langkah tersebut adalah sebagai berikut :

a. Keterbaruan pada penelitian ini adalah data yang akan diolah merupakan data lulusan siswa SMK Negeri 1 Wonorejo yang diambil dari bagian BKK yang memuat 8 kompetensi pokok siswa SMK meliputi *communication skills, critical and creative thinking, inquiry/reasoning skills, interpersonal skills, multicultural/multilingual literacy, problem solving, information/digital literacy dan technological skills* (Sutianah, 2020) sebagai data utamanya, dimana data ini belum pernah diteliti oleh peneliti lain sebelumnya.
b. Referensi atau jurnal yang digunakan dalam penyusunan penelitian ini dapat dilihat pada tabel 2.1

Tabel 2. 1 Daftar Jurnal

No	Sumber	Input (Variabel Bebas)	Output (Variabel Terikat)	Metode (Pre-Process)	Metode (Main)	Metode (Post-Process)	Hasil
1	Ankita Sinha et al. (2023)	Nilai akademik, minat, kepribadian, partisipasi lomba, sertifikat	Rekomendasi jalur karir / pekerjaan	*Data cleaning, One-Hot Encoding*	*Random Forest, SVM, Decision Tree, Adaboost*	Evaluasi akurasi model, perbandingan performa	Random Forest memiliki akurasi tertinggi dibanding model lain
2	Al-Dossari et al. (2020), ETASR Vol.10 No.6	Soft skills (e.g. communication, logic), technical skills (e.g. programming, databases), programming languages, academic major	*Recommended career path (Developer, Analyst, Engineer)*	*Data cleaning, encoding, missing value handling, skill normalization, job/major categorization*	*Supervised ML algorithms: KNN, Decision Tree, Bagging, Gradient Boosting, XGBoost*	Oversampling (balancing dataset), accuracy evaluation using confusion matrix	*XGBoost with balanced dataset achieved highest accuracy: 70.47%*
3	Chidi Ukamaka Betrand et al. (2025)	Minat, kepribadian, hasil tes IQ/EQ, atribut akademik	Rekomendasi karir yang sesuai	*Data cleaning, SelectKBest, feature engineering*	*Decision Tree, Random Forest, Naïve Bayes*	Evaluasi akurasi, *precision, recall, F1-score, confusion matrix*	Random Forest memiliki akurasi tertinggi dan digunakan untuk prediksi real-time
4	Nawawi et al. (2024)	Gender, SSC Percentage, SSC Board, HSC Percentage, HSC Board, HSC Subject, Degree Percentage, Undergrad Degree, Work Experience, Emp Test Percentage, Specialization, MBA Percent	Status *(Placed / Not Placed)*	*Data cleaning, normalisasi, split 80:20, cross-validation*	*Random Forest, Decision Tree, Naïve Bayes, KNN, SVM*	*Confusion Matrix, AUC/ROC, Feature Importance, Visualisasi Tree*	Random Forest terbaik: Akurasi 87%, AUC 0.93; fitur paling penting: SSC Percentage
5	Gendis A.D. et al. (2021)	Nilai, minat, jurusan, kepribadian, tes kompetensi, data Pendidikan	Rekomendasi pekerjaan atau jalur karir	*Data selection, data transformation*	*Decision Tree C4.5*	Visualisasi pohon keputusan, evaluasi akurasi	Model C4.5 memiliki akurasi 84.09% dalam merekomendasikan jalur karir siswa SMK
6	Idakwo et al. (2022)	Minat, bakat, nilai akademik, keterampilan interpersonal	Rekomendasi jalur karir	Pembersihan data,	*Naive Bayes, Decision Tree, K-*	Evaluasi model menggunakan	*Bagging Ensemble* menghasilkan akurasi

No	Sumber	Input (Variabel Bebas)	Output (Variabel Terikat)	Metode (Pre-Process)	Metode (Main)	Metode (Post-Process)	Hasil
			bidang *IT*	penghapusan atribut irrelevan (misalnya gender, email), normalisasi	*NN, SVM,* dan *Bagging Ensemble*	confusion matrix dan metrik (*Accuracy, Precision, Recall, F1 Score*)	terbaik sebesar 90.65%, mengungguli *Decision Tree* (82.01%) dan metode lainnya
7	Farhana, S. (2021)	*Gender, age, class level, academic data, research activities*	Kategori performa akademik staf universitas	Penyaringan data, pembagian data (*training & testing*)	Modified Naive Bayes	Evaluasi akurasi (dibandingkan dengan *NB* & *Decision Tree*)	Akurasi Modified *Naive Bayes*: 93.7%; lebih baik dari *Naive Bayes* dan *Decision Tree* pada kondisi tertentu
8	Sanil Gokarn et al. (2024)	Data minat, nilai akademik, keterampilan personal, hasil kuis bidang tertentu	Rekomendasi jalur karir (3 pilihan teratas)	Penggabungan dua dataset (nilai dan minat), normalisasi nilai numerik menjadi kategorikal, pembagian data 70:30	KNN, Decision Tree, Random Forest, SVM (dibandingkan)	Evaluasi akurasi model, *GridSearch* untuk *tuning hyperparameter*, integrasi ke portal web/mobile	*KNN* akurasi 94.10%, *Decision Tree* 89.06%, *SVM* 86.32%, *Random Forest* 81.05%
9	Kumari et al. (2021)	Nilai akademik, minat karir, tingkat pendidikan, jawaban kuis, atribut soft skill	Rekomendasi karir terbaik	*Cleaning data, label encoding* untuk fitur kategorikal, normalisasi	*Random Forest, Decision Tree, Naive Bayes*	Evaluasi performa model (akurasi, precision, recall, f1-score)	*Random Forest* memberikan hasil terbaik dengan akurasi 95%, lebih tinggi dibanding metode lain
10	Sangeetha and Sumathi (2018)	Nilai akademik, minat, kemampuan kognitif, tipe kepribadian	Rekomendasi jalur karir	Data cleaning, normalisasi, klasifikasi atribut	*Naive Bayes, C4.5, K-Means, Neural Network*	Perbandingan hasil klasifikasi, pengujian akurasi, visualisasi	*C4.5* menunjukkan hasil terbaik dalam akurasi dan efisiensi klasifikasi karir siswa
11	Vignesh et	Skor hasil kuis: *analytical*	Rekomendasi	Pembuatan	*K-Nearest*	Evaluasi	*KNN* akurasi:

No	Sumber	Input (Variabel Bebas)	Output (Variabel Terikat)	Metode (Pre-Process)	Metode (Main)	Metode (Post-Process)	Hasil
	al. (2021)	*skills, logical reasoning, mathematical skills, problem solving, programming, creativity, hardware skills*	jurusan/departemen (*CSE, ECE, EEE, MECH*)	dataset manual, klasifikasi skill sebagai core/sub-skill, normalisasi numerik	*Neighbor* untuk klasifikasi; K-Means untuk rekomendasi tambahan	menggunakan *confusion matrix, f-measure, success rate per cluster*	94.10%, *F-measure* tertinggi: *MECH* (0.9849); sistem rekomendasi berhasil menyarankan jurusan primer, sekunder, dan tersier
12	Purnomo & Sururi (2022)	Nilai Ujian Kompetensi, Nilai PKL, Disiplin, Tanggung Jawab, Sikap, Kemampuan Komunikasi	Kemampuan siswa bersaing di dunia kerja (mampu / belum mampu)	*Data Cleaning, Data Selection, Standar Deviasi, Mean*	*Naïve Bayes, K-Nearest Neighbor* (K=3 dan K=5)	*Confusion Matrix (Evaluasi:* Akurasi, *Precision, Recall)*	*KNN* Akurasi 98.22%, *Precision* 99.38%, *Recall* 98.77%; *Naive Bayes* Akurasi 94.67%, *Precision* 98.73%, *Recall* 95.71%
13	Adiputra et al. (2024)	Nilai akademik, kehadiran, aktivitas ekstrakurikuler	Kategori kinerja siswa (Baik, Cukup, Kurang)	Pembersihan data, normalisasi, encoding variabel kategorikal, split data	*Support Vector Machines (SVM), Naive Bayes, k-Nearest Neighbors (k-NN)*	*Evaluasi dengan confusion matrix,* akurasi, *presisi, recall, F1-score*	*SVM*: Akurasi 93.2%, *F1-Score* 93.1%; *Naive Bayes:* Akurasi 86.2%, *F1-Score* 86.4%; *K-NN*: Akurasi 81.0%, *F1-Score* 80%

Dari Gambar 2.1 dapat disimpulkan bahwa metode *random forest (RF)* dan *K-Nearst Neighbors* merupakan metode yang unggul dalam kasus – kasus klasifikasi pemetaan jalur karis lulusan SMK. Masing – masing metode memiliki kelebihan, random forest memiliki akurasi yang tinggi, dapat mengatasi *over fitting, Robust* terhadap *Data Noise* dan *Outlier*, dapat mengatasi data besar dan tinggi, serta dapat mengatasi data tidak seimbang (Breiman, 2001). Sedangkan metode K-Nearst Neighbors memiliki kelebihan mudah diimplementasikan, tidak membutuhkan pelatihan, dapat menggunakan berbagai metode jarak, fleksibel, dapat digunakan dengan data multikategorial (James et al., 2017)..

c. Hasil dari penelitian ini akan digunakan untuk memetakan jalur karir pada lulusan siswa SMK Negeri 1 Wonorejo.

BAB III

METODOLOGI PENELITIAN

Pada penelitian ini menggunakan pendekatan *comparative.* Pendekatan tersebut dipilih karena pada penelitian ini membandingkan beberapa algoritma klasifikasi pada pembelajaran mesin yakni membandingkan algoritma *random forest* dengan algoritma *K-nearest neighbors* sehingga diharapkan dapat menyelesaikan permasalahan pemetaan jalur karir siswa dengan akurasi yang tinggi.

3.1 Deskripsi Data

Pada penelitian ini, data yang digunakan diperoleh dari Bagian Layanan Bursa Kerja Khusus (BKK) SMK Negeri 1 Wonorejo. Data tersebut merupakan hasil tes yang dilakukan oleh pihak BKK terhadap siswa SMK dengan tujuan untuk memetakan potensi serta memberikan rekomendasi pekerjaan yang sesuai dengan kompetensi masing-masing siswa.

Data berbentuk tabel dengan total 240 baris (record) dan 13 kolom (atribut). Setiap baris merepresentasikan seorang siswa, sedangkan setiap kolom merepresentasikan variabel atau kriteria penilaian. Adapun struktur data meliputi:

1. **No** → Nomor urut data.
2. **NISN** → Nomor Induk Siswa Nasional.
3. **Nama** → Identitas siswa.
4. **Jurusan** → Program keahlian siswa di SMK.
5. **Communication Skills** → Kemampuan komunikasi siswa.

6. **Critical and Creative Thinking** → Kemampuan berpikir kritis dan kreatif.
7. **Inquiry/Reasoning Skills** → Kemampuan penalaran dan penyelidikan.
8. **Interpersonal Skills** → Kemampuan berinteraksi dengan orang lain.
9. **Multicultural/Multilingual Literacy** → Literasi budaya dan bahasa.
10. **Problem Solving** → Kemampuan memecahkan masalah.
11. **Information/Digital Literacy** → Kemampuan menggunakan informasi dan literasi digital.
12. **Technological Skills** → Keterampilan dalam teknologi.
13. **Pekerjaan** → Rekomendasi jenis pekerjaan yang sesuai (variabel target/kelas).

Metode pengumpulan data dilakukan melalui tes penilaian kompetensi yang diselenggarakan oleh BKK. Tes ini dirancang untuk mengukur aspek keterampilan komunikasi, berpikir kritis, penalaran, interpersonal, literasi budaya dan bahasa, pemecahan masalah, literasi digital, hingga keterampilan teknologi. Hasil tes kemudian diproses menjadi data tabular yang siap digunakan sebagai basis analisis.

Dengan karakteristik tersebut, data ini bersifat kuantitatif (nilai skor kompetensi) dan kategorikal (jurusan dan rekomendasi pekerjaan). Data inilah yang selanjutnya digunakan sebagai dataset utama dalam proses pemodelan dengan algoritma Random Forest dan K-Nearest Neighbors (KNN) untuk menghasilkan sistem pemetaan jalur karir lulusan SMK yang lebih akurat. Pada tabel 3.1 ditunjukkan contoh data yang digunakan pada penelitian ini.

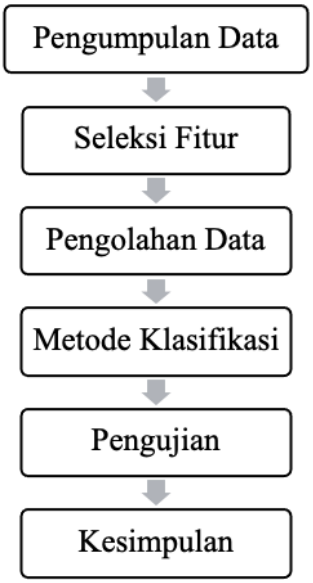
Tabel 3. 1 Data Hasil Tes Pemetaan Jalur Karir Lulusan SMK

No	NISN	Nama	Jurusan	Pekerjaan	*(1) communication skills*	*(2) critical and creative thinking*	*(3) inquiry/ reasoning skills*	*(4) interpersonal skills*	*(5) multicultural/multilingual literacy*	*(6) problem solving*	*(7) information/ digital literacy*	*(8) technological skills*
1	0058200054	Akhmad Masduki Mahfud	Akuntansi dan Keuangan Lembaga	admin	10	12	10	20	8	12	22	18
2	0042827704	Barirotus Sariroh	Akuntansi dan Keuangan Lembaga	guru	27	12	16	22	20	20	17	15
3	0053116382	CHURIN 'AINIYAH	Akuntansi dan Keuangan Lembaga	operator produksi	27	12	14	22	19	20	17	18
4	0021684209	DIAN JULIATUL SAFITRI	Akuntansi dan Keuangan Lembaga	admin	11	12	10	20	8	12	22	18
5	0048709464	HIMATUL ULYA	Akuntansi dan Keuangan Lembaga	admin	12	12	11	22	8	15	23	18
6	0045986351	INDAH SAFITRI	Akuntansi dan Keuangan Lembaga	operator produksi	27	12	16	22	20	22	17	15
7	0047241058	IZZA WARDATUL MUTFIA	Akuntansi dan Keuangan Lembaga	operator produksi	27	10	16	21	20	20	16	16
8	0046135110	M. Abid Aulia Akbar	Akuntansi dan Keuangan Lembaga	operator produksi	28	12	16	22	21	20	17	15
9	0049920477	MARIYATI	Akuntansi dan Keuangan Lembaga	guru	27	10	15	21	20	19	15	15
10	0052359406	NIKMATUL AFIYAH	Akuntansi dan Keuangan	operator produksi	27	12	16	22	20	20	17	15

No	NISN	Nama	Jurusan	Pekerjaan	*(1) communication skills*	*(2) critical and creative thinking*	*(3) inquiry/ reasoning skills*	*(4) interpersonal skills*	*(5) multicultural/multilingual literacy*	*(6) problem solving*	*(7) information/ digital literacy*	*(8) technological skills*
			Lembaga									
11	0046787312	NUR LAILA KHODIJAH	Akuntansi dan Keuangan Lembaga	admin	9	13	9	21	9	12	22	15
12	0045686382	Rizal Farid Maulana	Akuntansi dan Keuangan Lembaga	operator produksi	27	11	16	20	20	20	18	14
13	3051400926	RIZAL MUKHIBUR ROKHMAN	Akuntansi dan Keuangan Lembaga	operator produksi	26	12	16	22	20	20	17	15
14	0049471778	ROCHMANIA PUTRI ARUNITA	Akuntansi dan Keuangan Lembaga	guru	26	12	15	22	19	20	15	16
15	0051071805	SHOFIYAH	Akuntansi dan Keuangan Lembaga	admin	10	15	10	20	8	12	22	18
16	0045001570	SILVI AGUSTINA	Akuntansi dan Keuangan Lembaga	admin	11	12	10	20	8	12	22	17
...												
240	0043162503	SITI NUR RAHMAWATI	Akuntansi dan Keuangan Lembaga	designer	27	20	19	15	8	22	26	25

3.2 Desain Penelitian

Kerangka penelitian yang terdapat pada penelitian ini terdiri dari beberapa tahapan yang meliputi 1) pengumpulan data, 2) pemilihan fitur , 3) pengolahan data, 4) klasifikasi, 5) pengujian, dan 6) Kesimpulan.



Gambar 3. 1 Alur Penelitian

Pada penelitian ini berfokus menyelesaikan permasalahan sistem pemetaan jalur karir lulusan smk menggunakan metode klasifikasi. Pada gambar 3.1 ditunjukkan Langkah-langkah yang digunakan untuk menyelesaikan permasalahan tersebut. Langkah awal yang dilakukan adalah dengan mengumpulkan data. Kemudian data yang telah terkumpul akan dilakukan proses pengolahan data sehingga diharapkan data tersebut dapat digunakan pada tahap berikutnya. Tahap pemilihan fitur merupakan tahapan yang digunakan untuk menyeleksi fitur atau parameter yang akan dijadikan sebagai input untuk algoritma klasifikasi. Pada tahap klasifikasi akan menguji 2 algoritma yakni random forest dan algoritma k-nearest neighboor dengan menggunakan jumlah data yang sama. Hasil dari hasil klasifikasi tersebut akan dilanjutkan proses pengujian untuk setiap algoritma klasifikasi yang digunakan dengan metode

confution matriks. Pada tahap pengujian akan didapatkan nilai akurasi, *precision*, *recall* dan *F1 score* sehingga dapat disimpulkan metode klasifikasi apa yang terbaik untuk menyelesaikan permasalahan sistem pemeteaan jalur karir lulusan SMK.

3.2.1 Pemilihan Fitur

Proses seleksi fitur yang dilakukan pada penelitian ini adalah dengan memilih fitur-fitur yang mempengaruhi terhadap keputusan dalam menilai jenis pekerjaan yang tepat untuk siswa. Berdasarkan tujuan yang ingin dicapai pada penelitian ini maka fitur yang terpilih meliputi jurusan, pekerjaan, dan 8 kompetensi yang diharapkan dimiliki oleh lulusan SMK yang meliputi communication skills, critical and creative thinking, inquiry/reasoning skills, interpersonal skills, multicultural/multilingual literacy, problem solving, information/digital literacy, technological skills. Sedangkan fitur lainnya yang terdiri dari No, NISN, Nama akan dihapus karena fitur tersebut tidak memiliki makna dan hanya menunjukkan identitas data.

Fitur-fitur yang terpilih tersebut selanjutnya akan di bagi kedalam variable *independent* (X) dan variabel *dependen* (Y). fitur yang digolongkan sebagai Variable indepenen (X) antara lain jurusan, *communication skills, critical and creative thinking, inquiry/reasoning skills, interpersonal skills, multicultural/multilingual literacy, problem solving, information/digital literacy, technological skills.* Sedangkan yang menjadi variabel *dependen* (Y) sekaligus sebagai label atau klas untuk setiap data adalah fitur pekerjaan. Pada table 3.2 ditunjukkan hasil seleksi fitur.

Tabel 3. 2 Hasil Seleksi Fitur

No.	Nama Fitur	Status	Jenis Variabel
1.	No.	Dihapus	-
2.	NISN	Dihapus	-
3.	Nama	Dihapus	-
4.	Jurusan	Dipertahankan	*Independent (x)*
5.	Pekerjaan	Dipertahankan	*Dependen (y)*
6.	*Communication skills*	Dipertahankan	*Independent (x)*
7.	*Critical and creative thinking*	Dipertahankan	*Independent (x)*
8.	*Inquiry/reasoning skills*	Dipertahankan	*Independent (x)*
9.	*Interpersonal skills*	Dipertahankan	*Independent (x)*
10.	*Multicultural/multilingual literacy*	Dipertahankan	*Independent (x)*
11.	*Problem solving*	Dipertahankan	*Independent (x)*
12.	*Information/digital literacy*	Dipertahankan	*Independent (x)*
13.	*Technological skills*	Dipertahankan	*Independent (x)*

3.2.2 Pengolahan Data

Pengolahan data yang dilakukan pada penelitian ini digunakan untuk mengubah nilai dari setiap fitur menjadi nilai numeric atau angka. Hal tersebut dilakukan karena algoritma klasifikasi yang digunakan pada penelitian ini yang terdiri dari algoritma *random forest* dan *k-nearest neighboar* hanya bisa menerima masukan data berupa angka. Pada penelitian ini proses pengolah data dilakukan dengan mengubah nilai yang terdapat pada fitur jurusan dan pekerjaan yang bertipe kategori menjadi angka dengan Teknik *One Hot Encoding* untuk fitur jurusan dan *Label Encoding* untuk fitur pekerjaan. Teknik *One Hot Encoding* yang diterapkan pada fitur jurusan bekerja dengan cara mengubah nilai kategori pada fitur tersebut menjadi fitur baru pada variabel *independent* (X) dengan format nama fitur “jurusan_(nama kategori)”.

Tabel 3. 3 One Hot Encoding untuk Fitur Jurusan

No	Kategori	Fitur Baru
1	Akuntansi dan Keuangan Lembaga	Jurusan_Akuntansi dan Keuangan Lembaga
2	Multimedia	Jurusan_Multimedia
3	Teknik dan Bisnis Sepeda Motor	Jurusan_Teknik dan Bisnis Sepeda Motor

4	Teknik Kendaraan Ringan Otomotif	Jurusan_Teknik Kendaraan Ringan Otomotif
5	Teknik Komputer dan Jaringan	Jurusan_Teknik Komputer dan Jaringan

Pada tabel 3.4 ditunjukkan proses *One Hot Encoding* dengan membuat fitur baru serta pengisian nilai untuk setiap fitur baru tersebut. Kita dapat melihat bahwa fitur baru tersebut akan bernilai 1 jika nilai pada fitur jurusan sama dengan nama fiturnya dan selain itu akan bernilai 0. Selanjutnya fitur jurusan akan dihapus. Teknik *one hot encoding* diterapkan pada fitur jurusan karena nilainya berupa kategori yang tidak bertingkat.

Tabel 3. 4 Pelabelan Encoding pada Kolom Pekerjaan

No	Pekerjaan	No Encoding
1	Admin	0
2	Guru	1
3	Operator Produksi	2
4	Designer	3
5	Mekanik/Teknisi	4

Label encoding yang diterapkan pada fitur pekerjaan bekerja dengan cara mengubah nilai kategori menjadi nilai urut. Pada Tabel 3.4 ditunjukkan pengubahan data dari nilai kategori menjadi nilai angka. Teknik ini dipilih karena fitur pekerjaan merupakan fitur yang menjadi variabel *dependen* (Y) sehingga pada penelitian ini fitur tersebut hanya bertujuan sebagai label data. Pada Tabel 3.5 ditunjukkan hasil akhir dataset setelah dilakukan pengolahan data, dimana fitur yang dijadikan sebagai variable *independent* (X) bertambah jumlahnya menjadi 12 fitur.

Tabel 3. 5 Dataset Setelah Dilakukan Pengolahan Data

Jurusan_Akuntansi dan Keuangan Lembaga	Jurusan_Multimedia	Jurusan_Teknik dan Bisnis Sepeda Motor	Jurusan_Teknik Kendaraan Ringan Otomotif	Jurusan_Teknik Komputer dan Jaringan	pekerjaan	*(1) communication skills*	*(2) critical and creative thinking*	*(3) inquiry/reasoning skills*	*(4) interpersonal skills*	*(5) multicultural/ multilingual literacy*	*(6) problem solving*	*(7) information/digital literacy*	*(8) technological skills*
1	0	0	0	0	0	10	12	10	20	8	12	22	18
1	0	0	0	0	1	27	12	16	22	20	20	17	15
1	0	0	0	0	2	27	12	14	22	19	20	17	18
1	0	0	0	0	0	11	12	10	20	8	12	22	18
1	0	0	0	0	0	12	12	11	22	8	15	23	18
1	0	0	0	0	2	27	12	16	22	20	22	17	15
1	0	0	0	0	2	27	10	16	21	20	20	16	16
1	0	0	0	0	2	28	12	16	22	21	20	17	15
1	0	0	0	0	1	27	10	15	21	20	19	15	15
1	0	0	0	0	2	27	12	16	22	20	20	17	15
1	0	0	0	0	0	9	13	9	21	9	12	22	15
1	0	0	0	0	2	27	11	16	20	20	20	18	14
1	0	0	0	0	2	26	12	16	22	20	20	17	15
1	0	0	0	0	1	26	12	15	22	19	20	15	16
1	0	0	0	0	0	10	15	10	20	8	12	22	18
1	0	0	0	0	2	27	13	13	22	20	20	17	15
…													
0	0	0	0	1	2	27	12	16	21	20	20	16	15

Proses berikutnya setelah data diolah adalah melakukan pembagian dataset menjadi data latih (*training*) dan data uji (*testing*). Proses pembagian data ini dilakukan dengan mengambil data secara acak sebesar 80% dari seluruh data sebagai data latih. Sedangkan 20% sisa data dijadikan sebagai data uji.

3.2.3 Klasifikasi

Metode klasifikasi yang digunakan pada penelitian ini adalah algoritma *random forest* dan *K-nearest neightbors*. Algoritma tersebut tergolong sebagai algoritma machine learning dimana algoritma tersebut bekerja dengan membentuk model berdasarkan data latih. Adapun penjelasan setiap metode klasifikasi adalah sebagai berikut :

a. Random Forest

Random Forest adalah algoritma *ensemble learning* yang digunakan untuk tugas klasifikasi maupun regresi. Algoritma ini bekerja dengan membangun sejumlah pohon keputusan (*Decision Tree*) selama proses pelatihan, kemudian menggabungkan hasil prediksi dari masing-masing pohon untuk menghasilkan keputusan akhir. Untuk klasifikasi, hasil agregasi dilakukan dengan metode *majority voting*, sedangkan untuk regresi digunakan metode *averaging*.

Secara umum, algoritma *Random Forest* terdiri dari dua tahap utama, yaitu:

- Proses pelatihan (*training*): membentuk model dengan data latih.
- Proses pengujian (*testing*): melakukan prediksi terhadap data uji menggunakan model yang telah dibentuk.

Dapat kita lihat bahwa proses pelatihan algoritma *random forest* memiliki beberapa tahapan antara lain :

- ***Input Data***

  Dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ hasil dari proses pengolahan data, digunakan sebagai data latih.

- ***Bootstrap Sampling***

  Lakukan proses *bootstrap sampling*, yaitu membentuk B subset data secara acak (dengan pengembalian) dari dataset asli. Tiap subset ini akan digunakan untuk membentuk satu pohon keputusan. Pada penelitian ini jumlah pohon keputusan (*n_estimators*) sebanyak 10. Misalkan $D_b \subset D$ adalah subset bootstrap ke-b, di mana b=1,2,...,B

- **Pemilihan Fitur Acak dan Pembentukan Pohon Keputusan**

  Pada setiap simpul selama pembentukan pohon, dipilih secara acak sejumlah fitur $m \subset M$ (di mana M adalah total fitur dalam *dataset*). Kemudian dipilih fitur terbaik dari subset tersebut untuk melakukan pemisahan data. Pada penelitian ini jumlah maksimal fitur yang dipertimbangkan di setiap pemisahan simpul sebanyak 3.

- **Agregasi Hasil (*Ensembling*)**

  Setelah semua pohon terbentuk, data uji x diklasifikasikan dengan cara mengirimkan x ke masing-masing pohon dan menggabungkan hasilnya:

  - **Untuk klasifikasi:**

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), \dots, T_b(x)\} \quad (3.1)$$

Di mana $T_b(x)$ adalah prediksi dari pohon ke-b terhadap data x, dan mode adalah hasil voting terbanyak.

b. K-Nearest Neighbors

Algoritma *K-nearest neighbors* merupakan salah satu algoritma *machine learning* yang digunakan untuk melakukan proses klasifikasi dan regresi. Algoritma ini bekerja dengan memberikan keputusan berdasarkan dari kemiripan data. Algoritma *K-NN* tidak membutuhkan proses pembentukan model sehingga algoritma ini tergolong algoritma *lazy learning*. Pada penelitian ini algoritma *K-NN* bekerja dengan beberapa tahapan berikut :

- Input data :

Menentukan nilai k, dimana pada penelitian ini nilai k yang digunakan adalah 10. Menghitung jarak antara data uji yang akan diprediksi dengan seluruh data training. Pada penelitian ini teknik yang digunakan untuk menghitung jarak adalah *Euclidian Distance*. Rumus *Euclidian Distance* dapat dilihat pada persamaan 3.2 berikut :

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3.2)$$

Tahap berikutnya adalah mengurutkan data berdasarkan nilai jarak secara ascending. Hal ini dilakukan untuk mengetahui peringkat data latih yang memiliki kemiripan terdekat. Ambil data sejumlah k data teratas kemudian lakukan proses *voting* untuk menentukan hasil prediksi.

3.2.4 Pengujian

Pada tahap ini dilakukan proses pengujian terhadap model yang telah dibuat pada tahap sebelumnya. Pengujian ini dilakukan untuk mengetahui seberapa baik model dalam melakukan proses klasifikasi berdasarkan nilai akurasi atau error yang telah didapatkan. Pada penelitian ini pengujian dilakukan dengan mengukur

seberapa baik algoritma *random forest* dan *k nearest neighbors* dalam melakukan klasifikasi dengan menggunakan teknik confusion matrix. Data yang akan diuji untuk mengukur model adalah data *testing*, dimana data tersebut merupakan data yang tidak digunakan pada saat proses *training*. Jumlah data yang dijadikan sebagai data training pada penelitian ini adalah sebesar 20%.

Confusion matrix merupakan salah satu teknik yang digunakan untuk mengukur performa dari sebuah model. Teknik ini sering digunakan karena mampu memberikan gambaran kinerja dari sebuah algoritma dengan membandingkan hasil klasifikasi dengan nilai label actual data. *Confution matriks* juga memiliki kemampuan dalam mendeteksi ketidak seimbangan data. Selain itu, Teknik ini juga bekerja untuk mengevaluasi model dengan klas biner maupun multi klas. *Confusion matrix* memungkinkan evaluasi yang lebih rinci tentang bagaimana model melakukan klasifikasi, dengan memecah hasil prediksi menjadi empat kategori:

True Positives (TP): jumlah lulusan yang bekerja sebagai mekanik, dan sistem benar memprediksi mereka bekerja sebagai mekanik**.**

True Negatives (TN): jumlah lulusan yang tidak bekerja sebagai mekanik, dan sistem benar memprediksi mereka tidak bekerja sebagai mekanik.

False Positives (FP): jumlah lulusan yang tidak bekerja sebagai mekanik, dan sistem salah memprediksi mereka bekerja sebagai mekanik.

False Negatives (FN): jumlah lulusan yang bekerja sebagai mekanik, dan sistem benar memprediksi mereka tidak bekerja sebagai mekanik.

Dari *confusion matrix*, beberapa metrik penting dapat dihitung, seperti: akurasi, presisi, dan *recall*. Akurasi mengukur proporsi prediksi yang benar dari total prediksi yang dilakukan oleh model. Ini memberikan gambaran umum tentang seberapa baik model bekerja dalam memetakan jalur karir lulusan SMK. Akurasi tinggi menunjukkan bahwa model memiliki kemampuan baik dalam memetakan jalur karir lulusan SMK. Rumus akurasi adalah sebagai berikut :

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3.3)$$

Presisi mengukur proporsi prediksi positif yang benar (*true positives*) dari seluruh prediksi positif yang dibuat oleh model. Dalam sistem pemetaan jalur karir presisi membantu memastikan bahwa siswa dengan 8 kompetensi yang dimilikinya sesuai dengan prediksi dengan proporsi yang tinggi. Presisi bertujuan supaya rekomendasi jalur karir tidak salah sasaran. Rumus presisi dapat dilihat pada persamaan 3.4 berikut :

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (3.4)$$

Recall mengukur dari semua data yang sebenarnya *positif*, berapa banyak yang berhasil ditemukan model. Recall tinggi menunjukkan bahwa model efektif dalam memetakan jalur karir lulusan SMK. *Recall* berfungsi agar sistem tidak melewatkan lulusan yang sebenarnya punya potensi di jalur tertentu. Rumus *recall* dapat dilihat pada persamaan 3.5 berikut :

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (3.5)$$

3.2.5 Kesimpulan.

Setelah dilakukan proses pengujian model, tahap berikutnya adalah memberikan simpulan dari perbandingan performa algoritma *random forest dan k nearest neighboor* berdasarkan nilai akurasi yang didapatkan pada tahap pengujian. Hal ini dilakukan untuk memberikan Solusi terhadap permasalahan yang diangkat pada penelitian ini yaitu sistem pemetaan jalur karir lulusan SMK menggunakan algoritma *Random Forest* dan *K Nearest Neighbors*. Selain berdasarkan nilai akurasi yang dihasilkan pada tahap pengujian, pada penelitian ini juga akan menjelaskan seberapa baik performa dari kedua algoritma tersebut dilihat dari berbagai sisi yang dihasilkan oleh *confusion matrix*. Sehingga diharapkan dapat memberikan Gambaran yang jelas dan detail tentang performa kedua algoritma tersebut dalam memecahkan permasalahan sistem pemetaan jalur karir lulusan SMK.

3.3 Desain Eksperiment

Pada tahap ini dilakukan proses desain eksperimen dengan tujuan membandingkan kinerja performa dari 2 algoritma yaitu algoritma *random forest* dan *k-nearest neighbors (KNN)* dalam memecahkan permasalahan pemetaan jalur karir lulusan Sekolah Menengah Kejuruan (SMK). Eksperimen dirancang untuk mengukur dan mengevaluasi performa kinerja dari kedua algoritma klasifikasi tersebut dengan dataset yang sama menggunakan beberapa matrik evaluasi.

3.3.1 Alat Eksperimen

Pada penelitian ini menggunakan beberapa alat untuk melakukan eksperimen

antara lain :

- Bahasa pemrograman menggunakan python
- Lingkungan pengembangan yang digunakan adalah google colab.
- Library yang digunakan adalah scikit-learn untuk implementasi algoritma dan matrik evaluasi, pandas dan numpy untuk proses pengolahan data, serta matplotlib dan seaborn untuk visualisasi.

3.3.2 Pemilihan Parameter Terbaik

Proses pemilihan parameter (*hyperparameter tuning*) merupakan tahap penting untuk mengoptimalkan kinerja dari sebuah algoritma. Pada penelitian proses pemilihan parameter dilakukan dengan menggunakan Teknik *Grid Search*. Terdapat beberapa parameter untuk setiap algoritma yang digunakan pada penelitian ini yang meliputi algoritma *random forest* dan algoritma *K-nearest neighbors* sebagai berikut :

- Parameter random forest meliputi parameter n_estimator (jumlah pohon) dan max_depth (jumlah maksimal kedalaman pohon).
- Parameter *K-Nearest Neighbors* meliputi parameter k (jumlah tetangga terdekat) dan *distance matriks* (rumus penghitung jarak).

Proses pemilihan parameter dilakukan dengan membandingkan nilai dari hasil evaluasi menggunakan *accuracy, precision, recall, dan F1 Score*. Parameter terbaik yang dilihat dari hasil nilai evaluasi dijadikan sebagai parameter utama untuk masing-masing algoritma.

3.3.3 Kasus eksperimen

Pada penelitian ini menggunakan beberapa kasus eksperimen untuk membandingkan performa dari algoritma random forest dan algoritma k-nearest neighbors untuk mendapatkan gambaran yang lebih jelas dari kedua algoritma tersebut. Adapun kasus eksperimen tersebut antara lain :

1. Pengujian Hyperparameter Tunning.

   Pengujian ini dilakukan untuk mendapatkan model dengan konfigurasi terbaik. Pengujian dilakukan dengan Teknik cross validation (Kfold cross falidation. Teknik ini diterapkan pada data training dan matrik F1Score dijadikan sebagai acuan untuk pemilihan konfigurasai terbaik

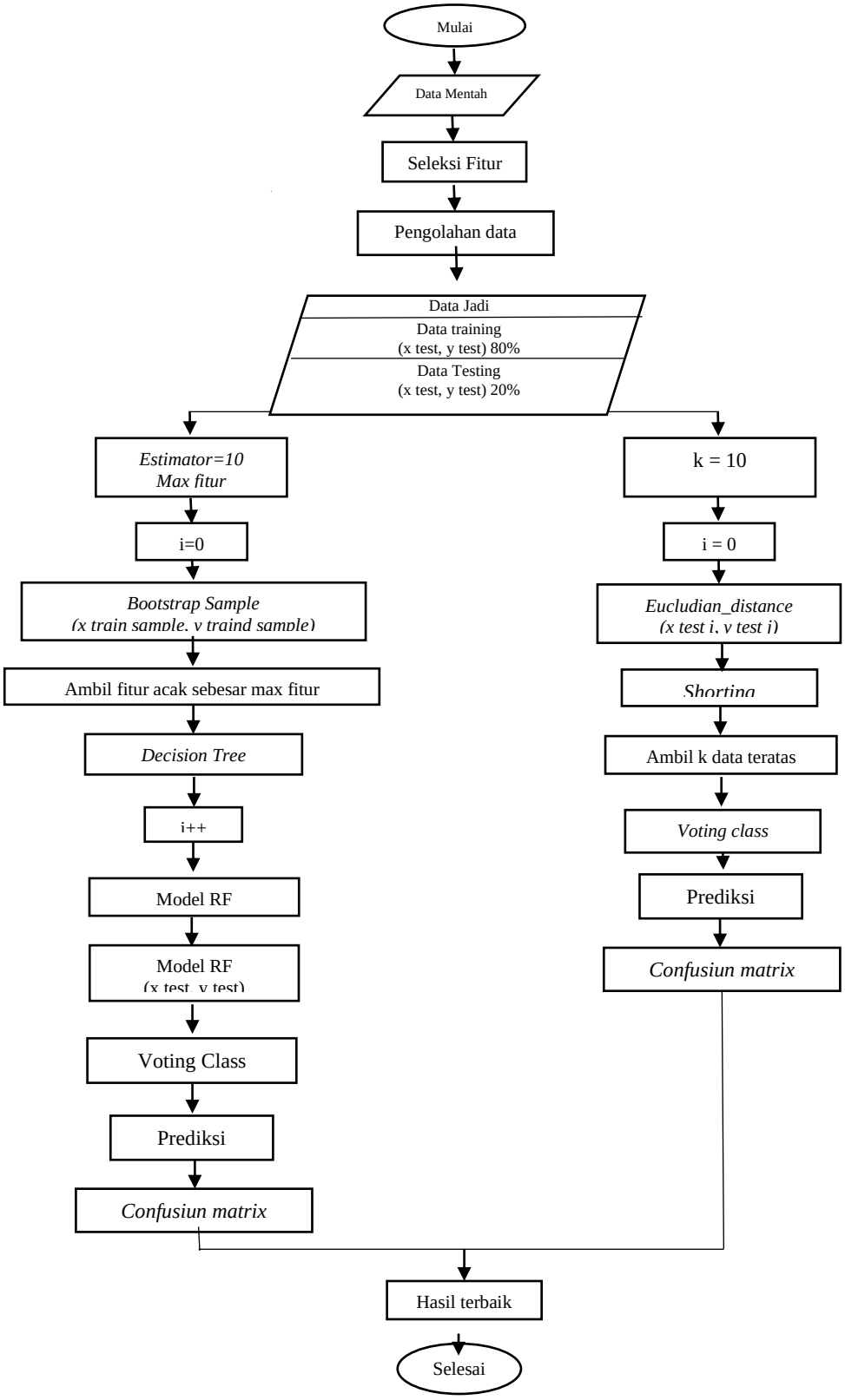
2. Pengujian berdasarkan kelompok jenis kelamin

   Pengujian ini dilakukan menggunakan model dengan konfigurasi parameter terbaik yang telah didapatkan oleh algoritma random forest dan KNN. Data yang digunakan pada penelitian ini dibagi menjadi 2 kelompok yakti jenis kelamin laki-laki dan Perempuan. Miodel akan melakukan pelatihan dan pengujian ulang menggunakan data masing-masing kelompok tersebut.

3. Pengujian berdasarkan kelompok jurusan

   Pengujian ini dilakukan menggunakan model dengan konfigurasi parameter terbaik yang telah didapatkan oleh algoritma random forest dan KNN. Data yang digunakan pada penelitian ini dibagi menjadi 5 kelompok yakni Teknik Kendraan Ringan Otomotif (TKRO), Teknik Komputer dan Jaringan(TKJ), Muktimedia(MM), Teknik dan Bisnis sepeda motor(TBSM), dan akuntansi Keuangan dan lembaga (AKL). Miodel akan melakukan pelatihan dan pengujian ulang menggunakan data masing-masing kelompok tersebut

3.4 Kerangka Konsep Penelitian



Gambar 3. 2 Kerangka Konsep Penelitian

Pada Gambar 3.2 dapat dilihat tahapan-tahapan yang akan dilakukan pada penelitian ini, langkah pertama adalah mengumpulkan data mentah yang diambil dari bagian BKK SMK Negeri 1 Wonorejo. Data mentah tersebut akan diseleksi fitur yaitu pemilihan data mana saja yang diperlukan dalam penelitian ini. Berdasarkan tujuan yang ingin dicapai pada penelitian ini maka fitur yang terpilih meliputi jurusan, pekerjaan, dan 8 kompetensi yang diharapkan dimiliki oleh lulusan SMK yang meliputi *communication skills, critical and creative thinking, inquiry/reasoning skills, interpersonal skills, multicultural/multilingual literacy, problem solving, information/digital literacy, technological skills.* Setelah dilakukan seleksi fitur maka data siap diolah.

Pada proses pengolahan data dilakukan dengan mengubah nilai yang terdapat pada fitur jurusan dan pekerjaan yang bertipe kategori menjadi angka dengan Teknik *One Hot Encoding* untuk fitur jurusan dan *Label Encoding* untuk fitur pekerjaan. Dari hasil proses pengolahan menghasilkan data jadi dimana akan terbagi data *training* dan data *testing*. *Data training* akan dijadikan oleh algoritma *random forest* dan algoritma *k nearest nieghbors* untuk membangun model. Algoritma *random forest* membangun model dengan membuat pohon keputusan sejumlah *n_estimator* dengan fitur sebesar *max_fitur* dimana pohon keputusan yang terbentuk merupakan model yang digunakan untuk memprediksi atau mengklasifikasi data uji. Sedangkan untuk algoritma *k nearest neighbors* membangun model dengan menyimpan *data training* untuk dijadikan sebagai acuan menghitung jarak dengan data uji serta menentukan nilai k untuk menghasilkan prediksi. Setelah model dibentuk dilakukan proses pengujian dengan menggunakan *data testing* sebesar 20%. Pengujian dilakukan dengan

menghitung nilai akurasi hasil klasifikasi yang dihasilkan oleh kedua algoritma tersebut dengan menggunakan *confusion matrix*. Nilai akurasi yang telah dihasilkan oleh masing-masing algoritma akan dibandingkan untuk menarik sebuah kesimpulan algoritma mana yang mampu memberikan nilai akurasi yang lebih baik dalam menyelesaikan permasalahan sistem pemetaan jalur karir lulusan SMK.

3.5 Instrumen Penelitian

Instrumen yang digunakan pada penelitian ini dapat dilihat pada Tabel 3.8 yang berisi *variable dependen* dan *independen*. *Variable independen* ini diproses yang nantinya akan menghasilkan suatu akurasi dimana akurasi tersebut merupakan *variabel dependen*.

Tabel 3. 6 Instrumen Penelitian

Variable Independent	***Main Process***	***Variable Dependen***
Jurusan, 8 Komptensi lulusan siswa SMK	*Random Forest* dan *K-Nears Neighbors*	Akurasi

BAB IV

ALGORITMA RANDOM FOREST

4.1 Pengujian Hyperparameter Tunning

Pada penelitian ini menggunakan algoritma random forest dimana akan dilakukan eksperimen *hyperparameter tunning* terhadap model pembelajaran mesin. *Hyperparameter tunning* untuk algoritma random forest dilakukan dengan mengkonfigurasi beberapa parameter yang meliputi *n_estimator* (jumlah pohon), *max_depth* (kedalaman pohon), dan *max_feature* (teknik pemilihan fitur). Data yang digunakan untuk uji coba *hyperparameter tunning* menggunakan *cross validation* dengan kombinasi data latih mulai dari 60% sampai 90% dari dataset. Evaluasi dilakukan menggunakan metrik utama yaitu *Accuracy, Precision, Recall*, dan *F1 Score*. Hasil evaluasi untuk konfigurasi terbaik dan jumlah persentase data training akan digunakan sebagai model untuk melakukan pengujian dengan data uji. Hasil evaluasi disajikan dalam bentuk tabel ringkasan untuk setiap konfigurasi. Ringkasan hasil evaluasi setiap kombinasi *hyperparameter* untuk algoritma *random forest* ditunjukkan pada Tabel 4.1. Tabel ini menyajikan nilai rata-rata (*mean*), standar deviasi (std), nilai minimum (min), dan maksimum (max) dari masing-masing metrik. Nilai rata-rata menunjukkan performa umum model, sedangkan standar deviasi menggambarkan konsistensi antar pembagian *cross-validation.* Berikut di sajikan kombinasi hasil hyperparameter tunning untuk setiap masing-masing data training:

1. Hasil Hyperparamerter Tunning dengan data training 60%

Dari hasil yang ditunjukkan pada tabel 4.1 dapat dilihat bahwa konfigurasi parameter terbaik berdasarkan hasil akurasi dan evaluasi *F1 Score* sebesar 0.97 dan 0.79 serta nilai presisi sebesar 0.79 dan nilai recall sebesar 0.80. Berdasarkan hal tersebut dapat dilihat bahwa parameter tidak memiliki pengaruh terhadap performa dari algoritma random forest dengan jumlah data training sebesar 60%.

Tabel 4. 1 Hasil Hyperparameter Tunning dengan Data Training 60%

parameter			**Accuracy**				**Prescision**				**Recall**				**F1 Score**			
max_depth	**max_features**	**n_estimators**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
	sqrt	10	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
	sqrt	15	0.92	0.96	0.96	0.94	0.79	0.74	0.79	0.77	0.79	0.77	0.80	0.79	0.79	0.75	0.79	0.78
	sqrt	20	0.92	0.98	0.96	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.79	0.80	0.79	0.79
	log2	10	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
	log2	15	0.92	0.96	0.96	0.94	0.79	0.74	0.79	0.77	0.79	0.77	0.80	0.79	0.79	0.75	0.79	0.78
	log2	20	0.92	0.98	0.96	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.79	0.80	0.79	0.79
5	sqrt	10	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
5	sqrt	15	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
5	sqrt	20	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79

parameter			Accuracy				Prescision				Recall				F1 Score			
max_depth	**max_features**	**n_estimators**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
5	log2	10	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
5	log2	15	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
5	log2	20	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
10	sqrt	10	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
10	sqrt	15	0.92	0.96	0.96	0.94	0.79	0.74	0.79	0.77	0.79	0.77	0.80	0.79	0.79	0.75	0.79	0.78
10	sqrt	20	0.92	0.98	0.96	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.79	0.80	0.79	0.79
10	log2	10	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
10	log2	15	0.92	0.96	0.96	0.94	0.79	0.74	0.79	0.77	0.79	0.77	0.80	0.79	0.79	0.75	0.79	0.78
10	log2	20	0.92	0.98	0.96	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.79	0.80	0.79	0.79

2. Hasil Hyperparamerter Tunning dengan data training 70%

Dari hasil yang ditunjukkan pada tabel 4.2 dapat dilihat bahwa konfigurasi parameter terbaik berdasarkan hasil akurasi dan evaluasi *F1 Score* terdapat pada baris pertama, keempat, ke tiga belas dan ke enam bleas sebesar 0.96 dan 0.84 serta nilai presisi sebesar 0.83 dan nilai recall sebesar 0.86 dengan konfigurasi max_depth sebesar none, 10 dan *n_estimator* sebanyak 10 dan

max_fitur dengan nilai sqrt dan log2. Berdasarkan hal tersebut dapat dilihat bahwa parameter max_fitur tidak memiliki pengaruh terhadap performa dari algoritma random forest dengan jumlah data training 70%.

Tabel 4. 2 Hasil Hyperparameter Tunning dengan Data Training sebesar 70%

parameter			Accuracy				Prescision				Recall				F1 Score			
max_depth	**max_features**	**n_estimators**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
	sqrt	10	0.96	0.96	0.96	0.96	0.90	0.79	0.79	0.83	0.99	0.80	0.80	0.86	0.93	0.79	0.79	0.84
	sqrt	15	0.95	0.96	0.95	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.79	0.79	0.79	0.79
	sqrt	20	0.95	0.96	0.95	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.79	0.79	0.79	0.79
	log2	10	0.96	0.96	0.96	0.96	0.90	0.79	0.79	0.83	0.99	0.80	0.80	0.86	0.93	0.79	0.79	0.84
	log2	15	0.95	0.96	0.95	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.79	0.79	0.79	0.79
	log2	20	0.95	0.96	0.95	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.79	0.79	0.79	0.79
5	sqrt	10	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
5	sqrt	15	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
5	sqrt	20	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
5	log2	10	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
5	log2	15	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
5	log2	20	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
10	sqrt	10	0.96	0.96	0.96	0.96	0.90	0.79	0.79	0.83	0.99	0.80	0.80	0.86	0.93	0.79	0.79	0.84

10	sqrt	15	0.95	0.96	0.95	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.79	0.79	0.79	0.79
10	sqrt	20	0.95	0.96	0.95	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.79	0.79	0.79	0.79
10	log2	10	0.96	0.96	0.96	0.96	0.90	0.79	0.79	0.83	0.99	0.80	0.80	0.86	0.93	0.79	0.79	0.84
10	log2	15	0.95	0.96	0.95	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.79	0.79	0.79	0.79
10	log2	20	0.95	0.96	0.95	0.95	0.79	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.79	0.79	0.79	0.79

3. Hasil Hyperparamerter Tunning dengan data training 80%

Dari hasil yang ditunjukkan pada tabel 4.3 dapat dilihat bahwa konfigurasi parameter terbaik berdasarkan hasil akurasi dan evaluasi *F1 Score* terdapat pada baris ketujuh, dan kesepuluh enam sebesar 0.97 dan 0.90 serta nilai presisi sebesar 0.83 dan nilai recall sebesar 0.86 dengan konfigurasi max_depth sebesar none, 10 dan *n_estimator* sebanyak 10 dan *max_fitur* dengan nilai sqrt dan log2. Berdasarkan hal tersebut dapat dilihat bahwa parameter maxfitur tidak memiliki pengaruh terhadap performa dari algoritma random forest dengan jumlah data training 80%.

Tabel 4. 3 Hasil Hyperparameter Tunning dengan Data Training sebesar 80%

parameter			Accuracy				Prescision				Recall				F1 Score			
max_depth	**max_features**	**n_estimators**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
	sqrt	10	0.99	0.94	0.93	0.95	1.00	0.86	0.84	0.90	0.90	0.86	0.85	0.87	0.93	0.86	0.84	0.88
	sqrt	15	0.99	0.93	0.93	0.95	1.00	0.79	0.79	0.86	0.90	0.79	0.79	0.83	0.93	0.79	0.79	0.84

	sqrt	20	0.97	0.93	0.93	0.94	0.79	0.79	0.79	0.79	0.80	0.79	0.79	0.79	0.80	0.79	0.79	0.79
	log2	10	0.99	0.94	0.93	0.95	1.00	0.86	0.84	0.90	0.90	0.86	0.85	0.87	0.93	0.86	0.84	0.88
	log2	15	0.99	0.93	0.93	0.95	1.00	0.79	0.79	0.86	0.90	0.79	0.79	0.83	0.93	0.79	0.79	0.84
	log2	20	0.97	0.93	0.93	0.94	0.79	0.79	0.79	0.79	0.80	0.79	0.79	0.79	0.80	0.79	0.79	0.79
5	sqrt	10	0.99	0.96	0.97	0.97	1.00	0.89	0.99	0.96	0.90	0.86	0.87	0.88	0.93	0.87	0.90	0.90
5	sqrt	15	0.99	0.94	0.96	0.96	1.00	0.79	0.79	0.86	0.90	0.80	0.80	0.83	0.93	0.79	0.79	0.84
5	sqrt	20	0.97	0.93	0.96	0.95	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.80	0.80	0.79	0.79	0.79
5	log2	10	0.99	0.96	0.97	0.97	1.00	0.89	0.99	0.96	0.90	0.86	0.87	0.88	0.93	0.87	0.90	0.90
5	log2	15	0.99	0.94	0.96	0.96	1.00	0.79	0.79	0.86	0.90	0.80	0.80	0.83	0.93	0.79	0.79	0.84
5	log2	20	0.97	0.93	0.96	0.95	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.80	0.80	0.79	0.79	0.79
10	sqrt	10	0.99	0.94	0.93	0.95	1.00	0.86	0.84	0.90	0.90	0.86	0.85	0.87	0.93	0.86	0.84	0.88
10	sqrt	15	0.99	0.93	0.93	0.95	1.00	0.79	0.79	0.86	0.90	0.79	0.79	0.83	0.93	0.79	0.79	0.84
10	sqrt	20	0.97	0.93	0.96	0.95	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.80	0.80	0.79	0.79	0.79
10	log2	10	0.99	0.94	0.93	0.95	1.00	0.86	0.84	0.90	0.90	0.86	0.85	0.87	0.93	0.86	0.84	0.88
10	log2	15	0.99	0.93	0.93	0.95	1.00	0.79	0.79	0.86	0.90	0.79	0.79	0.83	0.93	0.79	0.79	0.84
10	log2	20	0.97	0.93	0.96	0.95	0.79	0.79	0.79	0.79	0.80	0.79	0.80	0.80	0.80	0.79	0.79	0.79

4. Hasil Hyperparamerter Tunning dengan data training 90%

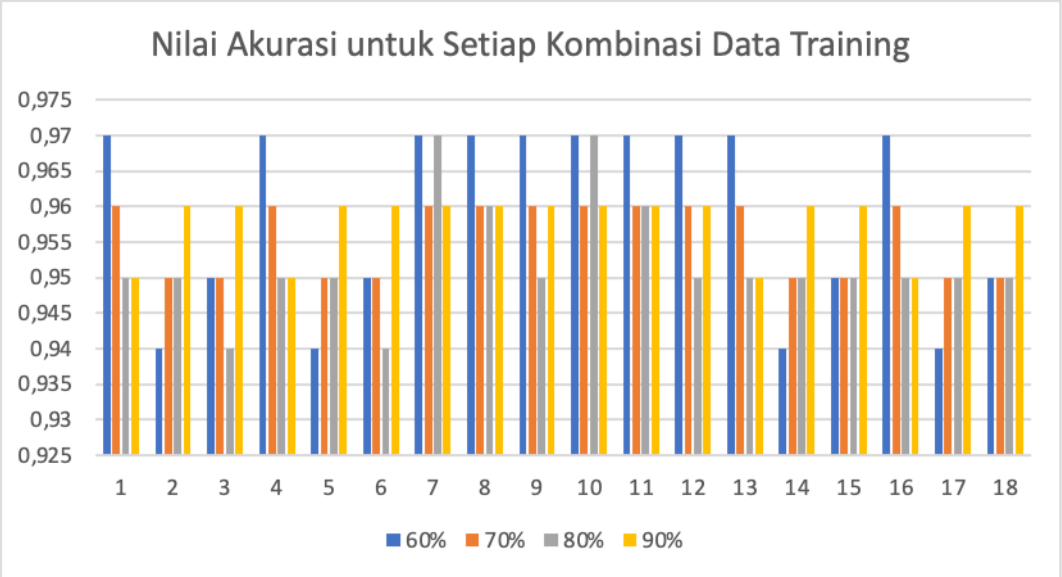
Dari hasil yang ditunjukkan pada tabel 4.4 dapat dilihat bahwa konfigurasi parameter terbaik berdasarkan hasil akurasi dan evaluasi *F1 Score* terdapat pada baris ke tujuh dan ke sepuluh sebesar 0.96 dan 0.84 serta nilai presisi sebesar 0.86 dan nilai recall sebesar 0.83 dengan konfigurasi max_depth sebesar 5, *n_estimator* sebanyak 10 dan *max_fitur* dengan nilai SQRT dan log2. Berdasarkan hal tersebut dapat dilihat bahwa parameter maxfitur tidak memiliki pengaruh terhadap performa dari algoritma random forest dengan jumlah data training sebasar 90%.

Tabel 4. 4 Hasil Hyperparameter Tunning dengan Data Training sebesar 90%

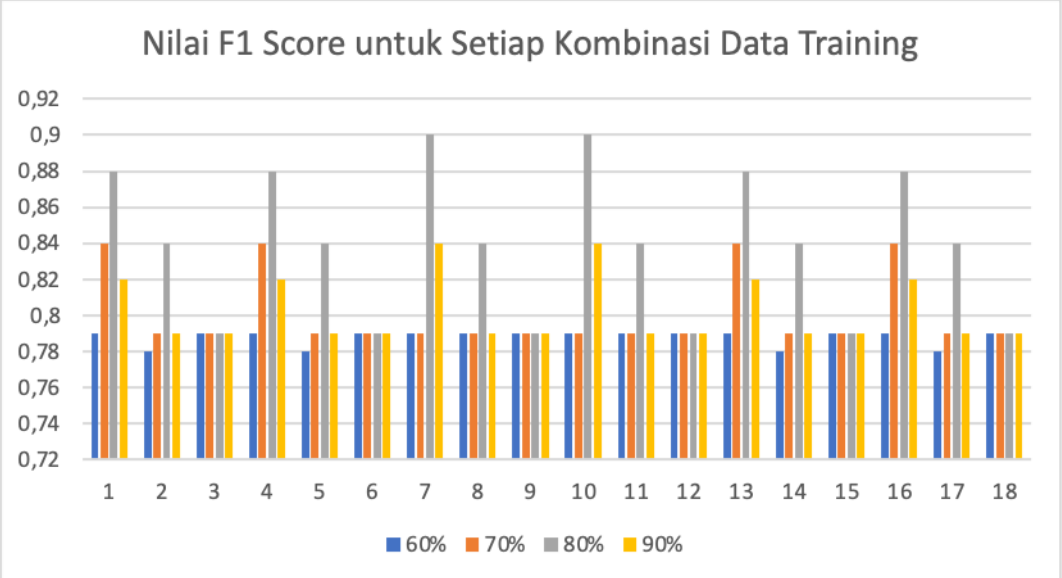
parameter			Accuracy				Prescision				Recall				F1 Score			
max_depth	**max_features**	**n_estimators**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
	sqrt	10	0.94	0.97	0.94	0.95	0.79	0.89	0.78	0.82	0.79	0.89	0.79	0.83	0.79	0.89	0.79	0.82
	sqrt	15	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
	sqrt	20	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
	log2	10	0.94	0.97	0.94	0.95	0.79	0.89	0.78	0.82	0.79	0.89	0.79	0.83	0.79	0.89	0.79	0.82
	log2	15	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
	log2	20	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
5	sqrt	10	0.97	0.98	0.94	0.96	0.79	0.99	0.78	0.86	0.80	0.90	0.79	0.83	0.79	0.93	0.79	0.84
5	sqrt	15	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
5	sqrt	20	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79

5	log2	10	0.97	0.98	0.94	0.96	0.79	0.99	0.78	0.86	0.80	0.90	0.79	0.83	0.79	0.93	0.79	0.84
5	log2	15	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
5	log2	20	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
10	sqrt	10	0.94	0.97	0.94	0.95	0.79	0.89	0.78	0.82	0.79	0.89	0.79	0.83	0.79	0.89	0.79	0.82
10	sqrt	15	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
10	sqrt	20	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
10	log2	10	0.94	0.97	0.94	0.95	0.79	0.89	0.78	0.82	0.79	0.89	0.79	0.83	0.79	0.89	0.79	0.82
10	log2	15	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
10	log2	20	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79

Untuk hasil pengujian hyperparameter tunning untuk algoritma random forest berdasarkan nilai akurasi dapat dilihat pada Gambar 4.1 dan berdasarkan nilai F1 score pada Gambar 4.2.



Gambar 4. 1 Grafik Nilai Akurasi Setiap Kombinasi Data Training



Gambar 4. 2 Grafik Nilai F1 Score Setiap Kombinasi Data Training

Berdasarkan hasil pengujian hyperparameter tunning dengan kombinasi data training dan data testing didapat hasil terbaik pada kombinasi 80% untuk data training dan 20% untuk data testing .Setelah dilakukan proses eksperimen dengan 80% data set atau data latih maka langkah selanjutnya adalah melakukan tes atau ujicoba terhadap 20% dataset atau data uji. Hasil dari ujicoba terhadap algoritma random forest dapat dilihat pada Tabel 4.5.

Tabel 4. 5 Hasil Klasifikasi Algoritma Random Forest Menggunakan Data Uji

No.	Aktual	Prediksi Random Forest
1	Admin	Admin
2	Mekanik/Teknisi	Mekanik/Teknisi
3	Designer	Designer
4	Mekanik/Teknisi	Mekanik/Teknisi
5	Designer	Designer
6	Mekanik/Teknisi	Mekanik/Teknisi
7	Mekanik/Teknisi	Mekanik/Teknisi
8	Mekanik/Teknisi	Mekanik/Teknisi
9	Guru	Guru
10	Admin	Admin
11	Admin	Admin
12	Admin	Admin
13	Mekanik/Teknisi	Operator Produksi
14	Mekanik/Teknisi	Mekanik/Teknisi
15	Mekanik/Teknisi	Mekanik/Teknisi
16	Mekanik/Teknisi	Mekanik/Teknisi
17	Admin	Admin
18	Designer	Designer
19	Admin	Admin
20	Mekanik/Teknisi	Mekanik/Teknisi
21	Mekanik/Teknisi	Mekanik/Teknisi
22	Mekanik/Teknisi	Mekanik/Teknisi
23	Operator Produksi	Mekanik/Teknisi
24	Designer	Designer
25	Mekanik/Teknisi	Mekanik/Teknisi
26	Mekanik/Teknisi	Mekanik/Teknisi
27	Mekanik/Teknisi	Mekanik/Teknisi
28	Designer	Designer

No.	Aktual	Prediksi Random Forest
29	Mekanik/Teknisi	Mekanik/Teknisi
30	Mekanik/Teknisi	Mekanik/Teknisi
31	Mekanik/Teknisi	Mekanik/Teknisi
32	Mekanik/Teknisi	Mekanik/Teknisi
33	Operator Produksi	Mekanik/Teknisi
34	Mekanik/Teknisi	Operator Produksi
35	Guru	Guru
36	Mekanik/Teknisi	Mekanik/Teknisi
37	Designer	Designer
38	Mekanik/Teknisi	Mekanik/Teknisi
39	Mekanik/Teknisi	Mekanik/Teknisi
40	Admin	Admin
41	Admin	Admin
42	Mekanik/Teknisi	Mekanik/Teknisi
43	Mekanik/Teknisi	Operator Produksi
44	Designer	Designer
45	Mekanik/Teknisi	Mekanik/Teknisi
46	Admin	Admin
47	Mekanik/Teknisi	Mekanik/Teknisi
48	Mekanik/Teknisi	Mekanik/Teknisi

Berdasarkan hasil yang ditunjukkan pada tabel 4.5 didapatkan metrik evaluasi F1 *score*, *precision*, *recall* dan *accuracy* untuk *random forest* adalah 0.905, 0.915, 0.895, dan 0.895.

4.2 Pengujian Algoritma Random Forest Berdasarkan Jenis Kelamin

Pengujian algoritma *Random Forest* dilakukan untuk mengukur performa model dalam mengklasifikasikan data berdasarkan kelompok jenis kelamin (Laki-laki/L dan Perempuan/P). Hasil evaluasi performa model disajikan pada Tabel 4.6

Tabel 4. 6 Hasil Pengujian Random Forest Berdasarkan Jenis Kelamin

Kelamin	Proses	Jumlah Data	Akurasi	Precision	Recall	F1 Skor
L	Training	146	1	1	1	1
L	Testing	37	1	1	1	1
P	Training	45	0.98	0.99	0.94	0.95
P	Testing	12	0.92	0.62	0,667	0.64

Model *Random Forest* menunjukkan performa yang sempurna untuk klasifikasi jenis kelamin laki-laki, baik pada data *training* (146 data) maupun data *testing* (37 data). Nilai Akurasi, *Precision*, *Recall*, dan F1 Skor mencapai 1 (atau 100%), yang mengindikasikan bahwa model mampu memprediksi semua data Laki-laki dengan benar tanpa adanya kesalahan klasifikasi pada kelas ini. Kontras dengan hasil tersebut, performa model untuk klasifikasi jenis kelamin Perempuan (P) menunjukkan adanya penurunan kinerja yang signifikan pada data *testing* (12 data). Meskipun data *training* (45 data) menunjukkan performa yang sangat baik dengan Akurasi 0.978 dan F1 Skor 0.959, F1 Skor pada data *testing* turun menjadi 0.641, dengan *Precision* hanya 0.619 dan *Recall* 0.667. Penurunan ini menunjukkan bahwa model mengalami kesulitan dalam menggeneralisasi pola data pada kelompok Perempuan, yang dikonfirmasi oleh nilai *Precision* yang rendah, mengindikasikan sekitar 38% dari data yang diprediksi sebagai P sebenarnya bukan P (*False Positive*). Secara keseluruhan, algoritma *Random Forest* sangat kuat untuk klasifikasi jenis kelamin Laki-laki, namun perlu dicermati bahwa penurunan kinerja pada kelompok Perempuan, terutama karena sampel *testing* yang relatif kecil (12 data), dapat menjadi indikasi adanya bias atau tantangan dalam generalisasi pada kelompok minoritas.

4.3 Pengujian Algoritma Random Forest Berdasarkan Kelompok Jurusan

Algoritma *Random Forest* juga diuji untuk mengukur performa model dalam mengklasifikasikan data berdasarkan kelompok jurusan. Hasil evaluasi disajikan pada Tabel 4.3.

Tabel 4. 7 Hasil Pengujian Random Forest Berdasarkan Jurusan

Jurusan	Proses	Jumlah Data	Akurasi	Precision	Recall	F1 Score
Teknik Kendaraan Ringan Otomotif	Training	60	1	1	1	1
Teknik Kendaraan Ringan Otomotif	Testing	15	0.93	0.75	0.75	0.74
Teknik Komputer dan Jaringan	Training	51	1	1	1	1
Teknik Komputer dan Jaringan	Testing	13	0.92	0.75	0.75	0.74
Multimedia	Training	46	1	1	1	1
Multimedia	Testing	12	1	1	1	1
Teknik dan Bisnis Sepeda Motor	Training	19	1	1	1	1
Teknik dan Bisnis Sepeda Motor	Testing	5	0.8	0.56	0.67	0.6
Akuntansi dan Keuangan Lembaga	Training	15	1	1	1	1
Akuntansi dan Keuangan Lembaga	Testing	4	0.75	0.56	0.67	0.6

Hasil pengujian menunjukkan bahwa algoritma *Random Forest* mencapai performa sempurna (Akurasi, *Precision*, *Recall*, F1 Skor = 1) pada seluruh kelas jurusan di data *training*, yang merupakan indikasi kuat adanya *overfitting* atau model terlalu menghafal data pelatihan. Dalam data *testing*, kinerja model bervariasi. Jurusan Multimedia (MM) mempertahankan performa sempurna dengan F1 Skor 1 pada 12 data *testing*, menjadikannya kinerja terbaik. Sementara itu, Teknik Kendaraan Ringan Otomotif (TKRO) dan Teknik Komputer dan

Jaringan (TKJ) menunjukkan kinerja yang serupa dengan F1 Skor 0.74, menandakan penurunan performa yang signifikan dibandingkan data *training*. Kinerja terendah tercatat pada Teknik dan Bisnis Sepeda Motor (TBSM) dan Akuntansi dan Keuangan Lembaga (AKL), keduanya memiliki F1 Skor 0.60 dan Akurasi masing-masing 0.80 dan 0.75. Kinerja yang rendah ini, ditandai dengan *Precision* 0.56 dan *Recall* 0.67, sangat mungkin disebabkan oleh jumlah data *testing* yang sangat kecil (hanya 5 data untuk TBSM dan 4 data untuk AKL), yang tidak cukup representatif dan memperparah masalah *overfitting*dari data *training* yang sempurna.

BAB V

ALGORITMA K-NEAREST NEIGHBORS

5.1 Pengujian Hyperparameter Tunning

Penelitian ini melakukan *hyperparameter tunning* untuk algoritma KNN dengan mengkonfigurasi beberapa parameter yang meliputi *metric* (jenis perhitungan jarak), *n_neighbors* (tetangga terdekat), dan *weights*. Data yang digunakan untuk uji coba *hyperparameter tunning* menggunakan *cross validation* dengan kombinasi data latih mulai dari 60% sampai 90% dari dataset. Evaluasi dilakukan menggunakan metrik utama yaitu *Accuracy*, *Precision, Recall,* dan *F1 Score*. Hasil evaluasi untuk konfigurasi terbaik dan jumlah persentase data training akan digunakan sebagai model untuk melakukan pengujian dengan data uj. Ringkasan hasil evaluasi setiap kombinasi *hyperparameter* untuk algoritma KNN ditunjukkan pada Tabel 5.1 yang menyajikan nilai rata-rata (*mean*), nilai minimum (min), dan maksimum (max) dari masing-masing metrik. Nilai rata-rata menunjukkan performa umum model, sedangkan standar deviasi menggambarkan konsistensi antar pembagian *cross-validation*. Berikut di sajikan kombinasi hasil hyperparameter tunning untuk setiap masing-masing data training :

1. Hasil hyperparameter tunning dengan dataset sebesar 60%

Dari hasil yang ditunjukkan pada Tabel 5.1 dapat dilihat bahwa konfigurasi parameter terbaik berdasarkan hasil akurasi dan evaluasi *F1 Score* sebesar 0.97 dan 0.79 serta nilai presisi sebesar 0.79 dan nilai recall sebesar 0.80. Dapat dilihat bahwa parameter metric, n_neighbors, dan weight tidak memiliki pengaruh terhadap performa dari algoritma K-nearest neighbors dengan jumlah data training 60%.

Tabel 5. 1 Hasil Hyperparameter Tunning dengan Data Training sebesar 60%

parameter			Accuracy				Precision				Recall				F1 Score			
metric	**n_neighbors**	**weight**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
euclidean	10	uniform	0.97	0.97	0.94	0.96	0.79	0.79	0.74	0.77	0.80	0.80	0.78	0.79	0.79	0.79	0.75	0.78
euclidean	10	distance	0.97	0.97	0.94	0.96	0.79	0.79	0.74	0.77	0.80	0.80	0.78	0.79	0.79	0.79	0.75	0.78
euclidean	15	uniform	0.97	0.94	0.95	0.95	0.79	0.75	0.79	0.78	0.80	0.70	0.80	0.77	0.79	0.71	0.79	0.77
euclidean	15	distance	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
euclidean	20	uniform	0.92	0.91	0.94	0.92	0.75	0.53	0.77	0.68	0.65	0.60	0.73	0.66	0.65	0.56	0.74	0.65
euclidean	20	distance	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
manhattan	10	uniform	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
manhattan	10	distance	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
manhattan	15	uniform	0.97	0.94	0.95	0.95	0.79	0.75	0.79	0.78	0.80	0.70	0.80	0.77	0.79	0.71	0.79	0.77
manhattan	15	distance	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79

parameter			Accuracy				Precision				Recall				F1 Score			
metric	**n_neighbors**	**weight**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
manhattan	20	uniform	0.92	0.92	0.94	0.93	0.75	0.74	0.77	0.75	0.65	0.65	0.73	0.68	0.65	0.65	0.74	0.68
manhattan	20	distance	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
minkowski	10	uniform	0.97	0.97	0.94	0.96	0.79	0.79	0.74	0.77	0.80	0.80	0.78	0.79	0.79	0.79	0.75	0.78
minkowski	10	distance	0.97	0.97	0.94	0.96	0.79	0.79	0.74	0.77	0.80	0.80	0.78	0.79	0.79	0.79	0.75	0.78
minkowski	15	uniform	0.97	0.94	0.95	0.95	0.79	0.75	0.79	0.78	0.80	0.70	0.80	0.77	0.79	0.71	0.79	0.77
minkowski	15	distance	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
minkowski	20	uniform	0.92	0.91	0.94	0.92	0.75	0.53	0.77	0.68	0.65	0.60	0.73	0.66	0.65	0.56	0.74	0.65
minkowski	20	distance	0.97	0.97	0.95	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79

2. Hasil hyperparameter tunning dengan dataset sebesar 70%.

Dari hasil yang ditunjukkan pada Tabel 5.2 dapat dilihat bahwa konfigurasi parameter terbaik berdasarkan hasil akurasi dan evaluasi *F1 Score* sebesar 0.96 dan 0.79 serta nilai presisi sebesar 0.79 dan nilai recall sebesar 0.80. Dapat dilihat bahwa parameter metric, n_neighbors, dan weight tidak memiliki pengaruh terhadap performa dari algoritma nearest neighbors dengan jumlah data training 70%.

Tabel 5. 2 Hasil Hyperparameter Tunning dengan Data Training sebesar 70%

parameter			Accuracy				Precision				Recall				F1 Score			
metric	**n_neighbors**	**weight**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
euclidean	10	uniform	0.95	0.96	0.96	0.96	0.74	0.79	0.79	0.77	0.78	0.80	0.80	0.79	0.75	0.79	0.79	0.78
euclidean	10	distance	0.95	0.96	0.96	0.96	0.74	0.79	0.79	0.77	0.78	0.80	0.80	0.79	0.75	0.79	0.79	0.78
euclidean	15	uniform	0.95	0.91	0.93	0.93	0.77	0.53	0.75	0.68	0.73	0.60	0.67	0.67	0.74	0.56	0.67	0.66
euclidean	15	distance	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
euclidean	20	uniform	0.91	0.91	0.91	0.91	0.53	0.53	0.54	0.54	0.60	0.60	0.60	0.60	0.56	0.56	0.57	0.56
euclidean	20	distance	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
manhattan	10	uniform	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
manhattan	10	distance	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
manhattan	15	uniform	0.96	0.91	0.93	0.93	0.79	0.53	0.75	0.69	0.80	0.60	0.67	0.69	0.79	0.56	0.67	0.68
manhattan	15	distance	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
manhattan	20	uniform	0.93	0.91	0.91	0.92	0.75	0.53	0.54	0.61	0.67	0.60	0.60	0.62	0.67	0.56	0.57	0.60
manhattan	20	distance	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79
minkowski	10	uniform	0.95	0.96	0.96	0.96	0.74	0.79	0.79	0.77	0.78	0.80	0.80	0.79	0.75	0.79	0.79	0.78
minkowski	10	distance	0.95	0.96	0.96	0.96	0.74	0.79	0.79	0.77	0.78	0.80	0.80	0.79	0.75	0.79	0.79	0.78
minkowski	15	uniform	0.95	0.91	0.93	0.93	0.77	0.53	0.75	0.68	0.73	0.60	0.67	0.67	0.74	0.56	0.67	0.66
minkowski	15	distance	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79

parameter			Accuracy				Precision				Recall				F1 Score			
metric	**n_neighbors**	**weight**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
minkowski	20	uniform	0.91	0.91	0.91	0.91	0.53	0.53	0.54	0.54	0.60	0.60	0.60	0.60	0.56	0.56	0.57	0.56
minkowski	20	distance	0.96	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.79	0.79	0.79

3. Hasil hyperparameter tunning dengan dataset sebesar 80%

Dari hasil yang ditunjukkan pada Tabel 5.3 dapat dilihat bahwa konfigurasi parameter terbaik berdasarkan hasil akurasi dan evaluasi *F1 Score* sebesar 0.96 dan 0.79 serta nilai presisi sebesar 0.79 dan nilai recall sebesar 0.80. Dapat dilihat bahwa parameter weight tidak memiliki pengaruh terhadap performa dari algoritma k-nearest neighbors dengan jumlah data training 80%.

Tabel 5. 3 Hasil Hyperparameter Tunning dengan Data Training sebesar 80%

Parameter			Accuracy				Precision				Recall				F1 Score			
metric	**n_neighbors**	**weight**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
euclidean	10	uniform	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
euclidean	10	distance	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
euclidean	15	uniform	0.90	0.92	0.94	0.92	0.52	0.53	0.76	0.60	0.60	0.60	0.70	0.63	0.55	0.56	0.71	0.61
euclidean	15	distance	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
euclidean	20	uniform	0.90	0.92	0.92	0.91	0.52	0.53	0.56	0.54	0.60	0.60	0.60	0.60	0.55	0.56	0.58	0.56

Parameter			Accuracy				Precision				Recall				F1 Score			
metric	**n_neighbors**	**weight**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
euclidean	20	distance	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
manhattan	10	uniform	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
manhattan	10	distance	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
manhattan	15	uniform	0.92	0.92	0.94	0.92	0.74	0.53	0.76	0.68	0.67	0.60	0.70	0.66	0.66	0.56	0.71	0.65
manhattan	15	distance	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
manhattan	20	uniform	0.90	0.92	0.92	0.91	0.52	0.53	0.56	0.54	0.60	0.60	0.60	0.60	0.55	0.56	0.58	0.56
manhattan	20	distance	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
minkowski	10	uniform	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
minkowski	10	distance	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
minkowski	15	uniform	0.90	0.92	0.94	0.92	0.52	0.53	0.76	0.60	0.60	0.60	0.70	0.63	0.55	0.56	0.71	0.61
minkowski	15	distance	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79
minkowski	20	uniform	0.90	0.92	0.92	0.91	0.52	0.53	0.56	0.54	0.60	0.60	0.60	0.60	0.55	0.56	0.58	0.56
minkowski	20	distance	0.96	0.98	0.96	0.97	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.79	0.80	0.79	0.79

4. Hasil hyperparameter tunning dengan dataset sebesar 90%.

Dari hasil yang ditunjukkan pada Tabel 5.4 dapat dilihat bahwa konfigurasi parameter terbaik berdasarkan hasil akurasi dan evaluasi *F1 Score* sebesar 0.96 dan 0.79 serta nilai presisi sebesar 0.79 dan nilai recall sebesar 0.80. Dapat dilihat bahwa parameter

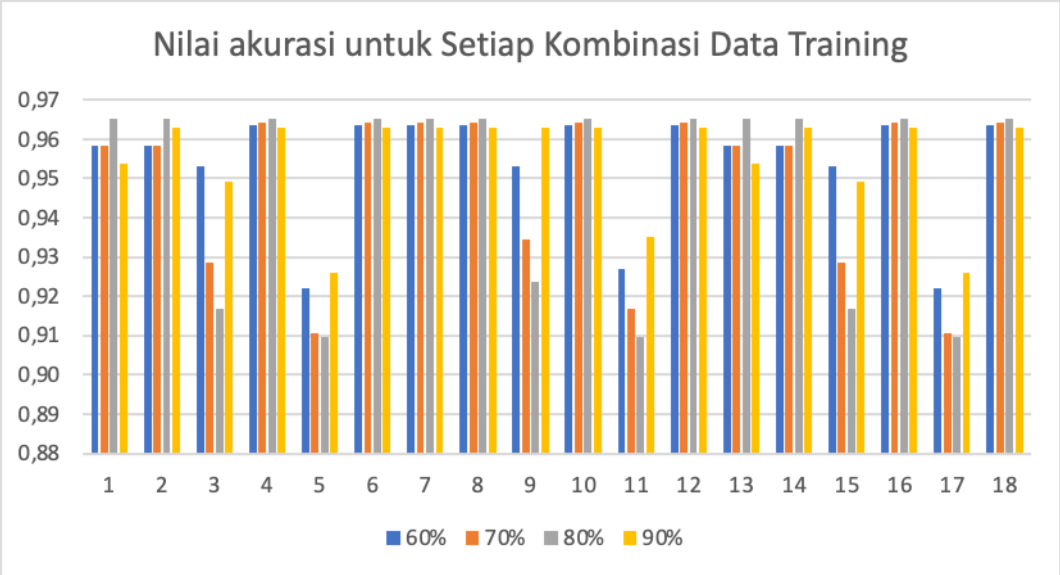
metric, n_neighbors, dan weight tidak memiliki pengaruh terhadap performa dari algoritma k-nearest neighbors dengan jumlah data training 90%.

Tabel 5. 4 Hasil Hyperparameter Tunning Algoritma K-Nearest Neigbors dengan data training sebesar 90%

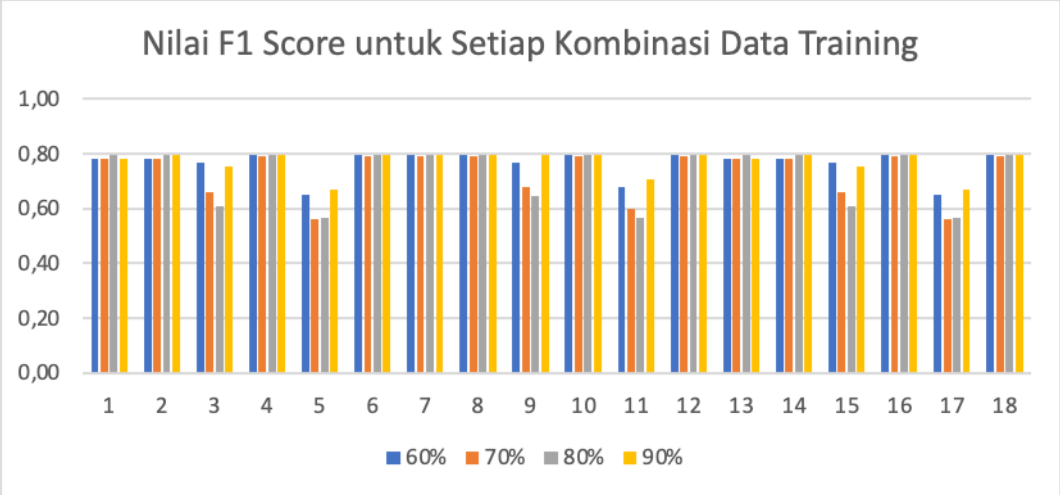
Parameter			Accuracy				Precision				Recall				F1 Score			
metric	**n_neighbors**	**weight**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
euclidean	10	uniform	0.97	0.94	0.94	0.95	0.79	0.78	0.75	0.77	0.80	0.78	0.78	0.79	0.80	0.78	0.76	0.78
euclidean	10	distance	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79
euclidean	15	uniform	0.96	0.96	0.93	0.95	0.77	0.79	0.76	0.77	0.75	0.80	0.70	0.75	0.76	0.79	0.71	0.75
euclidean	15	distance	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79
euclidean	20	uniform	0.93	0.93	0.92	0.93	0.74	0.75	0.75	0.75	0.65	0.70	0.65	0.67	0.65	0.71	0.65	0.67
euclidean	20	distance	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79
manhattan	10	uniform	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79
manhattan	10	distance	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79
manhattan	15	uniform	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79
manhattan	15	distance	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79
manhattan	20	uniform	0.94	0.94	0.92	0.94	0.76	0.77	0.75	0.76	0.70	0.75	0.65	0.70	0.71	0.76	0.65	0.71
manhattan	20	distance	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79
minkowski	10	uniform	0.97	0.94	0.94	0.95	0.79	0.78	0.75	0.77	0.80	0.78	0.78	0.79	0.80	0.78	0.76	0.78
minkowski	10	distance	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79
minkowski	15	uniform	0.96	0.96	0.93	0.95	0.77	0.79	0.76	0.77	0.75	0.80	0.70	0.75	0.76	0.79	0.71	0.75

Parameter			Accuracy				Precision				Recall				F1 Score			
metric	**n_neighbors**	**weight**	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean	split 0	split 1	split 2	mean
minkowski	15	distance	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79
minkowski	20	uniform	0.93	0.93	0.92	0.93	0.74	0.75	0.75	0.75	0.65	0.70	0.65	0.67	0.65	0.71	0.65	0.67
minkowski	20	distance	0.97	0.96	0.96	0.96	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80	0.80	0.79	0.79	0.79

Hasil pengujian hyperparameter tunning pada algoritma K-Nearest Neighbors untuk setiap kombinasi data training berdasarkan nilai akurasi dapat dilihat pada Gambar 5.1 dan berdasarkan nilai F1 Score dapat dilihat pada Gambar 5.2.



Gambar 5. 1 Grafik Nilai Akurasi untuk Setiap Kombinasi Data Training



Gambar 5. 2 Grafik Nilai F1 Score untuk Setiap Kombinasi Data Training

Berdasarkan hasil pengujian hyperparameter tunning dengan kombinasi data training dan testing didapat hasil tebaik pada kombinasi 80% untuk data training

dan 20% untuk data testing. Model terbaik pada proses training selanjutnya digunakan pada proses uji menggunakan data uji yang telah disediakan. Pada Tabel 5.4 ditunjukkan hasil klasifikasi setiap data uji untuk algoritma KNN.

Tabel 5. 4 Hasil Klasifikasi Algoritma K-NN Menggunakan Data Uji

No.	Aktual	Prediksi KNN
1	Admin	Admin
2	Mekanik/Teknisi	Mekanik/Teknisi
3	Designer	Designer
4	Mekanik/Teknisi	Mekanik/Teknisi
5	Designer	Designer
6	Mekanik/Teknisi	Mekanik/Teknisi
7	Mekanik/Teknisi	Mekanik/Teknisi
8	Mekanik/Teknisi	Mekanik/Teknisi
9	Guru	Guru
10	Admin	Admin
11	Admin	Admin
12	Admin	Admin
13	Mekanik/Teknisi	Mekanik/Teknisi
14	Mekanik/Teknisi	Mekanik/Teknisi
15	Mekanik/Teknisi	Mekanik/Teknisi
16	Mekanik/Teknisi	Mekanik/Teknisi
17	Admin	Admin
18	Designer	Designer
19	Admin	Admin
20	Mekanik/Teknisi	Mekanik/Teknisi
21	Mekanik/Teknisi	Mekanik/Teknisi
22	Mekanik/Teknisi	Mekanik/Teknisi
23	Operator Produksi	Mekanik/Teknisi
24	Designer	Designer
25	Mekanik/Teknisi	Mekanik/Teknisi
26	Mekanik/Teknisi	Mekanik/Teknisi
27	Mekanik/Teknisi	Mekanik/Teknisi
28	Designer	Designer
29	Mekanik/Teknisi	Mekanik/Teknisi
30	Mekanik/Teknisi	Mekanik/Teknisi
31	Mekanik/Teknisi	Mekanik/Teknisi
32	Mekanik/Teknisi	Mekanik/Teknisi
33	Operator Produksi	Mekanik/Teknisi

No.	Aktual	Prediksi KNN
34	Mekanik/Teknisi	Mekanik/Teknisi
35	Guru	Guru
36	Mekanik/Teknisi	Mekanik/Teknisi
37	Designer	Designer
38	Mekanik/Teknisi	Mekanik/Teknisi
39	Mekanik/Teknisi	Mekanik/Teknisi
40	Admin	Admin
41	Admin	Admin
42	Mekanik/Teknisi	Mekanik/Teknisi
43	Mekanik/Teknisi	Mekanik/Teknisi
44	Designer	Designer
45	Mekanik/Teknisi	Mekanik/Teknisi
46	Admin	Admin
47	Mekanik/Teknisi	Mekanik/Teknisi
48	Mekanik/Teknisi	Mekanik/Teknisi

Berdasarkan hasil yang ditunjukkan pada Tabel 5.4 didapatkan metrik evaluasi F1 *score, precision, recall* dan *accuracy* algoritma KNN adalah 0.94, 0.92, 0.96, dan 0.96.

5.2 Pengujian Algoritma *K-Nearest Neighbors* Berdasarkan Kelompok Jenis Kelamin

Pengujian algoritma *K-Nearest Neighbors* (KNN) dilakukan untuk mengevaluasi performa model dalam mengklasifikasikan data berdasarkan kelompok jenis kelamin (Laki-laki/L dan Perempuan/P). Hasil evaluasi disajikan pada Tabel 5.5 di bawah ini.

Tabel 5. 5 Hasil Pengujian K-Nearest Neighbors Berdasarkan Jenis Kelamin

Kelamin	Proses	Jumlah Data	Akurasi	Precision	Recall	F1 Skor
L	*Training*	146	0.95	0.73	0.76	0.74
L	*Testing*	37	0.95	0.91	0.88	0.88
P	*Training*	45	0.91	0.72	0.75	0.73
P	*Testing*	12	0.92	0.62	0.67	0.64

Pada klasifikasi jenis kelamin Laki-laki (L), performa model KN N pada data *testing* (37 data) mengalami peningkatan yang signifikan pada F1 Skor menjadi 0.88, dibandingkan dengan F1 Skor data *training* 0.74. Peningkatan ini didukung oleh *Precision* 0.91 dan *Recall* 0.88, yang menunjukkan kemampuan generalisasi yang baik dan stabil untuk kelompok mayoritas ini. Sementara itu, untuk klasifikasi jenis kelamin Perempuan (P), kinerja model pada data *testing* (12 data) menurun menjadi F1 Skor 0.64, padahal data *training* (45 data) memiliki F1 Skor 0.73. Hasil F1 Skor 0.64 ini tercatat identik dengan yang dihasilkan oleh algoritma *Random Forest* pada kelas Perempuan, menegaskan bahwa klasifikasi kelompok minoritas ini merupakan tantangan yang konsisten bagi kedua model, ditandai dengan *Precision* 0.62 dan *Recall* 0.67. Dalam perbandingan kinerja, *Random Forest* (F1) lebih unggul dalam klasifikasi kelompok Laki-laki, namun kedua algoritma menghasilkan F1 Skor yang sama (0.64) pada data *testing* kelompok Perempuan.

5.3 Pengujian Algoritma *K-Nearest Neighbors* Berdasarkan Kelompok Jurusan

Pengujian algoritma KNN untuk mengklasifikasikan data berdasarkan kelompok jurusan menghasilkan data yang disajikan pada Tabel 5.4 di bawah ini.

Tabel 5. 6 Hasil Pengujian K-Nearest Neighbors Berdasarkan Jurusan

Jurusan	Proses	Jumlah Data	Akurasi	Precision	Recall	F1 Skor
Teknik Kendaraan Ringan Otomotif (TKRO)	*Training*	60	0.98	0.73	0.75	0.74
TKRO	*Testing*	15	0.93	0.72	0.75	0.74
Teknik Komputer dan Jaringan (TKJ)	*Training*	51	0.96	0.74	0.75	0.74
TKJ	*Testing*	13	0.92	0.72	0.75	0.74

Multimedia (MM)	*Training*	46	0.91	0.54	0.60	0.57
MM	*Testing*	12	1	1	1	1
Teknik dan Bisnis Sepeda Motor (TBSM)	*Training*	19	0.89	0.61	0.67	0.64
TBSM	*Testing*	5	0.80	0.56	0.67	0.60
Akuntansi dan Keuangan Lembaga (AKL)	*Training*	15	0.80	0.40	0.50	0.45
AKL	*Testing*	4	0.75	0.56	0.67	0.60

Kinerja model KNN pada data *training* menunjukkan hasil yang jauh lebih rendah dan bervariasi dibandingkan *Random Forest*, dengan F1 Skor tertinggi 0.74 (TKJ) dan terendah 0.45 (AKL). Hal ini mengonfirmasi bahwa model KNN tidak mengalami *overfitting* yang parah pada data pelatihan. Meskipun demikian, pada data *testing*, kinerja KNN secara statistik sangat mirip dengan *Random Forest* untuk semua kelas jurusan. Jurusan Multimedia (MM) menjadi yang terbaik dengan F1 Skor sempurna 1. Sementara itu, TKRO dan TKJ menunjukkan F1 Skor yang sama yaitu 0.74. Kelompok data terkecil, TBSM dan AKL, keduanya kembali menunjukkan F1 Skor terendah, yakni 0.60. Kesamaan hasil antara KNN dan *Random Forest* pada data *testing* menunjukkan bahwa terlepas dari perbedaan kinerja di fase *training*, tantangan klasifikasi yang sama muncul pada data yang sebenarnya, terutama pada kelompok data yang minim sampel.

BAB VI

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan hasil penelitian mengenai sistem pemetaan jalur karir siswa lulusan SMK menggunakan algoritma *Random Forest* dan *K-Nearest Neighbors* (KNN), dapat diambil beberapa kesimpulan sebagai berikut:

1. Prediksi jalur karir lulusan SMK dengan algoritma Random Forest dan K-Nearest Neighbors (KNN) diperoleh melalui serangkaian langkah terstruktur, dimulai dari pengumpulan data hasil tes pemetaan jalur karir dari BKK SMK Negeri 1 Wonorejo sebanyak 240 data siswa dengan atribut jurusan dan 8 kompetensi utama, kemudian dilakukan seleksi fitur sehingga hanya jurusan dan delapan kompetensi yang digunakan sebagai variabel independen (X), sedangkan jenis pekerjaan (admin, guru, operator produksi, designer, mekanik/teknisi) dijadikan variabel dependen (Y) sebagai label karir yang diprediksi. Data kategori seperti *jurusan* dan *pekerjaan* diubah menjadi bentuk numerik melalui One Hot Encoding untuk jurusan dan Label Encoding untuk pekerjaan, lalu dataset dibagi menjadi 80% data latih dan 20% data uji sebagai dasar pembentukan dan pengujian model. Pada algoritma Random Forest, model prediksi karir dibangun dengan membuat banyak pohon keputusan dari data latih melalui proses bootstrap sampling dan pemilihan fitur secara acak pada tiap simpul, kemudian setiap siswa pada data uji dikirim ke seluruh pohon dan label pekerjaan diputuskan berdasarkan *majority voting* dari hasil prediksi semua pohon . Sementara itu

pada algoritma K-Nearest Neighbors, prediksi karir siswa dilakukan tanpa membentuk model eksplisit: profil siswa uji (jurusan dan skor 8 kompetensi) dihitung jaraknya terhadap seluruh data latih menggunakan Euclidean Distance, lalu diambil k tetangga terdekat (k = 10) dan jenis pekerjaan bagi siswa tersebut ditentukan berdasarkan voting kelas mayoritas dari tetangga terdekat tersebut. Kinerja kedua algoritma ini kemudian dievaluasi menggunakan confusion matrix dengan metrik akurasi, presisi, recall, dan F1-score, sehingga diperoleh konfigurasi parameter terbaik dan algoritma yang paling andal untuk memberikan prediksi jalur karir yang tepat bagi lulusan SMK berdasarkan jurusan dan profil kompetensinya.

2. Hasil prediksi dari kedua algoritma tersebut dapat diketahui menggunakan evaluasi kinerja model yaitu, *accuracy, precision, recall,* dan f1 *score* . Evaluasi kinerja model terbaik dari masing-masing algoritma menggunakan data uji (20% dari dataset) menunjukkan bahwa *Random Fores*t memperoleh nilai *accuracy* 0.90 yang berarti prediksi model benar, *precision* 0.92 yang berarti dari semua prediksi positif, 92% tepat sasaran, *recall* 0.90 yang berarti model berhasil mendeteksi 90% dari semua data positif yang sebenarnya dan *F1 Score* 0.90 yang berarti keseimbangan antara *precision* dan *recall* cukup baik sedangkan KNN memperoleh nilai *accuracy* 0.96 yang berarti 96% prediksi model benar, *precision* 0.92 yang berarti ketepatannya sama baiknya dengan *Random Forest*, *recall* 0.96 yang berarti lebih tinggi dari Random Forest, artinya KNN lebih mampu mengenali seluruh data positif, dan *F1 Score* 0.92 yang berarti menunjukkan keseimbangan antara *precision* dan *recall* yang lebih baik dari *Random Forest*. Hasil ini menunjukkan bahwa

kedua algoritma memberikan kinerja klasifikasi yang baik, dengan KNN sedikit lebih unggul dibandingkan *Random Forest* pada data uji, terutama pada metrik *recall* dan *F1 Score*. Dengan demikian, algoritma KNN dapat dipertimbangkan sebagai pilihan utama untuk implementasi awal sistem, sementara *Random Forest* dapat menjadi model alternatif yang lebih stabil ketika ukuran data diperbesar.

Selama proses penyusunan dan pelaksanaan penelitian ini, peneliti menghadapi beberapa kendala, antara lain:

1. **Keterbatasan jumlah dan cakupan data**, di mana dataset hanya berasal dari satu sekolah, yaitu SMK Negeri 1 Wonorejo, sehingga variasi karakteristik lulusan dan jalur karir masih terbatas.
2. **Ketidakseimbangan distribusi kelas pekerjaan** pada dataset, di mana beberapa jenis pekerjaan memiliki jumlah data yang lebih dominan dibandingkan kelas lainnya, sehingga berpotensi mempengaruhi performa model klasifikasi.
3. **Keterbatasan waktu penelitian**, terutama dalam melakukan eksplorasi dan pengujian lebih lanjut terhadap algoritma lain atau penerapan teknik optimasi lanjutan seperti *ensemble learning* dan *feature selection* tingkat lanjut.
4. **Keterbatasan akses data longitudinal**, sehingga sistem belum dapat mengakomodasi perubahan kompetensi lulusan atau perkembangan karir alumni dalam jangka panjang.

5. **Proses integrasi sistem dengan BKK secara langsung** belum dilakukan, sehingga sistem masih bersifat prototipe dan belum diimplementasikan secara penuh dalam lingkungan operasional sekolah.

6.2 Saran

Berdasarkan hasil penelitian ini, beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut adalah sebagai berikut:

1. **Perluasan Dataset**

   Dataset yang digunakan dalam penelitian ini masih terbatas. Disarankan untuk memperluas jumlah data siswa dari berbagai jurusan dan tahun kelulusan agar model dapat lebih general dan meningkatkan kemampuan generalisasi sistem.

2. **Penambahan Fitur**

   Fitur input yang digunakan masih terbatas pada atribut tertentu. Penambahan fitur seperti nilai akademik per mata pelajaran kejuruan, hasil asesmen minat bakat, riwayat magang, dan keterampilan tambahan dapat meningkatkan akurasi dan ketepatan rekomendasi jalur karir.

3. **Eksplorasi Algoritma Lain**

   Untuk meningkatkan performa sistem, disarankan melakukan eksperimen dengan algoritma klasifikasi lain seperti *Gradient Boosting, Support Vector Machine (SVM)*, atau *Artificial Neural Networks*, serta membandingkannya dengan *Random Forest* dan KNN

DAFTAR PUSTAKA

Adi Putra, Y., Nugroho, H. W., Trilokai, J., & Ilmu Komputer, F. (2024). *Application of Data Mining Techniques in Assessing the Performance of Vocational High School Students in Computer Engineering at SMK Negeri 1 Braja Selebah Using Support Vector Machines (SVM), Naive Bayes, and k-Nearest Neighbors (k-NN) Algorithms*. *5*(4), 1741–1747.

Agustiningsih, A., Findawati, Y., & Kautsar, I. A. (2023). Classification Of Vocational High School Graduates ' Ability In Industry Using Extreme Gradient Boosting (Xgboost), Random Forest , And Logistic Regression Klasifikasi Kemampuan Lulusan Smk Di Industri Menggunakan Extreme Gradient Boosting (Xgboost). *Jurnal Teknik Informatika (JUTIF)*, *4*(4), 977–985.

Al-Dossari, H., Al-Qahtani, Z., Nughaymish, F. A., Alkahlifah, M., & Alqahtani, A. (2020). CareerRec: A Machine Learning Approach to Career Path Choice for Information Technology Graduates. *Engineering, Technology and Applied Science Research*, *10*, 6589–6596. https://doi.org/10.48084/etasr.3821

Anton Wirawan, F. (2023). *Ketenagakerjaan Dalam Data Edisi 1 Tahun 2023* (Zulfiyandi (ed.); 1st ed.). Pusat Data dan Teknologi Informasi Ketenagakerjaan. https://satudata.kemnaker.go.id

Betrand, C., Aliche, O., Onukwugha, C., Ofoegbu, C., & Kelechi, D. (2025). Career Guidance System Using Decision Tree, Random Forest, and Naïve Bayes Algorithm. *International Journal of Science, Technology and Society*, *13*, 35–42. https://doi.org/10.11648/j.ijsts.20251302.11

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Farhana, S. (2021). Classification of Academic Performance for University Research Evaluation by Implementing Modified Naive Bayes Algorithm. *Procedia Computer Science*, *194*, 224–228. https://doi.org/10.1016/J.PROCS.2021.10.077

Gokarn, S., Taware, S., & Vartak, R. (2024). Smart Career Guidance System using Machine Learning. *International Journal of Science and Research (IJSR)*, *13*(6), 1140–1144. https://doi.org/10.21275/sr24612203820

Haque, R., Goh, H. N., Ting, C. Y., Quek, A., & Hasan, M. D. R. (2025). Leveraging LLMs for optimised feature selection and embedding in structured data: A case study on graduate employment classification. *Computers and Education: Artificial Intelligence*, *8*, 100356. https://doi.org/10.1016/J.CAEAI.2024.100356

Ibnu Katsir, ALU ASY SYAIKH, A. bin M. bin A., & M. Abdul Ghoffar [pnj]. (2011). *Tafsir Ibnu Katsir jilid 9: Juz 27–29 (Al-Mujādilah, Al-Hasyr, Al-Mumtahanah, Ash-Shaff, Al-Jumu'ah, Al-Munafiqun, At-Taghabun, Ath-Thalaq, At-Tahrim, Al-Mulk, Al-Qalam, Al-Haqqah)*. Pustaka Imam Asy-Syafi'i.

Idakwo, J., Joshua Babatunde, A., & Kolajo, T. (2022). Development Of A Hybrid Students' Career Path Recommender System Using Machine Learning Techniques.

In *FUW Trends in Science & Technology Journal, www.ftstjournal.com e-ISSN* (Vol. 7, Issue 3). www.ftstjournal.com

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). An Introduction to Statistical Learning in R. *Springer*, 1–463.

Mahmud Nawawi, H., Baitul Hikmah, A., Mustopa, A., & Wijaya, G. (2024). Model Klasifikasi Machine Learning untuk Prediksi Ketepatan Penempatan Karir. *Jurnal SAINTEKOM*, *14*(1), 13–25. https://doi.org/10.33020/saintekom.v14i1.512

Menteri Tenaga Kerja. (2016). Kemnaker No. 9 Tahun 2016. *Kemnaker*, *4*(2), 1–37.

Misaria Tarigan, Y., Putu, L., & Lestari, S. (2023). Efforts to Determine Career Choices for Vocational Students with the Career Information Service Module. *Bisma The Journal of Counseling*, *7*(1), 138–146. https://doi.org/10.23887/BISMA.V7I1.58909

Pandey, A., & L S, M. (2022). Career Prediction Classifiers based on Academic Performance and Skills using Machine Learning. *International Journal of Computer Science and Engineering*, *9*, 5–20. https://doi.org/10.14445/23488387/ijcse-v9i3p102

Peraturan Pemerintah RI. (2020). Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 50 Tahun 2020 Tentang Praktik Kerja Lapangan Bagi Peserta Didik. *Jurnal Pendidikan*, 2013–2015. https://peraturan.bpk.go.id/Home/Details/163849/permendikbud-no-50-tahun-2020

Purnomo, A., & Sururi, A. (2022). Prediksi Kemampuan Siswa Dalam Bersaing di Dunia Kerja Menggunakan Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor. *ICIT Journal*, *8*, 34–45. https://doi.org/10.33050/icit.v8i1.2171

rachmadani, esi vidia, Pane, syafrial fachri, & Harani, N. H. (2020). Algoritma C4.5 dan K-Nearest Neighbors (KNN) untuk Memetakan Matakuliah. In rolly maulana awangga (Ed.), *kreatif industri nusantara*. kreatif industri nusantara. https://books.google.co.id/books?hl=en&lr=&id=BGv9DwAAQBAJ&oi=fnd&pg=PR10&dq=buku+knn&ots=qogCp1WTfU&sig=4xy196YmJEb_0Dlvm4C_8-0nWCQ&redir_esc=y#v=onepage&q=buku knn&f=false

Peraturan Pemerintah RI. (2024). *Peraturan Menteri Ketenagakerjaan Republik Indonesia Nomor 18 Tahun 2024*. *Table 10*, 4–6.

Sinha, A., . G., & Singh, A. (2023). Student Career Prediction Using Algorithms Of Machine Learning. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4440156

Statistik, B. P., & Indonesia, R. (2022). *Keadaan Angkatan Kerja Di Indonesia Labor Force Situation in Indonesia*. Badan Pusat Statistik/BPS-Statistics Indonesia.

Sutianah, C. (2020). Pengembangan Karakter Wirausaha Siswa Melalui Pembelajaran Kewirausahaan Di Sekolah Menengah Kejuruan. *Jurnal Ekonomi, Sosial & Humaniora*, *2*(05), 96–103. https://jurnalintelektiva.com/index.php/jurnal/article/view/383%0Ahttps://jurnalintelektiva.com/index.php/jurnal/article/download/383/265

Vignesh, S., Shivani Priyanka, C., Shree Manju, H., & Mythili, K. (2021). An Intelligent Career Guidance System using Machine Learning. *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*, 987–990. https://doi.org/10.1109/ICACCS51430.2021.9441978

Wibisono, J., Suharsono, A., & Wijoyo, S. H. (2024). *Perbandingan Kinerja Metode Naive Bayes dan K- Nearest Neighbor untuk Klasifikasi Pekerjaan Berdasarkan Nilai Mata Kuliah (Studi Kasus: Alumni Program Studi Pendidikan Teknologi Informasi Fakultas Ilmu Komputer Universitas Brawijaya)*. Https://J-Ptiik.Ub.Ac.Id/Index.Php/j-Ptiik/Article/View/13589.