

BAB II

LANDASAN TEORI

2.1. Machine Learning

Machine Learning (ML) adalah cabang kecerdasan buatan yang memungkinkan komputer untuk belajar dari data tanpa diprogram secara eksplisit. ML menggunakan algoritma untuk mengidentifikasi pola dalam data, membuat prediksi, dan mengambil keputusan. Ini diterapkan dalam berbagai bidang, seperti analisis data, prediksi perceraian, dan pengenalan pola sosial.

2.1.1. Pengertian Naïve Bayes

Naïve Bayes adalah salah satu algoritma machine learning yang digunakan untuk tugas klasifikasi, yang berdasarkan pada Teorema Bayes dengan asumsi bahwa fitur-fitur dalam data bersifat independen satu sama lain. Meskipun asumsi independensi ini seringkali tidak realistis dalam banyak aplikasi dunia nyata, Naïve Bayes sering kali memberikan hasil yang sangat baik, bahkan pada dataset yang kompleks dan besar. Algoritma ini sangat efisien dan sederhana, serta sering digunakan untuk masalah klasifikasi seperti klasifikasi teks dan analisis sentimen (Maurya, Sharma, & Pandey, 2012).

Teorema Bayes mendasari metode ini, yang menghitung probabilitas suatu kejadian berdasarkan informasi yang sudah diketahui. Dalam konteks perceraian, Naïve Bayes digunakan untuk memprediksi apakah pasangan akan bercerai atau tidak, berdasarkan fitur-fitur seperti usia menikah, pendidikan, jumlah anak, dan

faktor-faktor lain yang relevan. Proses kerja Naïve Bayes melibatkan perhitungan probabilitas posterior untuk setiap kelas (misalnya "bercerai" atau "tidak bercerai"), kemudian memilih kelas dengan probabilitas tertinggi sebagai prediksi (Moumen et al., 2024).

Naïve Bayes memiliki keunggulan dalam hal kecepatan dan efisiensi, karena algoritma ini mampu memproses dan mengklasifikasikan data dengan sangat cepat, bahkan pada dataset yang besar. Algoritma ini juga dikenal karena kesederhanaan dan kemudahan implementasinya, serta sangat cocok untuk menangani data besar yang memiliki banyak fitur (Ahsan, 2023). Namun, algoritma ini memiliki keterbatasan, terutama karena asumsi independensi antar fitur yang seringkali tidak akurat dalam banyak kasus, terutama pada data yang memiliki hubungan kompleks antar fitur. Selain itu, Naïve Bayes lebih cocok digunakan pada data kategorikal dan kurang optimal untuk data numerik dengan variansi tinggi.

2.1.2. Pengertian Support Vector Machine

Support Vector Machine (SVM) adalah algoritma machine learning yang digunakan untuk tugas klasifikasi dan regresi. SVM bekerja dengan cara mencari hyperplane (garis atau bidang pemisah) yang optimal untuk memisahkan dua kelas dalam dataset. Tujuan utama dari SVM adalah untuk menemukan hyperplane yang memaksimalkan margin antara dua kelas, yang disebut dengan margin maksimal. SVM sangat efektif dalam mengklasifikasikan data yang tidak terpisahkan secara linier dengan menggunakan metode yang dikenal sebagai kernel trick.

Dalam konteks perceraian, SVM dapat digunakan untuk memprediksi apakah pasangan akan bercerai atau tidak berdasarkan berbagai faktor yang ada

dalam data, seperti usia menikah, pendidikan, jumlah anak, dan status ekonomi. Dengan menggunakan kernel yang tepat, SVM dapat mengatasi data yang memiliki hubungan non-linear antar variabel, yang sering ditemukan dalam analisis sosial.

Keunggulan utama dari SVM adalah kemampuannya untuk bekerja dengan data berdimensi tinggi, sehingga sangat cocok digunakan untuk kasus perceraian dengan banyak fitur. SVM juga memiliki kemampuan untuk menemukan pemisahan yang lebih baik antara dua kelas (misalnya, pasangan yang bercerai dan yang tidak bercerai) dibandingkan dengan algoritma lain seperti regresi logistik atau decision tree.

Namun, SVM juga memiliki beberapa kekurangan, seperti waktu pelatihan yang lama pada dataset yang sangat besar dan pemilihan kernel yang tepat sangat penting untuk memastikan hasil yang optimal. Meskipun demikian, dengan pengaturan kernel yang tepat, SVM dapat memberikan hasil klasifikasi yang sangat akurat (Cortes & Vapnik, 1995; Moumen et al., 2024; Ahsan, 2023).

2.1.3. Perbandingan Naïve Bayes dan SVM pada Data Sosial

Kedua algoritma memiliki keunggulan dan kelemahan masing-masing dalam konteks analisis perceraian. *Naïve Bayes* unggul dalam hal kecepatan dan efisiensi, serta sangat mudah diinterpretasikan. Algoritma ini juga memiliki kelebihan dalam hal kompleksitas komputasi yang rendah, yang membuatnya ideal untuk dataset besar dengan sumber daya terbatas. Namun, *Naïve Bayes* sensitif terhadap korelasi antar fitur, yang berarti jika fitur-fitur dalam data sangat saling bergantung, performanya bisa menurun (Moumen, A., Shafqat, A., Alraqad, T., Alshawarbeh, E., Saber, H., & Shafqat, R. (2024).

Sebaliknya, *Support Vector Machine (SVM)* lebih unggul dalam akurasi dan kemampuannya untuk menangani data yang kompleks. *Support Vector Machine (SVM)* dapat bekerja dengan data yang lebih beragam dan tidak terpisah secara linear, serta lebih robust terhadap *outlier*. Namun, *Support Vector Machine (SVM)* memerlukan tuning parameter yang lebih banyak, seperti pemilihan kernel (linear, RBF) dan parameter regularisasi (C, gamma), yang bisa meningkatkan komputasi yang lebih tinggi. Dengan demikian, meskipun lebih akurat, *Support Vector Machine (SVM)* membutuhkan waktu komputasi yang lebih lama dan bisa lebih sulit untuk dioptimalkan (Ahsan, M. (2023).

Pada data sosial seperti perceraian, *Support Vector Machine (SVM)* cenderung lebih akurat dalam memprediksi perceraian, terutama ketika data memiliki banyak variabel dan hubungan *non-linear*. Namun, *Naïve Bayes* lebih mudah diimplementasikan dan lebih cepat untuk digunakan ketika kecepatan dan interpretabilitas lebih penting daripada akurasi absolut (Moumen et al., 2024; Yogi et al., 2025).

2.2. Langkah-Langkah Machine Learning

Pipeline machine learning (ML) adalah serangkaian langkah sistematis yang dirancang untuk memastikan bahwa model machine learning berfungsi secara optimal dalam menganalisis data, termasuk prediksi perceraian menggunakan algoritma seperti *Naïve Bayes* dan *Support Vector Machine (SVM)*. Setiap tahapan dalam pipeline ini memainkan peran penting dalam meningkatkan kualitas model dan hasil prediksi yang dihasilkan. Dalam konteks penelitian perceraian,

pengolahan data yang tepat dan pemilihan teknik evaluasi yang baik akan mempengaruhi kinerja model dalam memprediksi kemungkinan perceraian dengan akurasi yang tinggi. Bagian ini akan membahas langkah-langkah utama dalam pipeline machine learning, dengan fokus pada pengaruh pra-pemrosesan data.

2.2.1. Langkah-Langkah Utama Pipeline Machine Learning

1. Pengumpulan Data

Pengumpulan data adalah langkah pertama dalam pipeline machine learning. Data yang dikumpulkan harus relevan dengan tujuan analisis dan berasal dari berbagai sumber, baik yang terstruktur maupun tidak terstruktur. Data yang relevan dengan perceraian bisa mencakup variabel-variabel seperti usia pasangan, status ekonomi, durasi pernikahan, dan faktor psikologis. Pengumpulan data yang komprehensif memungkinkan model untuk mempelajari pola-pola yang lebih kompleks dalam prediksi perceraian (Filippou et al., 2023; Rafliansyah & Anisa, 2024; Suryadevara, M., Rangineni, S., & Venkata, S. (2023).

2. Pra-pemrosesan Data

Pra-pemrosesan data mencakup langkah-langkah pembersihan dan transformasi data agar siap untuk digunakan dalam pelatihan model. Langkah ini mencakup penghapusan duplikasi data, penanganan nilai yang hilang (missing values), normalisasi, standarisasi, encoding kategori, dan seleksi fitur. Pra-pemrosesan yang tepat dapat meningkatkan kualitas data dan mempermudah model dalam melakukan klasifikasi. Misalnya, dalam analisis perceraian, variabel-variabel seperti usia dan durasi pernikahan

mungkin perlu dinormalisasi agar model tidak terpengaruh oleh perbedaan skala data. Pada algoritma *Support Vector Machine (SVM)*, normalisasi data sangat penting karena *Support Vector Machine (SVM)* sangat sensitif terhadap skala data, sedangkan *Naïve Bayes* lebih stabil namun tetap membutuhkan data yang bersih dan terstruktur untuk menghasilkan prediksi yang baik.

3. Pembagian Data (Training & Testing)

Setelah data siap, tahap selanjutnya adalah membaginya menjadi dua set pelatihan (*training set*) dan set pengujian (*testing set*). Pembagian data yang umum digunakan adalah rasio 70:30 atau 80:20, di mana 70% atau 80% data digunakan untuk pelatihan model, dan sisanya untuk pengujian. Pembagian ini penting agar model dapat belajar dari sebagian besar data dan diuji pada data yang tidak dilihat sebelumnya untuk menilai kemampuannya dalam generalisasi (Keelawat, P. (2023). Penting juga untuk melakukan stratifikasi pada pembagian data, terutama jika data tidak seimbang (misalnya lebih banyak pasangan yang tidak bercerai daripada yang bercerai), untuk memastikan bahwa proporsi kelas tetap seimbang di kedua set pelatihan dan pengujian (Gao Xu, Q., Wen, P., Shao, H., He, Y., & Huang, Q, 2023; Radford & Joseph, 2020; Sivakumar, M., Parthasarathy, S., & Padmapriya, T. (2024)

4. Pelatihan Model

Setelah data dibagi, langkah berikutnya adalah melatih model menggunakan algoritma yang dipilih, seperti *Naïve Bayes* atau *Support Vector Machine (SVM)*. Proses pelatihan melibatkan penyusunan model yang akan belajar dari data pelatihan untuk menemukan pola yang relevan. Pada tahap ini, *hyperparameter* model sering kali disesuaikan untuk mengoptimalkan performa. Misalnya, dalam *Support Vector Machine (SVM)*, pemilihan kernel yang tepat (seperti linear atau RBF) sangat penting untuk meningkatkan kemampuan model dalam menangani data non-linear. Sementara itu, pada *Naïve Bayes*, meskipun tidak memerlukan tuning *hyperparameter* yang rumit, pemilihan fitur yang relevan tetap penting untuk meningkatkan hasil (Akpinar, C., Koşar, Ö., & Durdu, A. (2025); Filippou et al., 2023).

5. Evaluasi Model

Setelah model dilatih, tahap selanjutnya adalah evaluasi model menggunakan data pengujian. Berbagai metrik evaluasi digunakan untuk menilai performa model, seperti *akurasi*, *precision*, *recall*, *F1-score*, *AUC*. Metrik ini memberi gambaran tentang seberapa baik model dalam memprediksi perceraian, baik dari segi ketepatan prediksi maupun kemampuan model dalam mendeteksi semua kasus perceraian yang relevan.

6. Iterasi dan Optimasi

Setelah evaluasi, langkah terakhir adalah iterasi dan optimasi. Berdasarkan hasil evaluasi, langkah-langkah tertentu dalam *pipeline* bisa diperbaiki, seperti memperbaiki pra-pemrosesan data, memilih fitur yang lebih baik, atau mengubah parameter model. Proses ini dilakukan berulang kali untuk meningkatkan kinerja model. *Tuning hyperparameter* dan penyesuaian pada seleksi fitur bisa memberikan hasil yang lebih optimal, terutama pada algoritma seperti *Support Vector Machine (SVM)* yang membutuhkan penyesuaian lebih lanjut untuk mencapai performa terbaik (Filippou et al., 2023; Rafliansyah & Anisa, 2024; Suryadevara, M., Rangineni, S., & Venkata, S. (2023).

2.2.2. Pengaruh Pra-pemrosesan Data

Pra-pemrosesan data memainkan peran sangat penting dalam meningkatkan kinerja model machine learning. Pada algoritma *Support Vector Machine (SVM)*, normalisasi dan standarisasi data sangat berpengaruh karena *Support Vector Machine (SVM)* sangat sensitif terhadap perbedaan skala antar fitur. Tanpa normalisasi, *Support Vector Machine (SVM)* mungkin tidak dapat memisahkan kelas dengan baik, yang dapat mengurangi akurasi prediksi. Sementara itu, *Naïve Bayes* lebih stabil terhadap skala data, tetapi tetap mendapat manfaat dari data yang bersih dan terstruktur, termasuk penanganan nilai yang hilang dan encoding fitur kategori dengan benar (Mhatre, M., Pandey, A., Rane, H., & Sahu, S. (2024).

Teknik pra-pemrosesan lainnya, seperti encoding untuk fitur kategorikal (misalnya, status pernikahan, pekerjaan pasangan) atau normalisasi untuk variabel

numerik (misalnya, usia, durasi pernikahan), juga penting untuk meningkatkan kualitas data. Selain itu, dalam kasus pra-pemrosesan teks (jika analisis perceraian melibatkan teks atau data kualitatif), langkah-langkah seperti *case folding*, *stemming*, dan *stopword removal* dapat meningkatkan akurasi model dalam memahami konteks perceraian (Pusean, N., Charibaldi, N., & Santosa, B. (2023).

2.2.3. Praktik Terbaik Pembagian Data

Pembagian data yang baik sangat penting untuk memastikan bahwa model dapat diuji secara adil dan dapat menggeneralisasi ke data yang tidak terlihat sebelumnya. Beberapa praktik terbaik dalam pembagian data antara lain:

1. Rasio umum: Pembagian data dengan rasio 70:30 atau 80:20 sering kali digunakan untuk data pelatihan dan pengujian. Pembagian ini dianggap cukup untuk memberikan model cukup data untuk belajar sambil memastikan bahwa model diuji dengan data yang tidak terlihat (Gao Xu, Q., Wen, P., Shao, H., He, Y., & Huang, Q, 2023; Liu, H., & Haig, E. (2017).
2. Stratifikasi: Dalam kasus data yang tidak seimbang, stratifikasi sangat penting untuk memastikan bahwa kedua set pelatihan dan pengujian memiliki distribusi kelas yang serupa, menghindari bias terhadap kelas mayoritas (Khalid, K., Wijaya, R., & Bijaksana, M. (2025); Liu, H., & Haig, E. (2017).
3. *Cross-validation*: Teknik *cross-validation* digunakan untuk meningkatkan reliabilitas estimasi performa model. Dengan membagi data menjadi beberapa subset dan melatih model pada setiap subset, *cross-validation* membantu dalam mengurangi *overfitting* dan memberikan gambaran lebih

akurat mengenai kemampuan model (Gao Xu, Q., Wen, P., Shao, H., He, Y., & Huang, Q, 2023; Kansab, S. (2025).

2.2.4. Dampak Teknik Persiapan Data pada Naïve Bayes & SVM

Pra-pemrosesan data dapat mempengaruhi model secara signifikan. *Support Vector Machine (SVM)* sangat diuntungkan dari normalisasi dan standarisasi, karena model ini sangat sensitif terhadap skala data, sementara *Naïve Bayes* lebih stabil terhadap skala data, meskipun tetap memerlukan data yang bersih dan terstruktur untuk kinerja yang optimal (Dina, A., Sarno, R., Anggraini, R., Haryono, A., & Septiyanto, A. (2024). . Dengan menerapkan pra-pemrosesan lengkap yang mencakup normalisasi, standarisasi, dan seleksi fitur, kinerja kedua algoritma dapat ditingkatkan secara signifikan (Alam, S., & Yao, N. (2018).

2.3. Perceraian

2.3.1. Definisi Perceraian

Perceraian adalah pemutusan hubungan hukum antara suami dan istri yang sah berdasarkan keputusan pengadilan. Perceraian dapat terjadi karena berbagai alasan, seperti ketidakcocokan, kekerasan dalam rumah tangga, perbedaan nilai hidup, masalah ekonomi, atau perselingkuhan. Setelah perceraian, pasangan tidak lagi terikat oleh kewajiban pernikahan, dan berbagai aspek hukum seperti pembagian harta, hak asuh anak, dan kewajiban nafkah harus diselesaikan. Perceraian berfungsi untuk memberikan kebebasan hukum kepada kedua belah pihak yang tidak lagi dapat melanjutkan hubungan pernikahan mereka.

2.3.2. Faktor-Faktor Mempengaruhi Perceraian

Faktor-faktor yang mempengaruhi perceraian dapat bervariasi antara pasangan, namun menurut berbagai lembaga dan para ahli, beberapa faktor utama yang sering teridentifikasi dalam penelitian perceraian antara lain:

1. Lembaga Pengadilan Agama

Menurut data yang tercatat di Pengadilan Agama, beberapa faktor yang paling banyak diidentifikasi sebagai penyebab perceraian meliputi ketidakstabilan ekonomi, yang sering menjadi penyebab utama dalam perceraian karena pasangan tidak dapat memenuhi kebutuhan dasar hidup mereka. Kekerasan dalam rumah tangga (KDRT), baik fisik maupun verbal, juga menjadi faktor dominan yang mengarah pada perceraian. Selain itu, perselisihan yang terus-menerus antara pasangan yang tidak dapat menyelesaikan masalah dengan cara yang sehat sering kali membuat hubungan rentan dan berisiko berakhir dengan perceraian. Ketidaksetiaan atau perselingkuhan juga merupakan faktor utama yang memicu perceraian, karena tindakan tersebut merusak kepercayaan yang dibangun dalam pernikahan.

2. Lembaga Kementerian Pemberdayaan Perempuan dan Perlindungan Anak (KPPPA)

Menurut Kementerian Pemberdayaan Perempuan dan Perlindungan Anak (KPPPA), faktor sosial dan budaya juga berperan penting dalam perceraian. Lembaga ini mencatat bahwa norma sosial sering menekan pasangan untuk mempertahankan pernikahan meskipun hal itu dapat mengabaikan

kesehatan emosional atau fisik mereka. Selain itu, peran gender yang tidak seimbang dalam keluarga, di mana wanita sering merasa terbebani dengan beban domestik dan kurangnya dukungan dari pasangan, juga berkontribusi pada meningkatnya kemungkinan perceraian.

3. Lembaga Penelitian Sosial

Dalam studi yang dilakukan oleh Badan Penelitian dan Pengembangan Sosial (BPS), beberapa faktor yang sering dikaitkan dengan perceraian antara lain tingkat pendidikan dan usia pernikahan. Pasangan dengan tingkat pendidikan yang lebih rendah cenderung memiliki kemungkinan perceraian yang lebih tinggi, karena perbedaan cara berpikir dan harapan hidup yang tidak selaras antara mereka. Selain itu, pasangan yang menikah di usia muda sering lebih rentan terhadap perceraian, karena mereka belum cukup matang secara emosional untuk menghadapi berbagai tantangan dalam pernikahan.

4. Lembaga Hukum dan Perundang-Undangan

Lembaga peradilan dan pemerintah juga menyatakan bahwa faktor hukum seringkali menjadi pemicu perceraian. Hal ini terutama berkaitan dengan kemudahan dalam proses hukum perceraian, di mana adanya prosedur yang relatif mudah untuk mengajukan perceraian di banyak negara atau wilayah memungkinkan pasangan yang menghadapi kesulitan dalam pernikahan memilih untuk mengakhiri hubungan mereka. Selain itu, ketidakpastian hukum mengenai hak asuh anak juga sering menjadi salah satu alasan perceraian, terutama ketika hukum tidak memberikan kejelasan atau

keadilan yang memadai terkait pembagian hak asuh anak setelah perpisahan.

2.3.3. Peran Machine Learning dalam Memahami Pola Perceraian

Machine learning menawarkan pendekatan baru dalam memprediksi perceraian dengan menganalisis data yang tersedia. Salah satu keuntungan utama dari penggunaan *machine learning* dalam konteks ini adalah kemampuan untuk mendeteksi pola *non-linear* dalam data perceraian. Algoritma seperti *Random Forest*, *SVM*, dan *Neural Networks* memiliki kemampuan untuk menangkap hubungan yang kompleks antara berbagai faktor penyebab perceraian yang mungkin tidak terdeteksi oleh metode statistik tradisional. Hal ini memungkinkan model untuk menghasilkan prediksi yang lebih akurat dan lebih dalam terkait faktor-faktor yang mempengaruhi perceraian (Arpino, B., Moglie, M., & Mencarini, L. (2021).

Dalam pengembangan model machine learning, seleksi fitur menjadi hal yang sangat penting. Teknik ini memungkinkan model untuk memilih atribut yang paling relevan dalam memprediksi perceraian, seperti durasi pernikahan, status ekonomi, atau faktor psikologis. Dengan memilih fitur yang paling signifikan, model dapat lebih efisien dan menghasilkan prediksi yang lebih baik. Teknik seleksi fitur berbasis korelasi atau information gain sering digunakan untuk tujuan ini, sehingga algoritma machine learning dapat memfokuskan pada variabel yang paling mempengaruhi (Kalra et al., 2022; Moumen et al., 2024).

Akurasi tinggi juga menjadi salah satu keuntungan utama penggunaan machine learning dalam prediksi perceraian. Algoritma seperti *Support Vector*

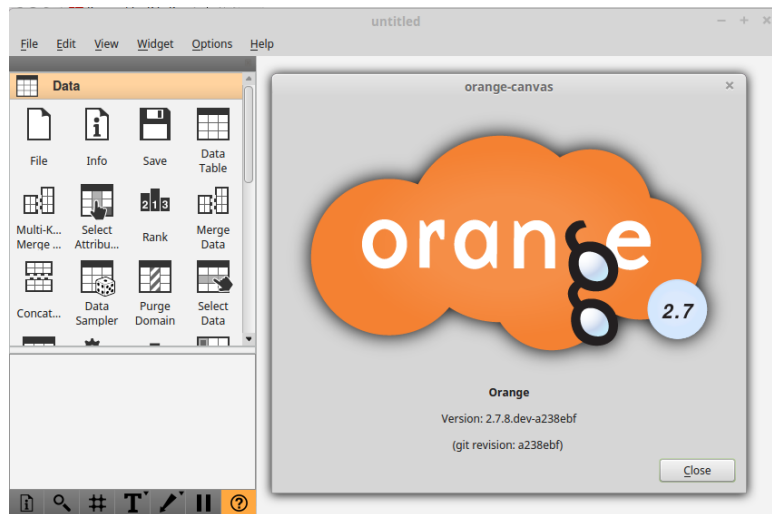
Machine (SVM), Perceptron, dan Random Forest telah terbukti menunjukkan tingkat akurasi prediksi yang sangat tinggi, dengan beberapa studi mencapai akurasi hingga 98-100%. Hal ini menunjukkan bahwa machine learning dapat menjadi alat yang sangat efektif dalam menganalisis data perceraian, memberikan wawasan yang lebih tajam dan presisi dalam prediksi, serta memungkinkan para praktisi dan pembuat kebijakan untuk merancang kebijakan yang lebih tepat sasaran dalam menangani perceraian (Ahsan, 2023; Kalra et al., 2022; Maurya et al., 2012).

2.4. Alat Bantu Pemrograman dan Tools Pendukung

Dalam penelitian machine learning (ML), pemilihan alat yang tepat sangat penting untuk memastikan bahwa model dapat diimplementasikan dengan efektif dan efisien. *Orange*, sebagai salah satu *platform open-source* untuk analisis data dan *machine learning*, menawarkan sejumlah keunggulan yang membuatnya sangat populer, terutama bagi pengguna yang tidak memiliki latar belakang pemrograman. *Orange* menyediakan antarmuka visual yang memungkinkan pengguna untuk mengembangkan dan menguji model machine learning tanpa menulis kode. Pada bagian ini, kami akan membahas keunggulan utama dari *Orange*, kemudahan penggunaannya, serta perbandingannya dengan alat-alat lain seperti *Scikit-learn* dan *Jupyter Notebooks* dalam konteks prediksi perceraian.

2.4.1. Keunggulan Orange untuk Machine Learning dan Prediksi Perceraian

Salah satu keunggulan utama *Orange* adalah antarmuka visual *drag-and-drop* yang memudahkan pengguna untuk membangun *pipeline machine learning* tanpa menulis kode.



Gambar 2.1 Platform Orange

Pengguna cukup menghubungkan berbagai *widget* untuk mengatur alur kerja dari pengumpulan data hingga evaluasi model. Fitur ini sangat bermanfaat bagi pengguna *non-programmer* yang ingin melakukan analisis *machine learning*, seperti dalam prediksi perceraian, tanpa harus memiliki pengetahuan mendalam tentang pemrograman (Oswald, C., Baranwal, S., Narayanan, S., & Bhattacharya, A. (2022).

Dengan kemudahan ini, *Orange* memungkinkan peneliti untuk lebih fokus pada analisis dan interpretasi hasil model daripada menghabiskan waktu untuk menulis kode (Thakur, E. (2023); Yu, N., Shin, D., Ryu, I., Yoo, T., & Koh, K. (2025).

Orange juga menyediakan fitur lengkap yang mencakup berbagai algoritma *machine learning*, seperti *Naïve Bayes*, *Support Vector Machine (SVM)*, *Decision Tree*, *Neural Network*, dan *Random Forest*, yang dapat langsung digunakan untuk membangun model prediksi perceraian. Selain itu, *Orange* memiliki alat untuk pra-pemrosesan data, seleksi fitur, evaluasi model, dan visualisasi hasil. Ini

memungkinkan pengguna untuk tidak hanya membangun model, tetapi juga menguji dan memahami hasil prediksi melalui berbagai metrik seperti *confusion matrix*, *ROC curve*, dan *precision-recall curve*. Fitur visualisasi ini sangat penting dalam konteks prediksi perceraian karena memberikan pemahaman yang lebih mendalam tentang bagaimana model bekerja dan di mana model mungkin perlu perbaikan (Popchev & Orozova, 2023; Thakur, 2023; Yu et al., 2025).

Keunggulan lain dari Orange adalah kemudahan replikasi dan dokumentasi. *Workflow* yang dibangun di *Orange* sangat mudah untuk direplikasi dan didokumentasikan, yang memungkinkan peneliti untuk mengulang eksperimen atau berbagi hasil dengan orang lain. Ini adalah keuntungan besar dalam penelitian yang melibatkan data yang kompleks dan sensitif seperti perceraian, di mana penting untuk memastikan bahwa hasil dapat diuji ulang dan dipertanggungjawabkan (Yu, N., Shin, D., Ryu, I., Yoo, T., & Koh, K. (2025)). Selain itu, *Orange* bersifat gratis dan *open-source*, yang berarti dapat diakses oleh siapa saja, termasuk peneliti dari latar belakang non-teknis atau organisasi dengan anggaran terbatas. Kemampuan ini membuat *Orange* menjadi pilihan yang sangat baik untuk penelitian di berbagai bidang, termasuk perceraian, kesehatan, dan pendidikan, di mana aksesibilitas menjadi salah satu faktor penting (Thakur, E. (2023); Yu, N., Shin, D., Ryu, I., Yoo, T., & Koh, K. (2025)).

2.4.2. Kemudahan Implementasi untuk Pengguna Non-Programmer

Salah satu alasan utama *Orange* sangat disukai oleh pengguna non-programmer adalah karena tanpa kode. Pengguna dapat melakukan hampir seluruh proses machine learning, mulai dari import data, pra-pemrosesan, pelatihan model,

hingga evaluasi hasil, tanpa perlu menulis satu baris kode pun. Ini memungkinkan mereka yang tidak memiliki keterampilan pemrograman untuk tetap menggunakan teknik machine learning yang canggih untuk analisis perceraian. Dengan demikian, Orange menyediakan alat yang sangat berguna bagi peneliti atau praktisi yang fokus pada pengolahan dan analisis data, namun tidak ingin terhambat oleh keterbatasan teknis dalam menulis kode (Thakur, E. (2023); Yu, N., Shin, D., Ryu, I., Yoo, T., & Koh, K. (2025)).

Widget interaktif adalah fitur kunci lain dari *Orange*. Pengguna dapat menghubungkan *widget* seperti *Data*, *Preprocess*, *Test & Score*, dan *Confusion Matrix* untuk membangun pipeline analisis data. Pengguna hanya perlu menarik dan menghubungkan widget sesuai dengan alur kerja yang diinginkan untuk membangun model machine learning yang berfungsi dengan baik. Keuntungan dari penggunaan *widget* ini adalah bahwa pengguna dapat melihat secara langsung perubahan yang terjadi pada model saat mereka mengubah parameter atau menambah elemen baru pada *workflow*, yang memudahkan pemahaman tentang cara kerja algoritma (Popchev, I., & Orozova, D. (2023); Thakur, E. (2023)).

Selain itu, parameter tuning di *Orange* sangat mudah dilakukan, karena semua pengaturan dilakukan melalui jendela pengaturan *grafis*, bukan melalui *script*. Hal ini memungkinkan pengguna untuk melakukan penyesuaian dengan mudah tanpa perlu menulis kode yang rumit, membuat *Orange* lebih mudah diakses dan digunakan oleh siapa saja (Yu, N., Shin, D., Ryu, I., Yoo, T., & Koh, K. (2025)). Visualisasi juga menjadi salah satu fitur kuat dari *Orange*. Hasil evaluasi model, seperti *confusion matrix*, *ROC curve*, dan *metrik* lainnya, dapat divisualisasikan

secara langsung. Visualisasi ini membantu peneliti untuk lebih mudah memahami kinerja model dan mengidentifikasi area yang perlu diperbaiki, yang sangat penting dalam konteks perceraian, di mana hasil model dapat memiliki dampak sosial yang signifikan (Popchev, I., & Orozova, D. (2023); Thakur, E. (2023)).

2.4.3. Perbandingan Orange dengan Scikit-learn dan Jupyter Notebooks

Ketika dibandingkan dengan *Scikit-learn* dan *Jupyter Notebooks*, *Orange* memiliki beberapa keunggulan dan kelemahan. *Scikit-learn* adalah salah satu pustaka *machine learning Python* yang paling fleksibel dan banyak digunakan, tetapi membutuhkan keterampilan pemrograman *Python*. Pengguna harus menulis kode untuk membangun model, mengonfigurasi algoritma, dan melakukan evaluasi. Meskipun sangat fleksibel, *Scikit-learn* bisa jadi lebih sulit digunakan oleh pemula yang tidak memiliki latar belakang pemrograman.

Jupyter Notebooks memungkinkan pengguna untuk menulis kode *Python* secara interaktif dan melihat hasilnya secara langsung. Ini sangat berguna untuk dokumentasi dan kolaborasi, namun juga membutuhkan keterampilan pemrograman dan kontrol manual yang lebih besar. *Jupyter* cocok untuk pengguna yang lebih berpengalaman dalam menulis kode dan membutuhkan fleksibilitas lebih dalam pemrograman.

Di sisi lain, *Orange* memberikan keuntungan besar bagi pengguna yang ingin menghindari pemrograman dan tetap menggunakan algoritma *machine learning* yang kuat. Dengan antarmuka grafis yang mudah digunakan, *Orange* memungkinkan peneliti untuk membangun dan menguji model *machine learning* tanpa menulis kode. Meskipun kurang fleksibel dibandingkan *Scikit-learn* dalam

hal kustomisasi algoritma lanjutan, *Orange* sangat ideal untuk penelitian prediksi perceraian, di mana kemudahan penggunaan dan aksesibilitas lebih diutamakan daripada fleksibilitas penuh dalam pengaturan model (Popchev, I., & Orozova, D. (2023)..; Thakur, E. (2023).

2.4.4. Alternatif Tools untuk Prediksi Perceraian

Selain *Orange*, ada beberapa platform lain yang juga dapat digunakan untuk prediksi perceraian:

1. *Google Vertex AI*: Platform berbasis cloud ini memungkinkan pengguna untuk melakukan machine learning tanpa menulis kode, mirip dengan *Orange*. Namun, *Orange* lebih unggul dalam fleksibilitas dan *open-source*, yang memungkinkan kustomisasi lebih besar dalam penelitian (Yu, N., Shin, D., Ryu, I., Yoo, T., & Koh, K. (2025).
2. *KNIME*: Platform visual yang mirip dengan *Orange*, *KNIME* menawarkan banyak fitur untuk data mining dan machine learning. Meskipun memiliki keunggulan serupa dengan *Orange*, *Orange* lebih populer di kalangan peneliti dan pendidikan karena lebih banyak digunakan dalam konteks penelitian (Thakur, E. (2023).
3. *Ktrain: Library Python* yang menawarkan pendekatan *low-code*, cocok untuk pengguna dengan sedikit pengalaman dalam coding (Maiya, A. (2020).
4. *Google Colab*: Platform berbasis *cloud* yang memungkinkan kolaborasi dan eksekusi kode *Python*, meskipun tetap membutuhkan kemampuan pemrograman (Cagnie et al., 2023).

2.5. Teknik Evaluasi Algoritma di Machine Learning

Evaluasi performa algoritma *machine learning (ML)* adalah tahap krusial dalam penilaian efektivitas model prediksi perceraian. Menggunakan berbagai metrik evaluasi, kita dapat mendapatkan gambaran menyeluruh tentang kinerja algoritma yang diterapkan pada data perceraian. Setiap metrik memberikan informasi yang berbeda mengenai kualitas model, baik dalam hal akurasi prediksi maupun kemampuan model dalam menangani berbagai jenis kesalahan. Pada bagian ini, kami akan membahas metrik utama yang digunakan dalam evaluasi klasifikasi dan menghubungkannya dengan konteks analisis perceraian.

2.5.1. Metrik Utama Evaluasi Klasifikasi

Berbagai metrik digunakan untuk mengevaluasi kinerja algoritma klasifikasi, termasuk *akurasi*, *precision*, *recall*, *AUC* dan *F1-score*. Setiap metrik memiliki tujuan dan relevansi tertentu, tergantung pada karakteristik data dan tujuan aplikasi. Akurasi adalah metrik evaluasi yang paling umum digunakan, yang mengukur persentase prediksi yang benar dari seluruh data. Metrik ini memberikan gambaran umum tentang seberapa baik model dalam mengklasifikasikan data secara keseluruhan (De Diego, I., Redondo, A., Fernández, R., Navarro, J., & Moguerza, J. (2022)). Namun, dalam konteks data yang tidak seimbang, di mana salah satu kelas (misalnya perceraian) jauh lebih sedikit daripada kelas lainnya, akurasi bisa memberikan gambaran yang menyesatkan karena model dapat memperoleh akurasi tinggi hanya dengan memprediksi kelas mayoritas.

Precision mengukur ketepatan model dalam memprediksi kelas positif (misalnya, pasangan yang bercerai). *Precision* mengacu pada proporsi prediksi positif yang benar-benar positif. Metrik ini sangat penting ketika *false positives* (prediksi salah untuk perceraian) memiliki konsekuensi besar, seperti memberikan saran perceraian pada pasangan yang sebenarnya tidak akan bercerai (Fareed et al., 2022).

Recall atau sensitivitas mengukur kelengkapan model dalam mendeteksi kasus positif (perceraian). Metrik ini penting jika *false negatives* (gagal mendeteksi perceraian) berisiko besar, misalnya jika intervensi dini untuk pasangan yang sedang dalam proses perceraian gagal dilakukan. Dalam konteks perceraian, *recall* membantu memastikan bahwa sebanyak mungkin kasus perceraian dapat terdeteksi, bahkan jika itu mengarah pada sejumlah kesalahan prediksi.

F1-score adalah harmonic mean antara *precision* dan *recall*. Metrik ini sangat berguna ketika data tidak seimbang dan memberikan gambaran yang lebih seimbang antara ketepatan dan kelengkapan model. *F1-score* penting karena kadang-kadang ada *trade-off* antara *precision* dan *recall*; ketika kita mencoba meningkatkan salah satu, yang lain cenderung menurun. Oleh karena itu, *F1-score* adalah metrik yang sangat relevan dalam konteks perceraian, di mana penting untuk mempertahankan keseimbangan antara *false positives* dan *false negatives* (Fareed et al., 2022).

Area Under Curve (AUC) merupakan salah satu metrik evaluasi yang digunakan untuk mengukur kemampuan model klasifikasi dalam membedakan dua kelas yang berbeda. Nilai AUC diperoleh dari kurva ROC (Receiver Operating

Characteristic) yang menggambarkan hubungan antara True Positive Rate (TPR) dan False Positive Rate (FPR) pada berbagai nilai threshold. Nilai AUC berada pada rentang 0 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan bahwa model memiliki kemampuan klasifikasi yang semakin baik dalam membedakan kelas positif dan negatif, sedangkan nilai yang mendekati 0,5 menunjukkan kemampuan klasifikasi yang rendah atau mendekati prediksi acak. Dalam penelitian ini, AUC digunakan untuk mengevaluasi kemampuan algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam membedakan pasangan yang berpotensi bercerai dan yang tidak bercerai

2.5.2. Peran F1-score dalam Evaluasi Prediksi Perceraian

F1-score sangat penting dalam evaluasi prediksi perceraian karena ia menggabungkan precision dan recall, memberikan gambaran yang lebih seimbang mengenai kinerja model. Dalam prediksi perceraian, di mana kita sering berhadapan dengan data yang tidak seimbang (misalnya lebih banyak pasangan yang tidak bercerai daripada yang bercerai), *F1-score* memastikan bahwa model tidak hanya mengoptimalkan akurasi dalam memprediksi pasangan yang tidak bercerai tetapi juga tidak mengabaikan pasangan yang berisiko bercerai (Flores, N., & Silva, S. (2021). Oleh karena itu, *F1-score* memberikan hasil evaluasi yang lebih representatif untuk situasi nyata di mana kita tidak hanya ingin meminimalkan kesalahan prediksi pada kelas mayoritas (pasangan yang tidak bercerai), tetapi juga mengidentifikasi dengan benar pasangan yang berisiko bercerai (Fareed et al., 2022).

2.5.3. Trade-off Precision dan Recall pada Klasifikasi Perceraian

Terdapat *trade-off* antara *precision* dan *recall* yang perlu dipertimbangkan dalam konteks prediksi perceraian. *Precision* tinggi sangat penting dalam situasi di mana *false positives* (prediksi perceraian yang salah) dapat berdampak besar. Misalnya, jika konselor atau pihak yang berwenang memberikan rekomendasi perceraian yang tidak diperlukan, hal ini bisa menyebabkan dampak negatif yang signifikan bagi pasangan tersebut. Oleh karena itu, model dengan *precision* tinggi lebih diutamakan dalam situasi seperti ini.

Di sisi lain, *recall* tinggi lebih penting jika *false negatives* (gagal mendeteksi perceraian) berisiko besar. Dalam konteks perceraian, jika pasangan yang berisiko bercerai tidak terdeteksi, mereka mungkin tidak menerima intervensi atau dukungan yang dibutuhkan untuk mencegah perceraian. Oleh karena itu, *recall* yang tinggi sangat diinginkan dalam kasus di mana deteksi lebih banyak kasus perceraian lebih penting daripada menghindari kesalahan dalam prediksi Rainio, O., Teuho, J., & Klén, R. (2024); Fareed et al., 2022).

Penting untuk menyesuaikan *trade-off* antara *precision* dan *recall* dengan tujuan aplikasi. Jika tujuan model adalah untuk memberikan rekomendasi perceraian kepada konselor atau pihak berwenang, maka *precision* mungkin lebih diutamakan. Namun, jika tujuan model adalah untuk deteksi dini perceraian dan intervensi yang lebih baik, maka *recall* menjadi lebih penting Rainio, O., Teuho, J., & Klén, R. (2024); Fareed et al., 2022).

2.6. Penelitian Terdahulu dan Kelebihan Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

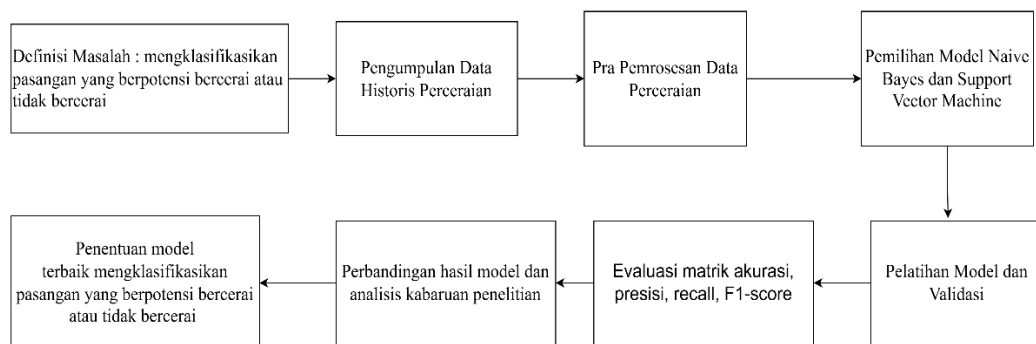
| Peneliti | Judul | Tahun | Data dan Metode | Hasil |
|----------------------------|---|--------------|---|--|
| Noghani et al. (2021) | <i>A Systematic Review of Quantitative and Qualitative Research on Divorce Factors</i> | 2021 | Tinjauan sistematis 70 studi kuantitatif dan kualitatif tentang faktor-faktor perceraian. | Menemukan kelompok faktor dominan: individu/psikologis, kekerasan, ekonomi, sosial, dan kultural. |
| Phiri et al. (2024) | <i>Predicting Social Phenomena: A Comparison of Naïve Bayes and SVM in Predicting Divorce</i> | 2024 | Penggunaan data perceraian dan analisis dengan algoritma Naïve Bayes dan SVM. | Faktor dominan yang ditemukan adalah kekerasan fisik dan verbal, ketidakcocokan pasangan, ketidakstabilan finansial, dan faktor sosial. |
| Ahsan (2023) | Divorce Prediction with Machine Learning: Insights and LIME Interpretability | 2023 | Dataset “divorce predictor,” menggunakan algoritma Naïve Bayes, SVM, serta teknik interpretasi fitur dengan LIME. | Akurasi tinggi pada Naïve Bayes dan SVM (hingga 98,57%), menunjukkan kedua algoritma memiliki performa yang baik dalam memprediksi perceraian. |
| Rahmadini & Santoso (2025) | <i>Machine Learning-Based Prediction of Divorce Verdicts Using Posita Data and Imbalanced Data Handling</i> | 2025 | Data posita dari 2.026 perkara perceraian di Pengadilan Agama Padang Sidempuan, dengan ekstraksi kata | Faktor dominan yang ditemukan adalah kekerasan domestik, mediasi, masalah ekonomi, dan peran keluarga dalam perceraian. |

| | | | | |
|----------------------|--|------|--|---|
| | | | kunci dan SMOTE. | |
| Moumen et al. (2024) | <i>Divorce prediction using machine learning algorithms in Ha'il region, KSA</i> | 2024 | Divorce Predictor Scale (DPS) berbasis teori Gottman dengan 54 item sebagai fitur. | Faktor dominan yang mempengaruhi perceraian adalah psikologi individu, kekerasan dalam rumah tangga, dan ketidakcocokan pasangan. |

Penelitian ini menawarkan kebaruan (novelty) yang signifikan dengan pendekatan yang lebih efisien dan replikatif dalam prediksi perceraian menggunakan algoritma machine learning. Salah satu keunggulan utama penelitian ini adalah penggunaan tools visual seperti Orange, yang memungkinkan pembangunan model tanpa menulis kode, membuatnya lebih mudah diakses oleh peneliti non-programmer dan mempermudah replikasi hasil. Dibandingkan dengan penelitian terdahulu yang menggunakan algoritma seperti *Support Vector Machine (SVM)* dan *Naïve Bayes*, penelitian ini juga mengintegrasikan LIME untuk meningkatkan interpretabilitas model, yang masih jarang dilakukan dalam penelitian perceraian sebelumnya. Selain itu, penelitian ini mengadopsi teknik oversampling yang lebih canggih untuk menangani data tidak seimbang, sebuah tantangan yang dihadapi dalam sebagian besar penelitian sebelumnya. Penelitian ini juga menguji model pada berbagai dataset dari berbagai negara, yang memperluas generalisasi hasil model, menjadikannya lebih robust dibandingkan penelitian terdahulu yang terbatas pada satu budaya atau negara.

2.7. Flowchart Kerangka Penelitian

Flowchart memiliki peran yang sangat penting dalam visualisasi alur penelitian machine learning. Sebagai alat bantu yang mengubah tahapan-tahapan penelitian yang kompleks menjadi langkah-langkah yang lebih mudah dipahami, flowchart dapat memberikan gambaran yang jelas mengenai seluruh proses, dari awal hingga akhir. Hal ini memungkinkan pembaca, tim peneliti, dan pemangku kepentingan untuk dengan mudah mengikuti alur kerja penelitian, meminimalkan kebingungannya, serta memastikan setiap tahap dilakukan secara sistematis dan efisien.



Gambar 2. 2 *Flowchart Kerangka Penelitian*

Flowchart penelitian ini menggambarkan tahapan-tahapan dalam penelitian prediksi perceraian menggunakan machine learning. Tahapan pertama adalah Definisi Masalah, yaitu mengklasifikasikan pasangan yang berpotensi bercerai atau tidak. Selanjutnya, dilakukan Pengumpulan Data Histori Perceraian untuk memperoleh data yang relevan. Setelah itu, dilakukan Pra Pemrosesan Data Perceraian untuk membersihkan dan menyiapkan data. Pemilihan Model dilakukan dengan memilih antara *Naïve Bayes* dan *Support Vector Machine (SVM)*. Kemudian, dilakukan Pelatihan Model dan Evaluasi menggunakan matrik akurasi,

presisi, recall, dan F1-score. Akhirnya, dilakukan Perbandingan Hasil Model dan Analisis Kebaruan Penelitian.