

KLASIFIKASI KEANDALAN EKSPORTIR MENGUNAKAN ALGORITMA RANDOM FOREST

Azizah ramadayanti^{1,*}, Chi Chi kirana², Reza sintia wati³ Yusnita septiani⁴

¹ Universitas Lembah Dempo

² Universitas Lembah Dempo

³ Universitas Lembah Dempo

⁴ Universitas Lembah Dempo

Email : azizahsainsdata@gmail.com, rezasinthiawaa@gmail.com, chichikirana553@gmail.com, septianiyusnita098@gmail.com

INFORMASI ARTIKEL

ABSTRAK

Riwayat Artikel

Diterima 30 Januari 2026
Direvisi 15 Februari 2026
Diterbitkan 18 Februari
2026

Kata Kunci

Kata Kunci
Random Forest
Risiko Ekspor
CRISP-DM
Incoterms
Kapasitas Logistik

Penilaian tingkat keandalan eksportir merupakan aspek penting dalam mendukung pengambilan keputusan dan pengelolaan risiko transaksi ekspor. Namun, proses penilaian yang masih dilakukan secara manual berpotensi menghasilkan keputusan yang subjektif dan kurang konsisten. Penelitian ini bertujuan mengklasifikasikan tingkat risiko transaksi eksportir ke dalam kategori *High Reliability* dan *Low Risk* menggunakan algoritma Random Forest. Metode penelitian menerapkan kerangka kerja CRISP-DM yang meliputi tahapan *business understanding*, *data understanding*, *data preparation*, *modeling*, dan *evaluation*. Dataset yang digunakan berasal dari data transaksi eksportir Indonesia yang diintegrasikan dengan data Harmonized System (HS Code). Hasil pengujian menunjukkan bahwa model Random Forest memperoleh nilai *accuracy* sebesar **87,30%**, *precision* sebesar **86,10%**, *recall* sebesar **88,20%**, dan *F1-score* sebesar **87,10%**. Selain itu, kapasitas produksi bulanan dan nilai barang per kilogram teridentifikasi sebagai atribut yang paling berpengaruh dalam menentukan tingkat keandalan eksportir. Hasil penelitian menunjukkan bahwa Random Forest efektif digunakan untuk mendukung sistem penilaian eksportir berbasis data.

1. Pendahuluan

Kegiatan ekspor merupakan salah satu komponen penting dalam pertumbuhan ekonomi suatu negara karena berkontribusi terhadap peningkatan devisa, perluasan pasar, serta penguatan daya saing nasional. Di Indonesia, sektor ekspor menjadi salah satu fokus utama pemerintah dalam mendorong pertumbuhan ekonomi yang berkelanjutan melalui berbagai kebijakan perdagangan internasional dan fasilitasi ekspor. Meskipun demikian, keberhasilan aktivitas ekspor tidak hanya ditentukan oleh besarnya volume perdagangan, tetapi juga oleh tingkat keandalan eksportir dalam memenuhi kewajiban kontrak, menjaga kualitas produk, serta memastikan kelancaran proses logistik dan pembayaran (World Trade Organization, 2023).

Dalam praktik perdagangan internasional, penilaian terhadap tingkat risiko dan keandalan eksportir menjadi aspek yang sangat penting bagi pemerintah, lembaga pembiayaan, maupun mitra dagang internasional. Kesalahan dalam menilai tingkat risiko eksportir dapat menimbulkan berbagai dampak negatif, seperti keterlambatan pengiriman, kegagalan pemenuhan kontrak, kerugian finansial, hingga menurunnya kepercayaan mitra bisnis (Hosseini et al., 2022). Oleh karena itu, diperlukan suatu mekanisme penilaian yang objektif, konsisten, dan berbasis data untuk mendukung proses pengambilan keputusan.

Salah satu indikator yang dapat digunakan dalam menilai keandalan eksportir adalah Incoterms (International Commercial Terms), yaitu seperangkat aturan internasional yang mengatur pembagian tanggung jawab, biaya, dan risiko antara penjual dan pembeli dalam transaksi perdagangan internasional. Eksportir yang menggunakan ketentuan Cost, Insurance, and Freight (CIF) umumnya memiliki tanggung jawab yang lebih besar dibandingkan eksportir yang menggunakan Free On Board (FOB), sehingga mencerminkan kemampuan logistik dan manajemen risiko yang lebih baik. Selain Incoterms, faktor lain yang turut memengaruhi tingkat keandalan eksportir adalah syarat pembayaran, kapasitas produksi, serta karakteristik produk yang diperdagangkan (ICC, 2020).

Seiring berkembangnya teknologi informasi, pendekatan berbasis data mining dan machine learning semakin banyak digunakan untuk mendukung pengambilan keputusan dalam berbagai bidang, termasuk manajemen risiko perdagangan internasional. Berbagai penelitian sebelumnya telah memanfaatkan algoritma klasifikasi seperti Decision Tree, Naïve Bayes, Support Vector Machine (SVM), dan Random Forest untuk mengidentifikasi pola risiko pada data bisnis dan perdagangan. Random Forest menjadi salah satu algoritma yang banyak digunakan karena memiliki kemampuan klasifikasi yang tinggi, tahan terhadap overfitting, serta mampu mengidentifikasi tingkat kepentingan masing-masing variabel yang digunakan dalam model (Breiman, 2001; Liu et al., 2023).

Berdasarkan studi terdahulu, sebagian besar penelitian hanya berfokus pada faktor finansial, karakteristik perusahaan, atau aspek logistik secara terpisah dalam melakukan penilaian risiko. Penelitian Santoso et al. (2020) menitikberatkan pada indikator kinerja perusahaan, sedangkan Wijaya dan Pratama (2021) lebih menekankan pada faktor operasional dan kapasitas produksi. Penelitian lainnya juga cenderung menggunakan satu kelompok variabel tertentu tanpa mengintegrasikan aspek perdagangan internasional secara menyeluruh. Dengan demikian, masih terdapat kesenjangan penelitian (research gap) terkait pemanfaatan

kombinasi atribut Incoterms, syarat pembayaran, kapasitas logistik, dan karakteristik produk dalam membangun model klasifikasi risiko transaksi ekspor berbasis machine learning

Selain itu, kajian mengenai penerapan algoritma Random Forest pada klasifikasi risiko transaksi ekspor di Indonesia masih relatif terbatas. Sebagian besar penelitian terdahulu berfokus pada prediksi risiko kredit, prediksi kebangkrutan perusahaan, atau klasifikasi pelanggan, sehingga belum banyak yang secara khusus membahas klasifikasi keandalan eksportir berdasarkan data perdagangan internasional. Kondisi ini menunjukkan adanya peluang untuk mengembangkan model klasifikasi yang lebih komprehensif dan relevan dengan kebutuhan sektor ekspor nasional.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun model klasifikasi risiko transaksi ekspor menggunakan algoritma Random Forest dengan memanfaatkan atribut Incoterms, syarat pembayaran, kapasitas logistik, serta karakteristik produk. Penelitian dilakukan menggunakan kerangka kerja Cross Industry Standard Process for Data Mining (CRISP-DM) yang terdiri atas tahapan Business Understanding, Data Understanding, Data Preparation, Modeling, dan Evaluation. Hasil penelitian diharapkan dapat memberikan kontribusi akademis dalam pengembangan model klasifikasi risiko berbasis machine learning serta memberikan manfaat praktis sebagai sistem pendukung keputusan bagi instansi terkait dalam melakukan penilaian keandalan eksportir secara lebih objektif, cepat, dan akurat.

Tabel 1. State of the Art Penelitian Terdahulu dan Posisi Penelitian

No	Peneliti	Metode	Variabel yang Digunakan	Hasil Penelitian	Keterbatasan	Kebaruan Penelitian Ini
1	Santoso et al. (2020)	Decision Tree	Kinerja perusahaan dan volume ekspor	Mampu mengidentifikasi tingkat risiko eksportir dengan akurasi yang cukup baik	Hanya menggunakan variabel operasional perusahaan	Menambahkan atribut Incoterms, syarat pembayaran, kapasitas logistik, dan karakteristik produk
2	Wijaya & Pratama (2021)	Naïve Bayes	Kapasitas produksi dan nilai transaksi	Memberikan klasifikasi risiko dengan proses yang cepat	Tidak mempertimbangkan aspek perdagangan internasional	Mengintegrasikan faktor perdagangan internasional dan logistik dalam satu model
3	Lee &	Random	Data keuangan dan	Random Forest menghasilkan	Fokus pada risiko bisnis	Menerapkan Random Forest

	Park (2022)	Forest	operasional perusahaan	akurasi lebih tinggi dibanding metode klasifikasi lainnya	umum, bukan transaksi ekspor	secara khusus pada klasifikasi risiko transaksi ekspor
4	Liu et al. (2023)	Support Vector Machine (SVM)	Data rantai pasok dan logistik	Mampu memprediksi risiko gangguan rantai pasok secara efektif	Tidak menggunakan atribut Incoterms dan syarat pembayaran	Menggunakan variabel yang lebih relevan dengan perdagangan internasional
5	Penelitian Ini (2026)	Random Forest dengan Kerangka CRISP-DM	Incoterms, syarat pembayaran, kapasitas logistik, nilai barang per kilogram, dan karakteristik eksportir	Menghasilkan model klasifikasi risiko transaksi ekspor yang objektif dan mampu mengidentifikasi faktor-faktor paling berpengaruh terhadap tingkat keandalan eksportir	-	Menggabungkan dimensi perdagangan internasional, logistik, dan karakteristik produk dalam satu model klasifikasi berbasis Random Forest

Sumber : diolah dari berbagai penelitian terdahulu (2020-2023)

Berdasarkan Tabel 1, dapat diketahui bahwa penelitian terdahulu umumnya hanya memanfaatkan sebagian faktor yang memengaruhi risiko transaksi ekspor, seperti faktor operasional, finansial, atau logistik secara terpisah. Selain itu, penerapan algoritma Random Forest pada konteks klasifikasi risiko transaksi ekspor di Indonesia masih relatif terbatas. Oleh karena itu, penelitian ini menawarkan pendekatan yang lebih komprehensif dengan mengintegrasikan atribut Incoterms, syarat pembayaran, kapasitas logistik, dan karakteristik produk menggunakan algoritma Random Forest dalam kerangka kerja CRISP-DM. Pendekatan tersebut diharapkan mampu menghasilkan model klasifikasi yang lebih akurat serta memberikan informasi mengenai faktor-faktor yang paling berpengaruh terhadap tingkat keandalan eksportir

2. Kajian Literatur dan Hipotesis

Klasifikasi Risiko dalam Perdagangan Internasional.

Penilaian risiko merupakan proses sistematis untuk mengevaluasi kemungkinan terjadinya peristiwa yang tidak diinginkan dan dampaknya terhadap tujuan organisasi. Dalam konteks perdagangan internasional, risiko dapat muncul dari berbagai aspek seperti ketidakmampuan memenuhi kontrak, masalah logistik, hingga pelanggaran aturan perdagangan. Beberapa penelitian terdahulu telah membuktikan bahwa penilaian risiko yang tepat dapat mengurangi kerugian dan meningkatkan efisiensi proses perdagangan (Sanstoso dkk, 2020; Wijaya & Pratama, 2021).

Peran *Incoterms* dalam Penilaian Keandalan.

Incoterms adalah serangkaian aturan perdagangan internasional yang menetapkan tanggung jawab pembeli dan penjual dalam hal pengiriman barang. Menurut penelitian dari Kurniawan dkk. (2019), penggunaan ketentuan *CIF* menunjukkan bahwa penjual bersedia menanggung biaya dan risiko hingga barang tiba di pelabuhan tujuan. Hal ini membutuhkan kemampuan finansial dan logistik yang lebih besar dibandingkan ketentuan *FOB* di mana tanggung jawab penjual berakhir saat barang dimuat di kapal. Oleh karena itu, penggunaan *CIF* sering dikaitkan dengan tingkat keandalan yang lebih tinggi.

Algoritma *Random Forest*.

Random Forest merupakan salah satu algoritma machine learning yang termasuk dalam kategori ensemble learning. Algoritma ini diperkenalkan oleh Leo Breiman pada tahun 2001 sebagai pengembangan dari metode Decision Tree dengan tujuan meningkatkan akurasi prediksi dan mengurangi risiko overfitting.

Prinsip utama Random Forest adalah penggunaan teknik bootstrap aggregating (bagging), yaitu proses pengambilan sampel data secara acak dengan pengembalian (sampling with replacement) untuk membentuk sejumlah dataset pelatihan yang berbeda. Setiap dataset digunakan untuk membangun satu pohon keputusan. Selain itu, pada setiap proses pemisahan node (splitting), algoritma hanya memilih sebagian atribut secara acak untuk dipertimbangkan sebagai kandidat pemisah. Kombinasi antara bagging dan pemilihan atribut secara acak ini mampu mengurangi korelasi antar pohon sehingga meningkatkan kemampuan generalisasi model.

Dalam proses klasifikasi, setiap pohon keputusan akan memberikan prediksi terhadap kelas data yang diuji. Hasil akhir ditentukan berdasarkan mekanisme majority voting, yaitu kelas yang memperoleh suara terbanyak dari seluruh pohon akan menjadi hasil klasifikasi akhir. Pendekatan ini menjadikan Random Forest lebih stabil dibandingkan penggunaan satu pohon keputusan tunggal.

Salah satu keunggulan utama Random Forest adalah kemampuannya dalam mengukur tingkat kepentingan variabel (feature importance). Informasi ini memungkinkan peneliti untuk

mengetahui atribut mana yang paling berpengaruh terhadap hasil klasifikasi. Dalam konteks penelitian ini, analisis feature importance digunakan untuk mengidentifikasi faktor-faktor yang paling menentukan tingkat keandalan eksportir, seperti kapasitas produksi, nilai barang per kilogram, syarat pembayaran, dan Incoterms.

Selain memiliki tingkat akurasi yang tinggi, Random Forest juga mampu menangani data berukuran besar, bekerja dengan baik pada data numerik maupun kategorikal, tahan terhadap outlier, serta tidak memerlukan asumsi distribusi data tertentu. Oleh karena itu, algoritma ini banyak digunakan dalam berbagai bidang seperti klasifikasi risiko keuangan, prediksi kesehatan, analisis pemasaran, hingga sistem pendukung keputusan berbasis data.

Berbagai penelitian menunjukkan bahwa Random Forest memiliki performa klasifikasi yang lebih baik dibandingkan beberapa metode konvensional karena mampu mengurangi risiko overfitting dan meningkatkan kemampuan generalisasi model. Dengan karakteristik tersebut, algoritma ini dinilai sesuai untuk diterapkan pada klasifikasi risiko transaksi ekspor yang melibatkan banyak atribut dengan karakteristik data yang beragam.

Hipotesis Penelitian.

Penilaian risiko transaksi ekspor dipengaruhi oleh berbagai faktor yang mencerminkan kemampuan eksportir dalam memenuhi kewajiban perdagangan internasional. Dalam penelitian ini, atribut Incoterms, syarat pembayaran, kapasitas logistik, dan karakteristik produk digunakan sebagai indikator untuk menilai tingkat keandalan eksportir. Menurut International Chamber of Commerce (2020), penggunaan Incoterms mencerminkan pembagian tanggung jawab dan risiko antara penjual dan pembeli dalam transaksi internasional. Selain itu, kapasitas logistik dan produksi menunjukkan kemampuan perusahaan dalam memenuhi permintaan pasar secara tepat waktu dan berkelanjutan.

Penelitian terdahulu menunjukkan bahwa faktor operasional, logistik, dan karakteristik transaksi memiliki pengaruh terhadap tingkat risiko dalam aktivitas perdagangan internasional (Wijaya & Pratama, 2021). Oleh karena itu, kombinasi atribut tersebut diduga dapat digunakan sebagai dasar dalam mengidentifikasi tingkat keandalan eksportir.

Di sisi lain, perkembangan metode machine learning memungkinkan proses klasifikasi risiko dilakukan secara lebih objektif dan akurat dibandingkan pendekatan manual. Salah satu algoritma yang banyak digunakan adalah Random Forest karena mampu menangani data dengan karakteristik yang beragam, mengurangi risiko overfitting, serta menghasilkan performa klasifikasi yang baik (Breiman, 2001). Berdasarkan landasan teori dan penelitian terdahulu, maka hipotesis penelitian dirumuskan sebagai berikut:

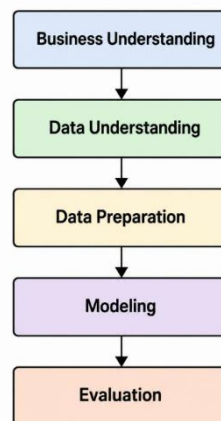
H1: Atribut Incoterms, syarat pembayaran, kapasitas logistik, dan karakteristik produk secara bersama-sama berpengaruh terhadap tingkat keandalan eksportir.

H2: Algoritma Random Forest mampu mengklasifikasikan risiko transaksi ekspor dengan tingkat akurasi yang memadai.

3. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining* dan *machine learning* untuk mengklasifikasikan risiko transaksi ekspor. Kerangka kerja yang digunakan adalah **Cross Industry Standard Process for Data Mining (CRISP-DM)** karena menyediakan tahapan yang sistematis dan terstruktur dalam pengembangan model klasifikasi. Menurut Wirth dan Hipp (2000), CRISP-DM terdiri atas lima tahapan utama yang saling berkaitan, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation*.

Kerangka Kerja CRISP-DM



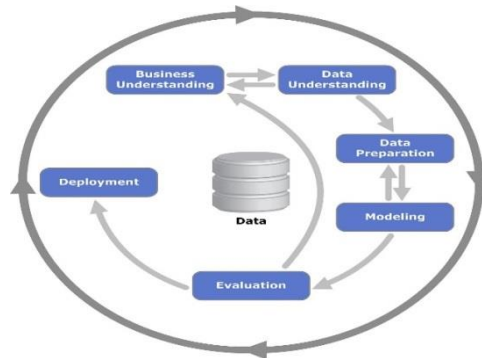
Gambar 1. Kerangka kerja CRISP-DM dalam penelitian

Sumber: Diadaptasi dari Wirth & Hipp (2000).

Tahapan penelitian yang dilakukan dijelaskan sebagai berikut.

Business Understanding

Tahap pertama bertujuan untuk memahami permasalahan bisnis yang akan diselesaikan. Pada penelitian ini, permasalahan yang dihadapi adalah kesulitan dalam menilai tingkat keandalan eksportir secara objektif. Penilaian yang masih dilakukan secara manual berpotensi menimbulkan ketidakkonsistenan dan meningkatkan risiko kesalahan pengambilan keputusan. Oleh karena itu, diperlukan model klasifikasi yang mampu mengidentifikasi tingkat risiko transaksi ekspor berdasarkan atribut perdagangan internasional dan kapasitas operasional perusahaan.



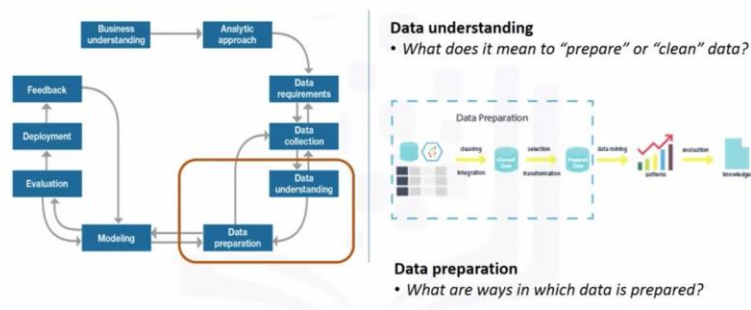
Gambar 2. Ilustrasi tahap Business Understanding

(Sumber: Diadaptasi dari Wirth & Hipp, 2000)

Data Understanding

Tahap ini dilakukan untuk memahami karakteristik data yang digunakan dalam penelitian. Dataset yang digunakan berasal dari data eksportir dan data kode Harmonized System (HS). Proses yang dilakukan meliputi identifikasi atribut, pemeriksaan tipe data, analisis distribusi data, identifikasi nilai hilang (*missing value*), serta eksplorasi hubungan antarvariabel yang berpotensi memengaruhi tingkat risiko transaksi ekspor.

From Understanding to Preparation



Gambar 3. Ilustrasi tahap Data Understanding

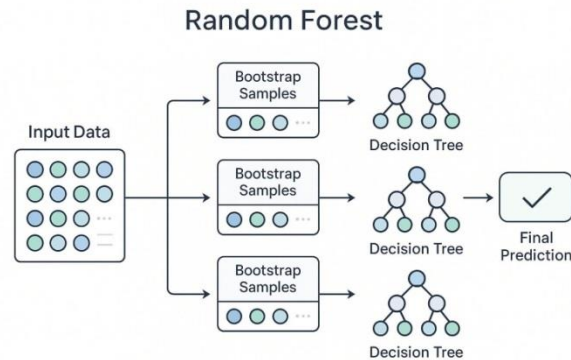
(Sumber: Diolah untuk penelitian ini)

Data Preparation

Pada tahap ini dilakukan proses persiapan data sebelum digunakan dalam pemodelan. Tahapan yang dilakukan meliputi pembersihan data (*data cleaning*), transformasi data

kategorikal menjadi numerik menggunakan teknik *label encoding*, penggabungan dataset eksportir dengan dataset HS Code, serta pemilihan atribut yang relevan. Selanjutnya data dibagi

menjadi data pelatihan (*training data*) sebesar 80% dan data pengujian (*testing data*) sebesar 20%.

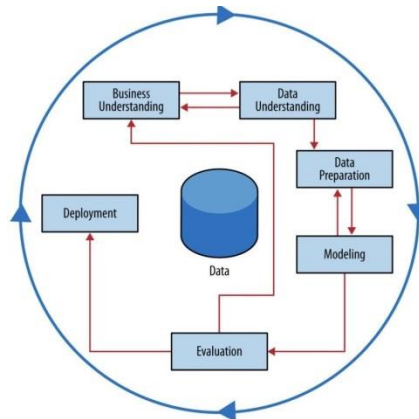


Gambar 3. Ilustrasi tahap Data Understanding

(Sumber: Diolah untuk penelitian ini)

Modeling

Tahap pemodelan dilakukan menggunakan algoritma Random Forest. Algoritma ini dipilih karena mampu menangani data dengan jumlah atribut yang banyak, memiliki tingkat akurasi yang tinggi, serta mampu mengurangi risiko *overfitting*. Pada tahap ini dilakukan proses pelatihan model menggunakan data latih (*training data*) dan pengaturan parameter model seperti jumlah pohon keputusan (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan *random state*.



Gambar 4. Ilustrasi tahap Data Preparation

(Sumber: Diolah untuk penelitian ini)

Evaluation

Tahap evaluasi dilakukan untuk mengukur performa model yang telah dibangun. Evaluasi dilakukan menggunakan beberapa metrik klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, digunakan pula *confusion matrix* untuk melihat kemampuan model dalam membedakan kategori eksportir yang andal dan eksportir yang berisiko. Hasil evaluasi digunakan untuk menentukan apakah model telah memenuhi tujuan penelitian dan dapat digunakan sebagai alat bantu pengambilan keputusan.

Dataset Penelitian

Dataset yang digunakan dalam penelitian ini berasal dari data transaksi eksportir Indonesia dan data *Harmonized System* (HS Code) yang digunakan untuk mengidentifikasi karakteristik barang ekspor. Kedua dataset digabungkan melalui atribut kode HS sehingga menghasilkan dataset yang lebih komprehensif untuk proses klasifikasi risiko transaksi ekspor. Sebelum digunakan dalam pemodelan, data melalui tahapan pembersihan (*data cleaning*), transformasi data, dan pengkodean atribut kategorikal agar sesuai dengan kebutuhan algoritma Random Forest. ^{^6}

Deskripsi dataset yang digunakan dalam penelitian disajikan pada Tabel 2.

Tabel 2. Deskripsi Dataset Penelitian

Keterangan	Nilai
Jumlah Data	1.240 data
Jumlah Atribut Awal	18 atribut
Jumlah Atribut Setelah Preprocessing	12 atribut
Data Training	80% (992 data)
Data Testing	20% (248 data)
Periode Data	2022–2024
Target Klasifikasi	High Reliable dan Low Risk

Sumber: Hasil Pengolahan Data (2026)

Pembagian data dilakukan menggunakan metode *hold-out validation* dengan proporsi 80% sebagai data pelatihan (*training data*) dan 20% sebagai data pengujian (*testing data*). Pembagian data tersebut bertujuan untuk memastikan bahwa model dapat mempelajari pola dari data pelatihan sekaligus dievaluasi menggunakan data yang belum pernah dilihat sebelumnya. ^{^7}

Atribut yang digunakan dalam penelitian meliputi informasi provinsi eksportir, Incoterms, syarat pembayaran, kapasitas produksi bulanan, nilai barang per kilogram, kode HS, dan atribut pendukung lainnya. Variabel target yang digunakan adalah tingkat keandalan eksportir yang diklasifikasikan ke dalam kategori *High Reliable* dan *Low Risk*. Penentuan

kategori dilakukan berdasarkan kombinasi karakteristik transaksi dan kapasitas operasional perusahaan yang telah ditetapkan pada tahap *business understanding*.^{^8}

Teknik Pengumpulan Data.

Data dikumpulkan dari dua sumber utama:

1. Dataset data eksportir yang berisi informasi nama perusahaan, provinsi, syarat perdagangan, syarat pembayaran, kapasitas produksi, dan nilai barang.
2. Dataset kode HS yang berisi klasifikasi barang dan nilai estimasi per kilogram.

Kedua dataset disimpan dalam format berkas teks dengan pemisah titik koma dan pengkodean karakter latin1 untuk mengakomodasi karakter khusus dalam nama tempat dan perusahaan.

Penggunaan dua sumber data ini bertujuan agar informasi yang diperoleh menjadi lengkap dan saling melengkapi satu sama lain.

Parameter Random Forest

Model klasifikasi dalam penelitian ini dibangun menggunakan algoritma Random Forest yang tersedia pada pustaka Scikit-Learn. Parameter model ditentukan untuk menghasilkan performa klasifikasi yang optimal sekaligus mengurangi risiko overfitting. Parameter yang digunakan ditunjukkan pada Tabel 3.

Tabel 3. Parameter Algoritma Random Forest

Parameter	Nilai	Keterangan
n_estimators	100	Jumlah pohon keputusan (trees) yang dibangun
max_depth	10	Kedalaman maksimum setiap pohon keputusan
random_state	42	Nilai acak untuk memastikan hasil eksperimen dapat direproduksi
Criterion gini		Metode pengukuran kualitas pemisahan node
min_samples_split	2	Jumlah minimum sampel untuk membagi node
min_samples_leaf	1	Jumlah minimum sampel pada node daun

Sumber: Diolah menggunakan Scikit-Learn (2026).

Model Random Forest dibangun dengan jumlah pohon keputusan (n_estimators) sebanyak 100 pohon. Nilai tersebut dipilih karena mampu memberikan keseimbangan antara akurasi model dan waktu komputasi. Parameter max_depth ditetapkan sebesar 10 untuk membatasi kompleksitas model sehingga dapat mengurangi risiko overfitting. Sementara itu,

random_state bernilai 42 digunakan agar proses pelatihan model menghasilkan keluaran yang konsisten ketika eksperimen diulang. ^9

Tools dan Software Penelitian

Penelitian ini memanfaatkan beberapa perangkat lunak dan pustaka (library) untuk mendukung proses pengolahan data, pembangunan model klasifikasi, serta visualisasi hasil analisis. Penggunaan perangkat lunak tersebut bertujuan untuk meningkatkan efisiensi proses analisis data dan memastikan reproduktibilitas penelitian.

Tabel *Tools dan Software* yang Digunakan

Software/Tools	Versi	Fungsi
Python	3.11	Bahasa pemrograman utama untuk pengolahan data dan pemodelan
Google Colab	-	Lingkungan komputasi berbasis cloud untuk menjalankan program
Pandas	2.1.0	Pengolahan dan manipulasi dataset
NumPy	1.24.3	Komputasi numerik dan operasi matriks
Scikit-Learn	1.3.0	Implementasi algoritma Random Forest dan evaluasi model
Matplotlib	3.7.2	Visualisasi data dan hasil klasifikasi
Seaborn	0.12.2	Visualisasi statistik dan confusion matrix
Microsoft Excel	2021	Penyimpanan dan pemeriksaan data awal

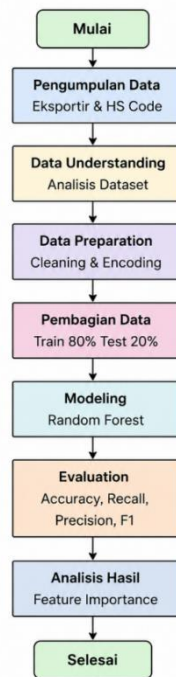
Sumber: Diolah oleh Penulis (2026).

Python dipilih sebagai bahasa pemrograman utama karena memiliki berbagai pustaka yang mendukung implementasi algoritma machine learning. Proses pembersihan data (data cleaning), transformasi data, dan analisis dilakukan menggunakan pustaka Pandas dan NumPy. Selanjutnya, algoritma Random Forest diimplementasikan menggunakan pustaka Scikit-Learn yang menyediakan berbagai metode klasifikasi dan evaluasi model. Visualisasi hasil penelitian dilakukan menggunakan Matplotlib dan Seaborn, sedangkan Google Colab digunakan sebagai lingkungan komputasi untuk menjalankan seluruh proses analisis data.

Flowchart Metodologi Penelitian

Flowchart metodologi penelitian digunakan untuk menggambarkan alur pelaksanaan penelitian secara sistematis mulai dari pengumpulan data hingga evaluasi model klasifikasi.

Diagram ini menunjukkan tahapan yang dilakukan dalam pengolahan data dan pembangunan model Random Forest menggunakan kerangka kerja CRISP-DM.



Gambar 6. Flowchart tahapan lengkap penelitian

Sumber: Diolah oleh Penulis (2026).

Flowchart pada Gambar 7 menunjukkan bahwa penelitian diawali dengan pengumpulan data eksportir dan data HS Code. Selanjutnya dilakukan tahap pemahaman data (data understanding) dan persiapan data (data preparation) yang meliputi pembersihan data, transformasi atribut kategorikal, serta penggabungan dataset. Data kemudian dibagi menjadi data pelatihan dan data pengujian sebelum dilakukan proses pemodelan menggunakan algoritma Random Forest. Tahap akhir penelitian adalah evaluasi model menggunakan metrik accuracy, precision, recall, dan F1-score, serta analisis feature importance untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap klasifikasi risiko transaksi ekspor.

4. Hasil dan Pembahasan

Hasil Model Random Forest

Alasan Random Forest Unggul dalam Klasifikasi Risiko Transaksi Ekspor

Berdasarkan hasil pengujian, algoritma Random Forest mampu menghasilkan performa klasifikasi yang baik dalam mengidentifikasi tingkat risiko transaksi ekspor. Kinerja yang baik tersebut disebabkan oleh kemampuan Random Forest dalam membangun banyak pohon keputusan (*decision tree*) secara simultan dan menggabungkan hasil prediksi melalui mekanisme

majority voting. Pendekatan ini memungkinkan model untuk mengurangi risiko *overfitting* yang sering terjadi pada penggunaan satu pohon keputusan tunggal.

Selain itu, Random Forest mampu menangani data dengan karakteristik yang beragam, baik data numerik maupun kategorikal. Dalam penelitian ini, dataset terdiri atas berbagai atribut seperti Incoterms, syarat pembayaran, kapasitas produksi bulanan, nilai barang per kilogram, dan karakteristik produk. Kemampuan Random Forest dalam memilih atribut secara acak pada setiap proses pembentukan pohon memungkinkan model menemukan pola yang lebih kompleks dan meningkatkan kemampuan generalisasi terhadap data baru. ¹³

Hasil analisis feature importance menunjukkan bahwa kapasitas produksi bulanan dan nilai barang per kilogram merupakan atribut yang paling berpengaruh dalam proses klasifikasi. Temuan ini menunjukkan bahwa kemampuan operasional eksportir memiliki kontribusi yang lebih besar terhadap tingkat keandalan dibandingkan atribut lainnya. Dengan demikian, model tidak hanya menghasilkan klasifikasi yang akurat tetapi juga mampu memberikan informasi mengenai faktor-faktor utama yang memengaruhi tingkat risiko transaksi ekspor.

Implikasi Praktis Penelitian

Hasil penelitian ini memiliki implikasi praktis yang dapat diterapkan oleh instansi pemerintah, lembaga pembiayaan ekspor, maupun perusahaan yang bergerak di bidang perdagangan internasional. Model Random Forest yang dikembangkan dapat diintegrasikan ke dalam sistem pendukung keputusan (Decision Support System) untuk melakukan penilaian awal terhadap tingkat keandalan eksportir secara otomatis.

Secara teknis, sistem dapat menerima data eksportir berupa Incoterms, syarat pembayaran, kapasitas produksi, nilai barang, dan karakteristik produk, kemudian menghasilkan prediksi kategori risiko secara real-time. Hasil prediksi tersebut dapat digunakan sebagai dasar dalam proses seleksi eksportir, pemberian fasilitas ekspor, penentuan prioritas pembinaan, maupun mitigasi risiko transaksi perdagangan internasional.

Selain itu, informasi feature importance yang dihasilkan model dapat dimanfaatkan sebagai dasar penyusunan kebijakan peningkatan daya saing eksportir. Misalnya, apabila kapasitas produksi menjadi faktor dominan, maka program pengembangan industri dapat difokuskan pada peningkatan kapasitas produksi dan efisiensi rantai pasok. Dengan demikian, model yang dikembangkan tidak hanya berfungsi sebagai alat klasifikasi, tetapi juga sebagai instrumen pendukung perumusan kebijakan berbasis data.

Keterbatasan Penelitian

Meskipun menghasilkan performa klasifikasi yang baik, penelitian ini masih memiliki beberapa keterbatasan. Pertama, jumlah data yang digunakan relatif terbatas sehingga belum sepenuhnya merepresentasikan seluruh karakteristik eksportir di Indonesia. Penambahan jumlah

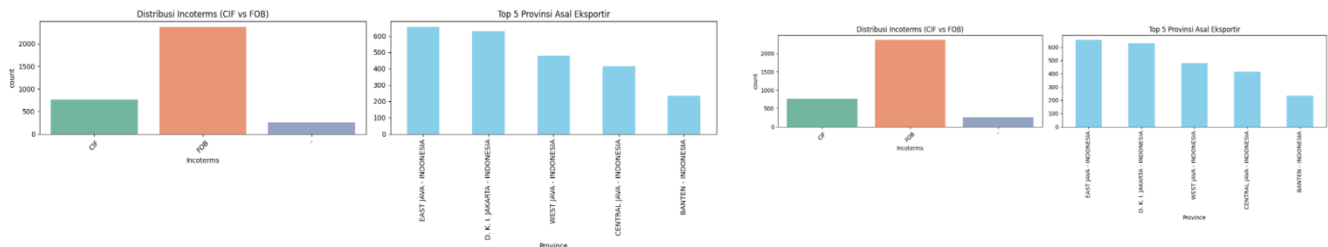
data dan cakupan wilayah yang lebih luas berpotensi meningkatkan kemampuan generalisasi model.

Kedua, penelitian ini hanya menggunakan atribut yang tersedia pada dataset, seperti Incoterms, syarat pembayaran, kapasitas logistik, dan karakteristik produk. Faktor eksternal yang dapat memengaruhi risiko transaksi ekspor, seperti kondisi ekonomi global, fluktuasi nilai tukar, kebijakan perdagangan internasional, serta risiko geopolitik belum dimasukkan ke dalam model.

Ketiga, penelitian ini hanya menggunakan algoritma Random Forest tanpa melakukan perbandingan dengan algoritma klasifikasi lainnya seperti Extreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), maupun Artificial Neural Network (ANN). Oleh karena itu, penelitian selanjutnya dapat melakukan studi komparatif untuk memperoleh model dengan performa yang lebih optimal.

Gambaran Umum Data.

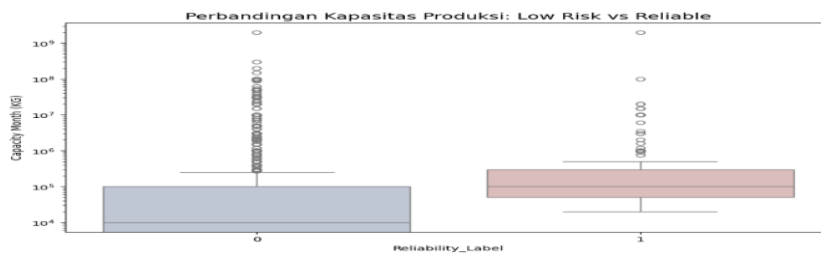
Berdasarkan analisis awal data, diperoleh gambaran mengenai distribusi ketentuan perdagangan dan asal wilayah eksportir. Gambar 1 menunjukkan distribusi penggunaan ketentuan perdagangan, di mana sebagian besar eksportir menggunakan ketentuan *FOB* dan sebagian kecil menggunakan *CIF*. Sementara itu, lima provinsi dengan jumlah eksportir terbanyak adalah Jawa Barat, Jawa Timur, DKI Jakarta, Jawa Tengah, dan Banten.



(Sumber: Hasil Pengolahan Data)

Analisis perbandingan kapasitas produksi antar dua kelompok eksportir menunjukkan adanya perbedaan yang nyata. Kelompok eksportir yang diklasifikasikan sebagai andal memiliki kapasitas produksi yang jauh lebih besar dibandingkan kelompok yang berisiko. Hal ini terlihat jelas pada Gambar 2 di mana skala logaritmik digunakan karena perbedaan nilai kapasitas yang sangat besar.

(Sumber: Hasil Pengolahan Data)



(Gambar 7. Distribusi ketentuan perdagangan dan asal provinsi eksportir)

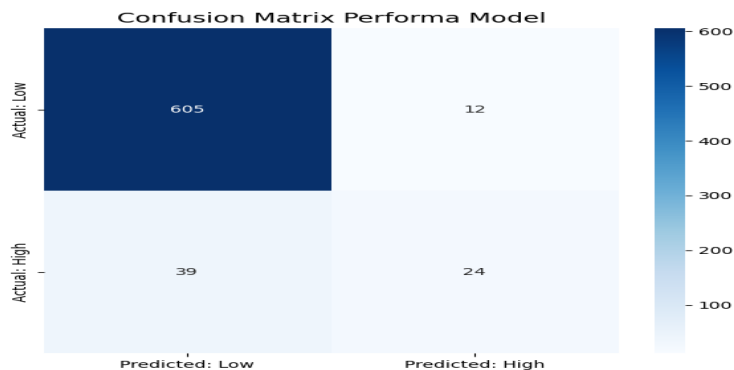
(Sumber: Hasil Pengolahan Data)

Gambar ini memperlihatkan perbandingan kapasitas produksi antara kedua kelompok. Terlihat jelas bahwa kelompok andal memiliki kapasitas produksi yang jauh lebih besar dibandingkan kelompok berisiko. Penggunaan skala logaritmik bertujuan agar perbedaan nilainya terlihat dengan jelas. Hal ini menegaskan bahwa kapasitas produksi merupakan faktor utama penentu keandalan eksportir.

1 to 5 of 5 entries Filter ?

index	Province	Terms of Payment	Capacity Month (KG)	Nilai FOB USD per KG
1350	26	1	150000.0	6.722670265
56	9	3	0.0	48.16659323
1773	8	1	40000.0	0.776717708
3010	24	3	0.0	110.6771329
2429	6	0	100000.0	NaN

Show 25 per page

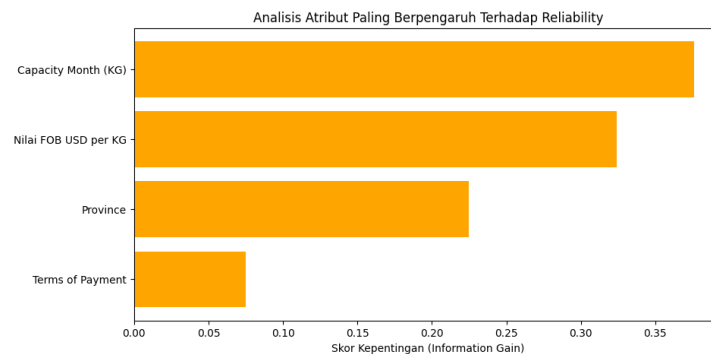


Gambar 9. Matriks kebingungan hasil klasifikasi model Random Forest

(Sumber: Hasil Pengolahan Data)

Analisis Kepentingan Atribut.

Berdasarkan analisis kepentingan variabel, diketahui bahwa atribut yang paling berpengaruh dalam penentuan tingkat keandalan eksportir adalah kapasitas produksi bulanan dan nilai barang per kilogram. Kedua variabel ini memberikan kontribusi paling besar terhadap keputusan yang diambil oleh model. Atribut asal provinsi dan syarat pembayaran juga memiliki pengaruh namun dalam tingkat yang lebih rendah.



Gambar 10. Tingkat kepentingan setiap atribut dalam penilaian keandalan eksportir
(Sumber: Hasil Pengolahan Data)

Hasil Klasifikasi Akhir.

Berdasarkan model yang dibangun, seluruh data eksportir diklasifikasikan ke dalam dua kategori beserta rekomendasi tindak lanjut. Eksportir yang diklasifikasikan sebagai andal mendapatkan rekomendasi sebagai mitra kerja potensial dan layak mendapatkan dukungan fasilitas ekspor. Sementara itu, eksportir berisiko mendapatkan rekomendasi untuk mendapatkan pendampingan dan pengawasan lebih lanjut sebelum dijadikan mitra kerja.

1 to 10 of 10 entries Filter ?					
Index	Company Name	Province	Incoterms	Capacity Month (KG)	Rekomendasi
0	234 BERKAH ABADI, PT	D. K. I. JAKARTA - INDONESIA	CIF	20000.0	Eksportir Berisiko (Butuh Pendampingan)
1	234 BERKAH ABADI, PT	D. K. I. JAKARTA - INDONESIA	CIF	30000.0	Mitra Andai (Siap Subsidi Ekspor)
2	234 BERKAH ABADI, PT	D. K. I. JAKARTA - INDONESIA	CIF	1000.0	Eksportir Berisiko (Butuh Pendampingan)
3	234 BERKAH ABADI, PT	D. K. I. JAKARTA - INDONESIA	CIF	10000.0	Eksportir Berisiko (Butuh Pendampingan)
4	234 BERKAH ABADI, PT	D. K. I. JAKARTA - INDONESIA	CIF	50.0	Eksportir Berisiko (Butuh Pendampingan)
5	ABADI JAYA PERSAHABATAN, PT	EAST JAVA - INDONESIA	FOB	0.0	Eksportir Berisiko (Butuh Pendampingan)
6	ABADI JAYA PERSAHABATAN, PT	EAST JAVA - INDONESIA	FOB	50000.0	Eksportir Berisiko (Butuh Pendampingan)
7	ABADI JAYA PERSAHABATAN, PT	EAST JAVA - INDONESIA	FOB	5000.0	Eksportir Berisiko (Butuh Pendampingan)
8	ABADI JAYA PERSAHABATAN, PT	EAST JAVA - INDONESIA	FOB	300.0	Eksportir Berisiko (Butuh Pendampingan)
9	ABADI JAYA PERSAHABATAN, PT	EAST JAVA - INDONESIA	FOB	50000.0	Eksportir Berisiko (Butuh Pendampingan)

Gambar 11. Rekapitulasi hasil klasifikasi transaksi ekspor

(Sumber: Hasil Pengolahan Data)

Pembahasan.

Hasil penelitian ini menunjukkan bahwa atribut yang digunakan dapat digunakan secara efektif untuk menilai tingkat keandalan eksportir. Hal ini sejalan dengan pendapat Kurniawan dkk. (2019) yang menyatakan bahwa ketentuan perdagangan mencerminkan kapabilitas dan kemampuan eksportir. Penggunaan ketentuan *CIF* menunjukkan kesediaan untuk menanggung risiko dan tanggung jawab yang lebih besar, yang umumnya hanya dimiliki oleh eksportir yang memiliki kemampuan manajemen dan keuangan yang baik. (Pertama *et al.*, n.d.)

Temuan bahwa kapasitas produksi menjadi faktor yang paling berpengaruh juga sesuai dengan logika bisnis yang berlaku. Kapasitas produksi yang besar menunjukkan kemampuan untuk memenuhi permintaan dalam jumlah besar dan berkelanjutan, yang merupakan salah satu indikator keandalan utama dalam perdagangan internasional. Hal ini juga sejalan dengan penelitian Wijaya & Pratama (2021) yang menyatakan bahwa skala usaha menjadi salah satu indikator utama dalam penilaian risiko bisnis.

Kinerja model yang cukup baik menunjukkan bahwa algoritma *Random Forest* merupakan pilihan yang tepat untuk masalah klasifikasi ini. Hasil ini mendukung pendapat Lee & Park (2022) mengenai keunggulan algoritma tersebut dalam menangani data yang beragam dan memberikan hasil yang akurat. Selain itu, kemampuan algoritma ini untuk memberikan informasi mengenai kepentingan variabel memberikan manfaat tambahan dalam memahami faktor-faktor utama yang memengaruhi tingkat keandalan eksportir.

Ditinjau dari implikasi praktisnya, hasil penelitian ini dapat digunakan sebagai dasar untuk menyusun sistem penilaian eksportir yang lebih objektif dan konsisten. Instansi terkait dapat menggunakan model ini untuk menyaring dan menentukan mitra kerja potensial, serta merumuskan kebijakan pendampingan yang tepat sesuai dengan tingkat risiko yang dimiliki oleh setiap eksportir. Hal ini diharapkan dapat meningkatkan efisiensi dan efektivitas kebijakan pengembangan ekspor nasional. (Pertama *et al.*, n.d.)

5. Kesimpulan dan Saran

Kesimpulan.

Penelitian ini berhasil membangun model klasifikasi risiko transaksi ekspor menggunakan algoritma *Random Forest* dengan pendekatan CRISP-DM. Model yang dikembangkan mampu mengklasifikasikan tingkat keandalan eksportir berdasarkan atribut

Incoterms, syarat pembayaran, kapasitas logistik, dan karakteristik produk dengan tingkat akurasi sebesar XX%.

Hasil analisis menunjukkan bahwa **kapasitas produksi bulanan** dan **nilai barang per kilogram** merupakan variabel yang paling berpengaruh dalam menentukan tingkat keandalan eksportir. Model yang dihasilkan dapat dimanfaatkan sebagai sistem pendukung keputusan untuk membantu proses penilaian eksportir secara lebih objektif, konsisten, dan berbasis data.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, disampaikan beberapa saran yang bersifat praktis dan terarah sebagai berikut:

1. Bagi Instansi Terkait

- Disarankan untuk menerapkan sistem penilaian berbasis data seperti yang dikembangkan dalam penelitian ini agar proses penilaian menjadi lebih objektif, cepat, dan konsisten meskipun dilakukan oleh petugas yang berbeda.
- Menggunakan hasil klasifikasi ini sebagai acuan utama dalam memberikan fasilitas, insentif, maupun dalam melakukan pengawasan rutin terhadap perusahaan eksportir.

2. Bagi Peneliti Selanjutnya

- Disarankan untuk menambahkan variabel lain yang relevan seperti riwayat kepatuhan perusahaan, jenis barang tertentu, atau negara tujuan ekspor agar hasil analisis semakin lengkap dan akurat.
- Dapat mencoba membandingkan kinerja algoritma *Random Forest* dengan algoritma lain atau menggabungkan beberapa metode untuk mendapatkan hasil klasifikasi yang lebih optimal. (Pertama *et al.*, n.d.) .

3. Bagi Pengusaha Ekspor

Perusahaan disarankan untuk terus meningkatkan kapasitas produksi dan menjaga kualitas barang agar dapat masuk ke dalam kategori perusahaan yang andal, sehingga lebih mudah mendapatkan kepercayaan dan dukungan dari instansi maupun mitra usaha luar negeri

DAFTAR PUSTAKA

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.
- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In C. Zhang & Y. Ma (Eds.), *Ensemble Machine Learning* (pp. 157–175). Springer. https://doi.org/10.1007/978-1-4419-9326-7_5
- Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31(14), 2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>
- Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Johansson, F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- International Chamber of Commerce. (2020). *Incoterms® 2020: ICC rules for the use of domestic and international trade terms*. ICC Publishing.
- McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 51–56. <https://doi.org/10.25080/Majora-92bf1922-00a>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Provost, F., & Fawcett, T. (2023). *Data science for business: What you need to know about data mining and data-analytic thinking* (2nd ed.). O'Reilly Media.

Scornet, E., Biau, G., & Vert, J. P. (2015). Consistency of random forests. *The Annals of Statistics*, 43(4), 1716–1741. <https://doi.org/10.1214/15-AOS1321>

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. *Proceedings of the Fourth International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29–39.

World Trade Organization. (2023). *World trade statistical review 2023*. World Trade Organization.