

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang dan Identifikasi Masalah**

Industri logistik dan pengiriman barang merupakan fondasi vital dalam menopang aktivitas perdagangan internasional. Di Indonesia, PT Samudera Indonesia Tbk menonjol sebagai salah satu perusahaan pelayaran utama yang memfasilitasi pengiriman barang ekspor ke berbagai destinasi global. Dalam konteks ini, biaya pengiriman menjadi elemen krusial yang tidak hanya berdampak pada efisiensi operasional perusahaan, tetapi juga memengaruhi keputusan pelanggan dan arah strategi bisnis secara keseluruhan.

Akan tetapi, penetapan biaya pengiriman dalam praktiknya sering kali dihadapkan pada tantangan, khususnya terkait akurasi prediksinya. Berbagai faktor memengaruhi besaran biaya ini, termasuk jenis komoditas, berat dan volume muatan, negara tujuan, tipe kapal yang digunakan, biaya asuransi, serta bea cukai. Menurut (Anjelina Surya Dewi et al., 2024), menggaris bawahi bahwa kelancaran ekspor-impor juga dipengaruhi oleh faktor seperti ketepatan dokumen, bea cukai, jenis muatan, fasilitas pelabuhan, dan waktu pengiriman, yang secara tidak langsung berkontribusi pada biaya logistik. Selain itu, biaya untuk pengurusan dokumen ekspor, dan tarif pengiriman yang bergantung pada jarak serta layanan, merupakan komponen biaya ekspor yang menuntut perhitungan cermat. Apabila proses penentuan biaya ini tidak didukung oleh

sistem prediksi yang mumpuni, perusahaan berisiko mengalami kerugian akibat estimasi yang tidak tepat, alokasi sumber daya yang suboptimal, dan penurunan kepuasan pelanggan karena informasi biaya yang tidak sesuai.

Variasi karakteristik logistik pada setiap jenis produk turut menambah kompleksitas dalam memprediksi biaya pengiriman. Seperti produk *frozen food* memerlukan pengiriman dengan suhu terkontrol menggunakan kapal jenis *Reefer*, sementara produk curah cair membutuhkan kapal tipe *Tanker*. Produk berisiko tinggi seperti barang elektronik seringkali memerlukan asuransi tambahan. Semua perlakuan spesifik ini secara langsung memengaruhi struktur biaya pengiriman. Oleh karena itu, pendekatan prediktif harus mampu mengadaptasi perbedaan logistik berdasarkan jenis komoditas dan tipe kapal yang digunakan.

Sejalan dengan kemajuan teknologi, implementasi algoritma *machine learning* menawarkan solusi prospektif untuk membangun model prediksi biaya yang lebih akurat. Salah satu metode yang umum diterapkan adalah *Linear Regression*, yang dipilih karena kemampuannya menjelaskan hubungan matematis linear antar variabel dalam pola data serta kemudahan interpretasinya. Riset oleh (Shiela et al., 2024), menunjukkan bahwa metode regresi linear efektif memprediksi tren profitabilitas logistik dengan tingkat kesalahan yang rendah, menjadikannya relevan untuk kebutuhan perencanaan biaya pengiriman. Senada dengan itu, penelitian (Hidayat & Wahyudi, 2024), juga menegaskan akurasi tinggi algoritma regresi linear

dalam memprediksi tren pengiriman logistik, memberikan wawasan berharga untuk optimasi layanan pengiriman.

Namun, *Linear Regression* memiliki keterbatasan dalam menangani data yang kompleks dan non-linear, terutama ketika variabel-variabelnya menunjukkan hubungan interaktif atau tidak proporsional. Untuk mengatasi kelemahan ini, metode alternatif seperti *Random Forest Regression* dapat dimanfaatkan. Algoritma ensemble ini tidak hanya tangguh terhadap overfitting, tetapi juga efisien dalam memproses variabel kategorikal maupun numerik, serta mampu mengidentifikasi bobot kepentingan fitur (*feature importance*). Studi (Pambudi et al., 2020), melaporkan bahwa *Random Forest* mencapai akurasi 76,6% dalam memprediksi status pengiriman barang, dan penelitian (Zebua et al., 2022), menunjukkan nilai R2 sebesar 85,87% dalam prediksi tarif penerbangan yang dinamis dan kompleks. Lebih jauh, studi (Pambudi et al., 2020), menegaskan bahwa *Random Forest* menunjukkan akurasi terbaik sebesar 79,70% dalam memprediksi status pengiriman barang, menyoroti keunggulannya pada data logistik.

Mempertimbangkan kompleksitas faktor biaya pengiriman dan karakteristik data logistik yang beragam, penelitian ini berfokus pada perbandingan kinerja dua algoritma regresi, yaitu *Linear Regression* dan *Random Forest Regression*, dalam mengembangkan model prediksi biaya pengiriman ekspor. Tujuan akhirnya adalah menghasilkan model yang presisi, yang dapat menjadi landasan pengambilan

keputusan biaya, sekaligus mendorong peningkatan efisiensi pengelolaan logistik dalam industri pelayaran.

### **1.2 Rumusan Masalah**

Bagaimana membangun dan membandingkan model prediksi biaya pengiriman ekspor pada PT Samudera Indonesia menggunakan algoritma Linear Regression dan Random Forest Regression berdasarkan fitur logistik seperti jenis barang, jenis kapal, berat, volume, asuransi, dan bea cukai?

### **1.3 Batasan Masalah**

Pada penelitian prediksi biaya pengiriman ekspor ini, ruang lingkup permasalahan dibatasi agar penelitian lebih terarah dan sesuai dengan tujuan yang ingin dicapai. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan merupakan data historis pengiriman ekspor pada periode Januari hingga Maret 2025 yang diperoleh dari PT Samudera Indonesia.
2. Penelitian ini hanya membandingkan dua algoritma machine learning, yaitu Linear Regression dan Random Forest Regression.
3. Fitur (variabel input) yang digunakan dalam model prediksi didasarkan pada data yang tersedia dalam dataset, yaitu berat barang, volume, jenis barang (komoditas), negara tujuan, jenis kapal, biaya asuransi, dan bea cukai.
4. Evaluasi performa model dilakukan menggunakan metrik regresi yaitu R-squared ( $R^2$ ), Mean Absolute Error (MAE), dan Root Mean Squared Error (RMSE).

5. Penelitian tidak membahas pengaruh kondisi eksternal yaitu cuaca, inflasi, atau kebijakan perdagangan yang mungkin turut memengaruhi biaya pengiriman.
6. Sistem prediksi ini tidak mencakup pelacakan posisi kapal secara *real-time* atau manajemen rute transit yang spesifik. Fokus penelitian terbatas pada prediksi biaya akhir berdasarkan data historis yang tersedia, tanpa melacak pergerakan kapal antar pelabuhan transit.

#### **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini adalah untuk:

1. Membangun model prediksi biaya pengiriman ekspor pada PT Samudera Indonesia dengan menggunakan algoritma machine learning Linear Regression dan Random Forest Regression berdasarkan data historis pengiriman.
2. Mengidentifikasi pengaruh variabel-variabel logistik yaitu berat barang, volume, jenis barang (komoditas), negara tujuan, jenis kapal, biaya asuransi, dan bea cukai terhadap total biaya pengiriman ekspor.
3. Membandingkan performa kedua algoritma (Linear Regression dan Random Forest Regression) dalam hal akurasi dan keandalan prediksi biaya pengiriman menggunakan metrik evaluasi R-squared, MAE, dan RMSE.
4. Menentukan fitur-fitur paling berpengaruh dalam memodelkan biaya pengiriman ekspor untuk mendukung pengambilan keputusan operasional yang lebih efisien.

## **1.5 Manfaat Penelitian**

Manfaat yang dapat diperoleh dari penelitian ini adalah:

1. Menyediakan estimasi biaya pengiriman yang lebih akurat dan berbasis data, sehingga dapat meningkatkan efisiensi operasional dan perencanaan biaya di sektor logistik ekspor.
2. Memberikan pemahaman yang lebih komprehensif mengenai hubungan antara karakteristik logistik (jenis barang dan jenis kapal) dengan total biaya pengiriman.
3. Menyajikan perbandingan performa dua algoritma machine learning dalam konteks prediksi biaya, sehingga dapat menjadi acuan dalam pemilihan metode analitik yang tepat di masa mendatang.
4. Menjadi referensi akademik dan praktis dalam penerapan machine learning untuk prediksi biaya logistik, khususnya pada industri pelayaran dan ekspor-impor.

## **1.6 Metode Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan pemanfaatan algoritma machine learning dalam membangun model prediksi biaya pengiriman ekspor. Fokus dari metodologi ini adalah membandingkan performa antara Linear Regression dan Random Forest Regression, serta menganalisis fitur-fitur logistik yang paling berpengaruh terhadap biaya pengiriman. Adapun tahapan metodologi dalam penelitian ini adalah sebagai berikut:

## 1. Pengumpulan Data

Data yang digunakan merupakan data primer dan sekunder. Data primer diperoleh dari dokumentasi internal PT Samudera Indonesia Tbk (Januari–Maret 2025) yang mencakup: berat barang, volume, negara tujuan, jenis barang (komoditas), jenis kapal, biaya asuransi, bea cukai, dan total biaya pengiriman sebagai variabel target. Selain itu, dilakukan wawancara informal dengan staf operasional untuk memahami hubungan variabel logistik dengan biaya pengiriman secara praktis. Studi literatur juga dilakukan terhadap jurnal ilmiah yang relevan untuk mendukung pemilihan metode algoritma dan evaluasi model.

## 2. Pra-Pemrosesan Data

Data yang diperoleh akan melalui tahap pembersihan dan transformasi, dengan langkah-langkah sebagai berikut:

- a. Menghapus data kosong, duplikat, atau nilai tidak valid.
- b. Encoding variabel kategorikal seperti negara tujuan, jenis kapal, dan komoditas menggunakan teknik Label Encoding atau One-Hot Encoding.
- c. Normalisasi fitur numerik (seperti berat, volume, biaya) agar skala antar variabel setara.
- d. Analisis korelasi dan feature importance untuk memilih fitur yang paling relevan terhadap biaya pengiriman.

### 3. Pemisahan Dataset

Dataset yang digunakan dalam penelitian ini dibagi menjadi dua bagian, yaitu 80% sebagai data pelatihan (training set) dan 20% sebagai data pengujian (testing set). Pembagian ini dilakukan untuk memastikan bahwa evaluasi performa model dilakukan secara objektif, di mana model dilatih pada sebagian data dan diuji pada data yang belum pernah dilihat sebelumnya untuk mengukur akurasi dan generalisasi prediksi secara adil.

### 4. Pelatihan Model Linear Regression

Penelitian ini akan melatih dan membandingkan dua algoritma prediksi, yaitu Linear Regression dan Random Forest Regression. Linear Regression digunakan sebagai model dasar (baseline) karena mampu menjelaskan hubungan matematis secara linier antara fitur-fitur input dan variabel target, yaitu total biaya pengiriman. Sementara itu, Random Forest Regression digunakan untuk menangani data yang bersifat kompleks dan non-linier. Algoritma ini mampu mengungkap interaksi antar fitur secara lebih mendalam dan memberikan informasi mengenai pentingnya masing-masing fitur (feature importance), sehingga dapat membantu dalam analisis variabel yang paling berpengaruh terhadap hasil prediksi.

#### **Evaluasi Model :**

Untuk menilai performa kedua model, digunakan beberapa metrik evaluasi regresi, yaitu R-squared ( $R^2$ ) untuk mengukur seberapa baik model menjelaskan variasi data,

Mean Absolute Error (MAE) untuk menghitung rata-rata selisih absolut antara nilai prediksi dan nilai aktual, serta Root Mean Squared Error (RMSE) yang memberikan gambaran terhadap tingkat kesalahan prediksi dalam satuan yang sama dengan target output. Evaluasi ini dilakukan agar diperoleh gambaran objektif mengenai akurasi dan keandalan masing-masing model dalam memprediksi biaya pengiriman ekspor.

Selain itu, pada tahap pelatihan model dilakukan pembagian data dengan lebih cermat agar hasil evaluasi tidak hanya terlihat baik pada data latih, tetapi juga tetap konsisten ketika diuji menggunakan data yang belum pernah dipakai sebelumnya. Pengujian juga dilakukan lebih dari satu kali untuk memastikan perbandingan antara model dasar dan model pembanding tidak menunjukkan selisih yang berlebihan.

Dalam proses evaluasi, Linear Regression dipilih karena lebih mudah dipahami dan dapat menunjukkan pengaruh langsung dari setiap fitur terhadap biaya pengiriman. Di sisi lain, Random Forest dinilai lebih mampu membaca pola data yang lebih beragam dan tidak linier. Oleh karena itu, perbandingan kedua model tidak hanya dilihat dari nilai metriknya saja, tetapi juga dari sejauh mana model tersebut dapat diterapkan secara praktis.

Jika ditemukan perbedaan hasil yang terlalu besar antara data latih dan data uji, penyesuaian sederhana pada parameter model dilakukan agar performanya tetap stabil dan tidak menimbulkan hasil prediksi yang menyimpang saat digunakan.

## 5. Deployment

Model yang telah dibangun dapat diimplementasikan ke dalam aplikasi berbasis web menggunakan framework seperti Flask atau Streamlit. Tujuan dari pengembangan ini adalah agar tim operasional dapat dengan mudah melakukan estimasi biaya pengiriman secara mandiri tanpa harus bergantung pada perhitungan manual. Aplikasi web tersebut akan menerima input dari pengguna berupa berat barang, volume, jenis barang, jenis kapal, biaya asuransi, bea cukai, dan negara tujuan, kemudian menghasilkan output berupa estimasi total biaya pengiriman yang dihitung berdasarkan model prediksi yang telah dilatih sebelumnya.

## 6. Pengujian dan Evaluasi Sistem

Tahap akhir berupa uji coba sistem secara fungsional dan evaluasi akurasi prediksi. Pengujian melibatkan data testing serta validasi manual dari pihak perusahaan (jika memungkinkan), untuk memastikan sistem dapat digunakan secara nyata dan memberikan hasil prediksi yang sesuai.

## **1.7 Sistematika Penulisan**

Penulisan skripsi ini disusun secara sistematis ke dalam beberapa bab agar memudahkan pembaca dalam memahami alur penelitian, dari mulai latar belakang permasalahan, landasan teori, hingga hasil dan kesimpulan. Setiap bab memiliki peran penting dalam menyampaikan isi penelitian secara terstruktur dan logis. Adapun sistematika penulisan dalam skripsi ini adalah sebagai berikut:

### **BAB I PENDAHULUAN**

Bab ini berisi penjelasan mengenai latar belakang permasalahan yang melandasi dilakukannya penelitian, serta urgensi dari pengembangan sistem prediksi biaya pengiriman ekspor. Di dalamnya dijelaskan pula identifikasi dan rumusan masalah, tujuan dan manfaat penelitian, batasan ruang lingkup, metode pengumpulan data, serta gambaran umum sistematika penulisan skripsi.

### **BAB II LANDASAN TEORI**

Bab ini membahas teori-teori dasar yang mendukung penelitian, yaitu konsep logistik ekspor, biaya pengiriman, jenis komoditas, dan jenis kapal. Selain itu, dijelaskan pula algoritma machine learning yang digunakan, yaitu Linear Regression dan Random Forest Regression, termasuk prinsip kerja, kelebihan, kekurangan, serta metrik evaluasi seperti  $R^2$ , MAE, dan RMSE.

### **BAB III METODOLOGI PENELITIAN**

Bab ini menjelaskan pendekatan yang digunakan dalam penelitian ini, mulai dari pengumpulan data, pra-pemrosesan data, pembagian dataset, pelatihan model, hingga evaluasi kinerja model. Proses ini bertujuan untuk membangun dan membandingkan dua model prediksi biaya pengiriman berdasarkan data historis dari perusahaan.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini memaparkan hasil prediksi biaya pengiriman menggunakan Linear Regression dan Random Forest Regression, dengan perbandingan akurasi berdasarkan metrik evaluasi. Analisis fitur utama yang memengaruhi biaya dilakukan dengan mengacu pada data lapangan dan wawancara untuk memperkuat interpretasi hasil model.

### **BAB V PENUTUP**

Bab ini menyimpulkan bahwa kedua algoritma memiliki performa berbeda dengan fitur-fitur tertentu yang dominan dalam prediksi biaya. Saran diberikan untuk pengembangan penelitian dan penerapan sistem prediksi ini di industri logistik agar lebih efektif dan akurat.