

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Berlian menjadi salah satu jenis komoditas dengan nilai ekonomi yang tinggi sekaligus merupakan salah satu jenis materi paling kompleks di Bumi dengan perbandingan 58 kali lebih keras dari benda apapun [1], [2], dan penggunaannya sangat terbatas. Salah satu alasan konsumen untuk membeli atau memakai berlian adalah persepsi akan kelangkaan dan biaya yang tinggi atau bisa dengan kata lain berlian digunakan sebagai perhiasan yang sangat mewah. Meskipun berlian telah menjadi simbol kekayaan selama ribuan tahun, penggunaannya secara substansial berubah pada abad ke-20 ketika cincin pertunangan berlian menjadi tren yang digunakan oleh banyak pasangan di seluruh dunia. *De Beers*, penyedia sekaligus penjual berlian terbesar di dunia, telah menggunakan *tagline* "a diamond is forever" sejak 1948 untuk mempromosikannya [3].

Di seluruh dunia, permintaan untuk berlian yang dijual secara terpisah dan telah diproses mencapai 25 miliar USD sedangkan untuk penjualan berlian dalam bentuk perhiasan mencapai 79 miliar USD [4]. Di pasar perhiasan, biasanya berlian dijual di toko perhiasan. Namun dengan adanya *e-commerce*, semakin banyak berlian ditawarkan secara *online*, dengan tujuan untuk mendapatkan basis konsumen yang lebih signifikan [5]. Seiring dengan meningkatnya permintaan baru untuk berlian, perkembangan *e-commerce* diyakini dapat memudahkan calon pembeli untuk dapat membandingkan harga satu berlian pada sebuah toko dengan harga berlian pada toko yang lain.

Berlian seringkali ditemukan dalam bentuk perhiasan yang sudah jadi, namun tidak sedikit juga investor yang melihat berlian sebagai aset investasi, menurut Reza Fahmi, Senior Vice President, Head of Retail, Product Research & Distribution Divion, PT. Henan Putihrai Asset Management "Berlian memiliki pangsa investornya tersendiri. Sebab, berinvestasi dalam berlian memiliki risiko yang rendah.". Analisis DCFX Futures Lukman Leong beranggapan serupa, harga

berlian sangat bervariasi. Ia mencotohkan, untuk berlian warna umum, yakni kuning dengan clarity umum yaitu VS1 dengan berat kurang dari 0,5 karat di level 1.986 USD. Lalu, kadar 0,51-0,99 karat sebesar 3.242 USD, kadar 1,00-1,49 karat sebesar 6.002 USD, kadar 1,50-1,99 sebesar 8.789 USD, kadar 2,00-2,99 sebesar 12.161 USD. Kemudian, 3,00-3,99 sebesar 16.862 USD, kadar 4,00-4,99 sebesar 21.535 USD, dan kadar di atas 5 karat sebesar 28.759 USD keatas [6].

Jika dibandingkan dengan komoditas lain yang umum seperti emas, berlian memiliki nilai yang tidak terpengaruh oleh keadaan ekonomi global seperti inflasi, dan tidak memiliki harga acuan tertentu seperti halnya emas. Harga Berlian dipengaruhi oleh berbagai faktor, seperti karat, ukuran, warna, serta tingkat kejernihan. Tidak adanya acuan harga berlian seperti halnya harga emas menyebabkan sebuah perusahaan jual beli berlian memerlukan sistem untuk memprediksi harganya.

Di era modern ini teknologi sudah berkembang sangat pesat, tentu hal ini didasari dengan keinginan manusia untuk mempermudah segala urusannya. Salah satu teknologi yang sedang banyak dibutuhkan saat ini ialah kecerdasan buatan. Kecerdasan buatan dapat dikatakan sudah menjadi kebutuhan primer dalam dunia industri, banyak sekali manfaat yang dapat dicapai dari penerapan kecerdasan buatan. Bicara tentang kecerdasan buatan tentunya tidak lepas dari suatu metode yang bernama Pembelajaran Mesin. Pembelajaran mesin dapat didefinisikan secara luas sebagai metode komputasi yang menggunakan pengalaman untuk meningkatkan kinerja atau membuat prediksi yang akurat [7]. Banyak sekali implementasi dari pembelajaran mesin ini, salah satunya ialah untuk memprediksi sesuatu berdasarkan data yang kemudian diproses sehingga mencapai output yang diinginkan. Bicara tentang pembelajaran mesin maka tidak akan lepas dari metode, teknik, serta algoritma yang membangunnya, ada tiga metode pembelajaran mesin yang umum digunakan yaitu *supervised learning*, *unsupervised learning*, *semi-supervised learning*, dan *reinforcement learning*, dari ketiga metode tersebut masing masing memiliki teknik yang beragam, ada beberapa teknik yang familiar dengan metode pembelajaran mesin yaitu teknik regresi, klasifikasi, klusterisasi, segmentasi, dan lainnya.

Salah satu metode *supervised learning* yang populer adalah pohon keputusan karena konsepnya yang mudah dipahami dan mirip dengan cara berpikir manusia. Pohon keputusan dapat dibagi menjadi dua jenis, yaitu pohon tunggal dan *ensemble tree* (pohon gabungan). *Ensemble tree* adalah pengembangan dari pohon tunggal yang menghasilkan beberapa pohon klasifikasi dari data dan menggunakan hasil prediksi dari setiap pohon untuk membuat keputusan. *Boosting* dan *bagging* adalah dua teknik yang umum digunakan pada *ensemble tree*, yang memiliki perbedaan pada cara pembentukan pohon. *Boosting* (Adaboost) membangun pohon secara sekuensial, sementara *bagging* membangun pohon secara paralel. Contoh metode *boosting* adalah *adaptive boosting*, sedangkan contoh metode *bagging* adalah *Random Forest* [8].

Metode *ensemble tree* seperti *random forest* dan *adaptive boosting* sering digunakan dalam berbagai bidang seperti *data mining* dan *machine learning* untuk melakukan klasifikasi, prediksi, dan analisis data. Sedangkan KNN (*K-Nearest Neighbors*) sendiri masuk kedalam kelompok algoritma *instance-based learning* (juga dikenal sebagai algoritma *lazy learning*), KNN dapat dikatakan sebagai salah satu algoritma *instance-based learning* terbaik karena walaupun dari segi kompleksitas jelas algoritma *ensemble tree* memiliki kompleksitas yang lebih tinggi dibandingkan *lazy learning* seperti KNN, namun KNN memiliki keunggulan tersendiri yaitu kemudahan penggunaan, interpretabilitas, fleksibilitas, waktu pelatihan yang cenderung lebih singkat, algoritma ini juga dapat menangani masalah multi-kelas dengan baik [9]. Ketiga algoritma tersebut layak dibandingkan sebagaimana ketiganya merupakan perwakilan terbaik dari masing masing kelompok regresi, *random forest* dengan *ensemble tree* berbasis *bagging* (paralel), Adaboost dengan *ensemble tree* berbasis *boosting* (sekuensial), serta KNN dengan *lazy learning* yang tentunya dari segi kompleksitas algoritma ini kalah dari algoritma *ensemble*, namun masih dapat bersaing dalam aspek yang lain.

Pemilihan ketiga algoritma tersebut diperkuat pula dengan beberapa penelitian terdahulu dengan algoritma dan studi kasus yang serupa, dan terbukti bahwa banyak penelitian yang mengunggulkan ketiganya dibandingkan yang lain, sebagai contoh pada penelitian yang dilakukan oleh Waad Alsuraihi, dkk. [10]

yang membandingkan algoritma *random forest*, regresi linear, *neural network*, dan *gradient boosting* dalam prediksi harga berlian yang hasilnya algoritma *Random Forest* memiliki performa terbaik diantara ketiganya berdasarkan perbandingan MSE dan RMSE-nya, selanjutnya pada penelitian yang dilakukan oleh Avinash Chandra Pandey, dkk. [11] yang menggunakan algoritma *adaboost regressor* dalam melakukan peramalan harga perhiasan emas dan berlian. Selanjutnya penelitian yang dilakukan oleh Prof. Amruta. A, dkk. [12], dimana penelitian tersebut menggunakan algoritma *KNN Regressor* untuk melakukan peramalan harga berlian.

Penelitian ini akan menggunakan *dataset* harga berlian yang bersumber dari *Kaggle* [12]. Pada pengukuran performa nantinya akan dilihat dari beberapa faktor seperti MSE (*Mean Square Error*), RMSE (*Root Mean Square Error*), R2-Score (*R Squared*), akurasi prediksi, serta durasi pelatihan dari ketiga algoritma tersebut. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam melakukan peramalan harga berlian atau kasus yang serupa di masa mendatang.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan sebelumnya, maka rumusan masalah dalam penelitian ini ialah algoritma manakah diantara ketiga algoritma yaitu *Random Forest*, *K-Nearest Neighbors* dan *Adaboost Regressor* yang memiliki kinerja terbaik dalam peramalan harga berlian?

## 1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini ialah untuk menemukan algoritma terbaik diantara tiga buah algoritma yaitu *Random Forest*, *K-Nearest Neighbors* dan *Adaboost Regressor* yang dapat diterapkan untuk peramalan harga berlian.

## 1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberi manfaat sebagai berikut :

### 1. Bagi Pengguna

Mendapat opsi yang tepat ketika dihadapi dengan pemilihan algoritma untuk penyelesaian suatu masalah prediksi terutama dalam memprediksi harga berlian.

### 2. Bagi Penulis

Memberikan pengalaman dan pengetahuan baru tentang bagaimana cara yang tepat untuk membandingkan suatu algoritma dengan algoritma lain pada studi kasus tertentu.

### 3. Bagi Akademik

Dapat dijadikan sebagai referensi bagi bidang studi yang tertarik memecahkan masalah serupa dengan menciptakan perbandingan algoritma menggunakan studi kasus yang serupa ataupun berbeda.

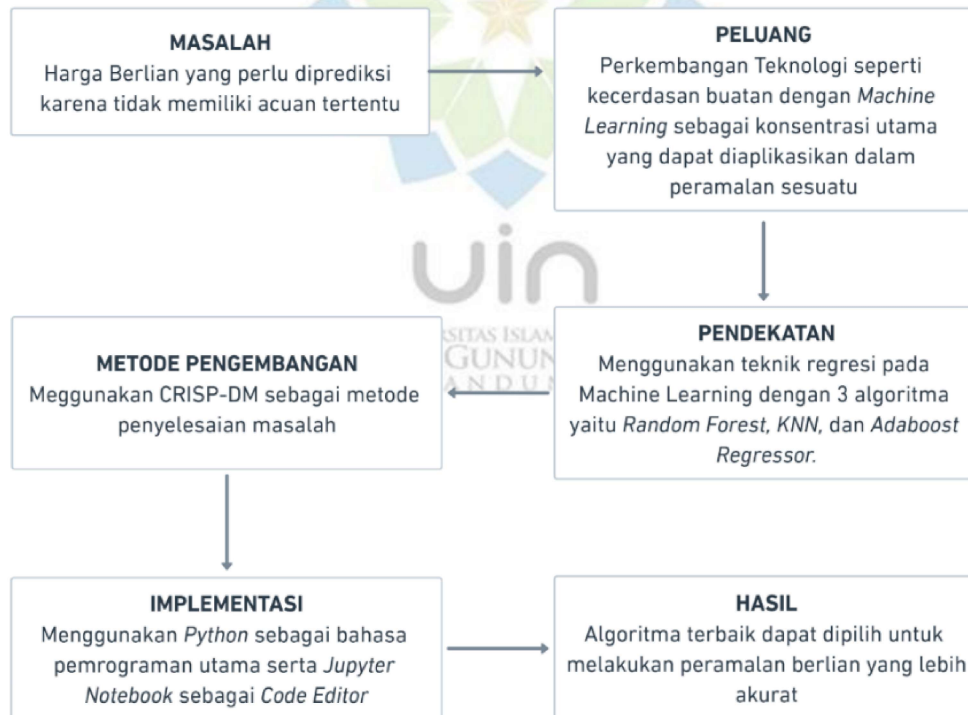
## 1.5 Batasan Masalah

Adapun batasan masalah yang akan peneliti terapkan pada penelitian kali ini ialah :

- a. Algoritma yang dibandingkan hanya terbatas pada *Random Forest*, KNN (*K-Nearest Neighbors*), dan *Adaboost Regressor*.
- b. Perbandingan yang dilakukan hanya terbatas pada kinerja algoritma yaitu berupa durasi pelatihan, MSE, RMSE, *R2-Score*, sampai menghasilkan prediksi harga Berlian.
- c. *Dataset* yang digunakan berasal dari situs *Kaggle* [13] yang memiliki 53.940 baris dan 10 kolom.

## 1.6 Kerangka Pemikiran

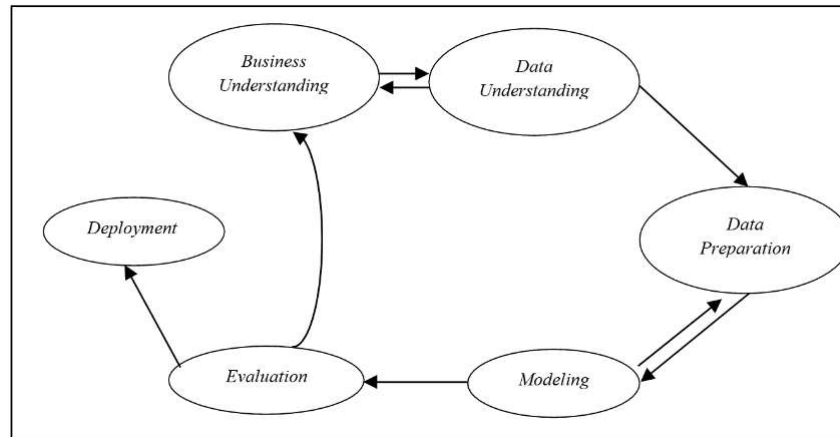
Berlian termasuk kategori materi yang sangat keras, berlian juga seringkali digunakan untuk menunjukkan simbol kekayaan seseorang dengan dijadikannya sebagai perhiasan dalam bentuk kalung, cincin, dan sebagainya. Barang mewah seperti berlian sebaiknya memiliki acuan harga agar tidak ada ketimpangan harga pada satu toko dengan toko yang lain, namun justru sebaliknya berlian belum memiliki acuan harga seperti halnya emas. Ketiga algoritma yaitu *random forest*, *k-nearest neighbors* (KNN), dan *adaboost regressor* dipilih untuk dibandingkan kinerjanya pada penelitian ini. Penilaian akan dilakukan dengan membandingkan berbagai parameter, termasuk akurasi, performa metrik pengujian, akurasi prediksi, dan faktor-faktor lain yang dapat mempengaruhi hasil dan kinerja algoritma atau metode tersebut. Dalam membandingkan ketiga algoritma tersebut, peneliti akan memastikan bahwa proses pengujian dan evaluasi dilakukan secara objektif dan cermat agar dapat menghasilkan kesimpulan yang akurat.



**Gambar 1.1 Kerangka Pemikiran**



## 1.7 Metodologi Penelitian



**Gambar 1.2 Alur CRISP-DM [14]**

Untuk menyelesaikan masalah yang ada maka diperlukan perbandingan algoritma yang nantinya dapat disesuaikan dengan kebutuhan. Proses pengembangan model dengan beberapa algoritma ini menggunakan metode CRISP-DM. Proses pada tahapan ini yaitu *business understanding*, *data understanding*, dan *data preparation*. Tahapan ketiga merupakan tahapan *modeling* dan *evaluation*. Proses pada tahapan ini juga merupakan bagian dari metodologi CRISP-DM [14]. Proses pada tahapan ini yaitu, proses *modeling*, proses *evaluation* dan proses *deployment*.

### a. *Business Understanding*

Pada tahap pertama yaitu dilakukan pemahaman bisnis serta penjelasan terhadap masalah *data mining* sehingga dapat mencapai tujuan. Pada tahap *business understanding* dilakukan proses penentuan tujuan bisnis, memahami situasi dan kondisi pada saat penelitian dan menetapkan sebuah tujuan dari penelitian yang dilakukan ke dalam permasalahan yang diselesaikan dengan pembelajaran mesin.

### b. *Data understanding*

Tahap *data understanding* merupakan tahap persiapan, dilakukan pengecekan terhadap data yang digunakan, mengumpulkan data awal serta melakukan identifikasi pada kualitas data. Dalam *data understanding*, data yang digunakan akan melalui proses deskripsi dari setiap variabelnya.

c. *Data Preparation*

Pada *Data Preparation* dilakukan persiapan data dengan menyesuaikan *dataset* agar dapat sesuai dengan kebutuhan yang akan digunakan saat tahap pemodelan. Pada tahap *Data Preparation* pertama akan dilakukan *data selection* untuk melakukan seleksi terhadap data yang dibutuhkan, *data preprocessing* untuk mempersiapkan data mentah sehingga siap untuk menjadi data yang dapat digunakan pada pemodelan, dan *data transformation* untuk melakukan transformasi data.

d. *Modeling*

Pada tahap ini, pemodelan menggunakan teknik regresi. Teknik pemodelan regresi yang digunakan adalah algoritma *Random Forest*, KNN, dan *Adaboost Regressor*. Selama proses pemodelan, parameter pada algoritma regresi dioptimalkan untuk meningkatkan kualitas model.

e. *Evaluation*

Setelah melakukan tahap pemodelan, tahap selanjutnya adalah evaluasi. Evaluasi dilakukan untuk mengevaluasi kualitas dari model yang telah dibangun dan memeriksa apakah model tersebut sesuai dengan tujuan dan kebutuhan penggunaannya. Evaluasi juga dapat digunakan untuk memperbaiki dan meningkatkan kinerja model jika ditemukan kelemahan atau kekurangan pada tahap pemodelan.

f. *Deployment*

Setelah tahap evaluasi selesai dilakukan untuk menilai secara detail hasil dari seluruh model, selanjutnya dilakukan tahap publikasi model. Pada tahap ini biasanya melibatkan *software development* seperti *website* atau *mobile app*.



## **1.8 Sistematika Penulisan**

Sistematika penulisan laporan tugas akhir mencakup urutan penulisan bab, konten yang disajikan pada setiap bab, serta hubungan antara satu bab dengan bab lainnya. Contoh tata penulisan laporan sebelumnya dapat memberikan gambaran tentang hal tersebut.

### **BAB I PENDAHULUAN**

Bab ini berisi tentang latar belakang penelitian, rumusan masalah penelitian, tujuan penelitian, batasan masalah penelitian, kerangka penelitian, dan sistematika penulisan. Pendahuluan juga berfungsi sebagai pengantar yang dapat membantu pembaca untuk memahami konteks dan urgensi penelitian yang dilakukan serta mengapa penelitian tersebut perlu dilakukan.

### **BAB II KAJIAN LITERATUR**

Pada bab ini dijabarkan apa saja literatur atau penelitian terdahulu, konsep-konsep, teori-teori, dan model-model yang menjadi landasan dalam perancangan sistem. Kajian literatur juga dapat membantu peneliti untuk menemukan kelemahan atau kesenjangan dalam penelitian terdahulu yang dapat dijadikan dasar untuk mengembangkan penelitian selanjutnya.

### **BAB III METODOLOGI PENELITIAN**

Metodologi penelitian merupakan tahapan sistematis yang dirancang untuk memperoleh data secara objektif dan valid dalam rangka menjawab pertanyaan penelitian yang telah ditetapkan. Bab ini berisi penjelasan bagaimana rangkaian metodologi yang digunakan atau implementasinya pada penelitian.

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

Bab IV dari laporan tugas akhir ini berisi dua hal utama. Yang pertama adalah presentasi hasil penelitian yang didapatkan melalui tahapan penelitian yang telah dilakukan. Yang kedua adalah pembahasan dari hasil penelitian tersebut, yang diarahkan untuk menjawab pertanyaan yang dirumuskan pada awal penelitian.

## **BAB V SIMPULAN DAN SARAN**

Bab ini berfokus pada penarikan kesimpulan dari hasil penelitian yang telah diperoleh dan analisis dari pembahasan sebelumnya. Selain itu, bab ini juga memberikan saran untuk penelitian selanjutnya yang dapat dilakukan untuk meningkatkan kualitas dari penelitian tersebut.

