

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Di era digital ini, keberlangsungan sebuah bisnis dipengaruhi oleh hasil analisis data sehingga di era ini banyak perusahaan yang memerlukan seorang ahli dalam menganalisis data. Oleh karena itu PT. Hacktivate Teknologi Indonesia menghadirkan solusi bagi siapapun yang ingin memahami dan mengembangkan keterampilan pengolahan data dengan menggunakan bahasa pemrograman *Python*. *Python for Data Science* adalah salah satu program yang diadakan oleh PT. Hacktivate Teknologi Indonesia. Dalam program ini peneliti diajari dasar - dasar yang diperlukan untuk menjadi seorang *Data Scientist*. Program ini diperuntukkan untuk learner yang hendak mempelajari bahasa pemrograman *python* dan hendak mengimplementasikan statistik, *machine learning*, dan visualisasi informasi/data lewat *toolkit python* yang populer seperti *pandas*, *numpy*, *matplotlib*, *scikit-learning*, *seaborn*, dan lain-lain supaya memperoleh wawasan mengenai data. Pada course terakhir learner akan mempelajari hal-hal yang berkaitan dengan *Machine learning* yang akan mendukung kompetensi *learner* di bidang *Data Science*. Adapun harapan yang diinginkan setelah mengikuti *Online Program* ini hingga akhir ialah learner akan menjadi seorang *Data Scientist* yang mampu menghadapi dunia kerja di bidang *Data Science*.

Jasa transportasi *online* telah menjadi salah satu kebutuhan dasar manusia karena mempermudah manusia melakukan aktivitas berpergian ke suatu tempat. Dengan adanya jasa transportasi *online*, konsumen dapat merasakan berbagai manfaat diantaranya tidak takut tersesat, keamanan terjamin, kenyamanan, hemat waktu, dll. Kemunculan jasa transportasi *online* tentu karena adanya perusahaan-perusahaan yang menyediakan jasa tersebut seperti *uber* dan *lyft*. Bersamaan dengan berjalannya waktu, data yang dihasilkan oleh pengguna transportasi *online* *uber* dan *lyft* semakin banyak, seperti data riwayat perjalanan konsumen dan harga untuk setiap perjalanan yang ditempuh oleh konsumen. Data-data tersebut dicatat oleh perusahaan dan dijadikan sebagai dataset. Dengan dataset/data-data riwayat tersebut dapat digunakan sebagai referensi untuk pembuatan *machine learning*

dalam membandingkan harga antara dua jasa transportasi *online* tersebut dan memprediksi harga perjalanan sebelum dimulainya perjalanan yang akan ditempuh oleh konsumen, sehingga konsumen dapat memperkirakan dan menyiapkan anggaran terlebih dahulu.

Seperti yang dijelaskan sebelumnya, untuk memprediksi harga perjalanan jasa transportasi *online* uber dan lyft, dapat dilakukan dengan menggunakan *machine learning* dengan referensi dataset riwayat perjalanan konsumen. *Machine learning* (ML) dapat diartikan sebagai program komputer dan algoritma matematika yang diperoleh dengan cara mempelajari data dan membuat prediksi di masa depan. Dalam *machine learning*, untuk memperoleh kecerdasan dapat dilakukan dengan dua hal yaitu *training data*/pelatihan data dan *testing data*/pengujian data. Dataset riwayat perjalanan konsumen uber dan lyft, akan di *training* dan di *testing* sehingga menghasilkan *output*/harga prediksi yang mendekati harga aslinya. *Output* yang dikeluarkan oleh *machine learning* akan menjadi acuan konsumen untuk membandingkan harga antara dua jasa transportasi *online* tersebut dan memperkirakan anggaran yang harus dikeluarkan oleh konsumen.

## **1.2 Rumusan Masalah**

1. Bagaimana nilai *mean absolute error*, *mean square error*, dan *explained variance score* dari pemodelan *linear regression* dan *random forest*?
2. Pemodelan manakah yang cocok digunakan untuk memprediksi harga taksi *online* uber dan lyft?

## **1.3 Tujuan Kegiatan**

1. Mengetahui nilai *mean absolute error*, *mean square error*, dan *explained variance score* dari pemodelan *linear regression* dan *random forest*.
2. Mengetahui pemodelan yang cocok digunakan untuk memprediksi harga taksi *online* uber dan lyft.

#### **1.4 Manfaat Kegiatan**

Informasi yang terdapat di dalam dataset [Uber and Lyft Dataset Boston, MA | Kaggle](#) [1] dapat digunakan untuk memprediksi harga kedua taksi *online* tersebut. Bagi konsumen diharapkan dari penelitian ini dapat membandingkan harga, memilih jasa transportasi yang diinginkan, dan mengatur anggaran dengan menggunakan program prediksi harga transportasi *online* uber dan lyft.