Machine Learning:

The general machine learning framework is outlined below:

1. **Prediction Engineering:** State the business need, translate into a machine learning problem, and generate labeled examples from a dataset.
2. **Feature Engineering:** Extract predictor variables — features — from the raw data for each of the labels.
3. **Modeling:** Train a machine learning model on the features, tune for the business need, and validate predictions before deploying to new data.

Type of Machine Learning:

The dataset comes under Supervised Learning since it has labelled data. Basically supervised learning is a learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples(data) so that supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data. Supervised learning classified into two categories of algorithms:

● **Classification**: A classification problem is when the output variable is a category, such as “Red” or “blue” or “disease” and “no disease”.

● **Regression**: A regression problem is when the output variable is a real value, such as “dollars” or “weight”. In this case our Supervised Learning is a Regression algorithm where we have to predict the Prices of the Airline tickets.

The Process of Prediction Engineering:

Prediction engineering requires guidance both from the business viewpoint to figure out

the right problem to solve as well as from the data scientist to determine how to translate the business needs into a machine learning problem. The inputs to prediction engineering are the *parameters* which define the prediction problem for the business requirement , and the historical dataset for finding examples of what we want to predict.

Feature Engineering:

Feature engineering, the second step in the machine learning pipeline, takes in the label times from the first step — prediction engineering — and a raw dataset that needs to be refined. Feature engineering means building features for each label while filtering the data used for the feature based on the label's cutoff time to make valid features. These features and labels are then passed to modeling where they will be used for training a machine learning algorithm.

The Machine Learning Modeling Process:

The outputs of prediction and feature engineering are a set of label times, historical examples of what we want to predict, and features, predictor variables used to train a model to predict the label. The process of modeling means training a machine learning algorithm to predict the labels from the features, tuning it for the business needs, and validating it on holdout data.

Machine Learning Algorithms:

The Algorithms used to build a model and predict the outcomes are:

1. Linear Regression: Since it is a regression problem the first model which we use to fit and predict will be Linear Regression. The main features which we can relate using Linear Regression are duration and Price in the dataset. Duration

being an independent feature and Price is a dependent feature.

2. We will regularize the above regression using Lasso, Ridge, ElasticNet along with GridsearchCV for hyperparameter tuning for regularization techniques

3. Random Forest Regressor: A random forest is a meta estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.