|  |  |  |  |
| --- | --- | --- | --- |
| **ML Type** | **What the Model Learns** | **Travel Example** | **Typical Algorithms (famously used)** |
| **Supervised Learning** | The model learns from labeled data (data with correct answers). | Predict a passenger’s *satisfaction score* from features like age, seat class, and delay minutes | Linear/Logistic Regression, Random Forest, XGBoost |
| **Unsupervised Learning** | The model finds patterns or groups in data without any labels. | Discover *traveler segments* (e.g., “business commuters”, “vacation families”) by clustering booking behavior | K‑Means, DBSCAN, Hierarchical Clustering |
| **Semi‑Supervised Learning** | Combines a small set of labeled data with a large set of unlabeled data | Classify millions of *trip‑review sentences* as positive/negative using a few thousand hand‑labeled reviews | Self‑Training, Label Propagation |
| **Reinforcement Learning** | A sequence of actions that maximizes long‑term reward through trial & error | An airline pricing engine that *adjusts ticket prices* in real time to maximize revenue while keeping load factors high | Q‑Learning, Deep Q‑Networks (DQN), Policy Gradient |
| **Self‑Supervised Learning** | Learns useful representations by predicting part of the input from other parts | Train a language model on *billions of itineraries* to autocomplete travel queries | Contrastive Learning, Masked Language Modeling |

|  |  |  |
| --- | --- | --- |
| **Use Case** | **Input Features** | **Output (Label)** |
| Predict ticket price | Destination, Season, Booking time, Flight class | Ticket price (Regression) |
| Predict likelihood of flight cancellation | Weather, Departure time, Airline, Route | Yes / No (Classification) |
| Predict customer satisfaction | Age, Flight class, Delay time, Wi-Fi, Service score | Satisfied / Not satisfied (Classification) |

**Examples:**

**Classification –** Predict categories or labels

Will the traveler be satisfied? (Yes/No)

**Regression** – Predicts continuous values What will the ticket price be? ($350, $420)

**Feature Engineering?**  
Feature engineering means changing or creating data so that a computer can understand it better and make smarter decisions.

Feature engineering means selecting, transforming or creating new features in the dataset to help the model learn better.

**Important?**

Improves accuracy - Help models learn meaningful patters

Reduce Noise - Reduces irrelevant or misleading data

Help the model to understand real-world problems

**Types of features:**

Numerical - Age or Salary - Used in Calculations.

- Distance, age, Delay in Minutes

Categorical - Gender or Country →used for grouping or classification

- class (Business/Economy)

Time based - Order date or login time helps find trends and patens

- Booking Date or Departure time - Weekday, Hour, or Is - weekend

Scaling → Numerical

Encoding → Categorical

Extracting components like day or hour from time features.

**Feature selection** → Removes irrelevant columns

**feature creation** → Discover new insights → combining Quantity & Price to create 'Total purchase’

- Build New columns

Creating Advance Booking days by subtracting Booking date From Departure date.

**Feature Selection:**

Choose only Relevant useful features

Techniques:

Correlation matrix (for numerical)

Chi- Square test (Categorical)

Feature importance from tree models

Domain Knowledge (what matters in real-world context)

**Feature Transformation:**

Normalization/ Scaling - Numerical

Encoding - Label or One-hot for categorical

Datetime breakdown - Extract Day, hour, month from timestamps

Binning - Age groups, Distance groups

Ex: Convert Departure Time into Morning /Afternoon/Evening

**Feature encoding:**

Different ML algorithms need features in numeric format

one Hot Encoding - Creates binary Columns for categories

Label Encoding - Converts categories to integer codes

Frequency Encoding - Replaces category with how often it appears.

EX: Encode "Airline" using One Hot so each airline becomes its own column

**Iterative process:** Feature engineering is not only one time It happens iteratively using:

EDA (Exploratory Data Analysis)

Model & Training & Evaluation

Model Tuning

**Select Features**

1. **Filter Methods** → Based on statistical Methods

→ Correlation Coefficient (Pearson's r)

→ Chi-Square Test

→ Mutual Information (Measures dependency between variables (numerical or categorical)

2. **Wrapper Methods** → Evaluate Subsets of features

→ Forward Selection - Start with no features, add one by one

→ Backward Elimination - Start with features, remove one by one

→ Recursive Feature Elimination (RFE) - Iterative train model, remove least important feature

3. **Embedded Methods** → Feature selection happens during model training.

Lasso Regression (L1 Regularization) → Shrinks less important feature coefficients to zero

Tree based Models (Random Forest, XGBoost) → provide feature importance scores

Elastic Net → Combines L1 & L2 regulation regularization

**Ex:** Predict flight delay

Features: Weather, Airline, Day of week, Flight Distance, Airport Traffic, Wi-fi Rating, etc.

→ Use Correlation to drop features unrelated to delays (ex: Wi-Fi Rating might have no effect)

→ Use Random Forest to get importance Scores and Keep top 10 features

→ Use RFE to further refine features for best predictive power.

|  |  |  |
| --- | --- | --- |
| **Topic** | **Description (One Word)** | **Examples / Notes** |
| What is Feature Engineering? | Creating features | Data transformation for model learning |
| Why is it Important? | Accuracy | Improves model prediction, reduces noise |
| Types of Features | Numerical, Categorical, Time | Age, Gender, Booking Date |
| Feature Transformation | Scaling, Encoding, Extraction | Normalize, One-hot, Extract day/hour |
| Feature Encoding | Numeric Conversion | One-Hot, Label, Frequency |
| Feature Creation | New Features | Total Purchase, Booking Days |
| Feature Selection | Relevant Features | Correlation, Chi-square, Feature importance |
| Feature Selection Methods | Filter, Wrapper, Embedded | Pearson’s r, RFE(Recursive Feature Elimination), Lasso |
| Iterative Process | Repeat | EDA → Model → Select → Tune |
| Example Use Case | Flight Delay | Drop Wi-Fi, Random Forest, RFE |