Introduction To
Azure Machine Learning

Expected Learning Outcomes

Azure Machine Learning

→ **Supervised Machine Learning**

→ **Machine Learning Algorithms in Azure ML**

→ **Workflow for Supervised Machine Learning**

→ **Algorithm Selection**
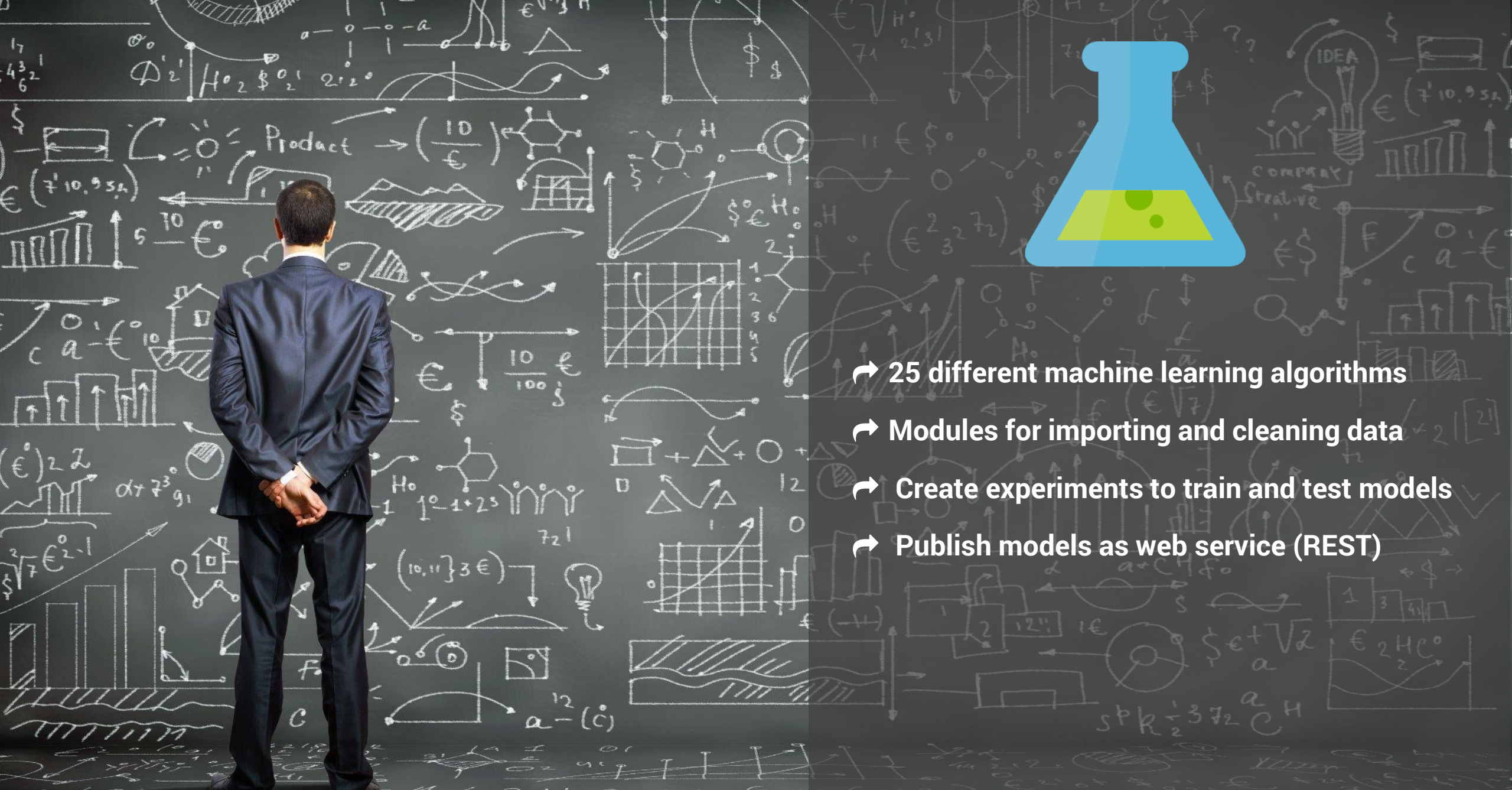
Supervised

Machine Learning

- **Unlabeled Data**
- **Labeled Data**

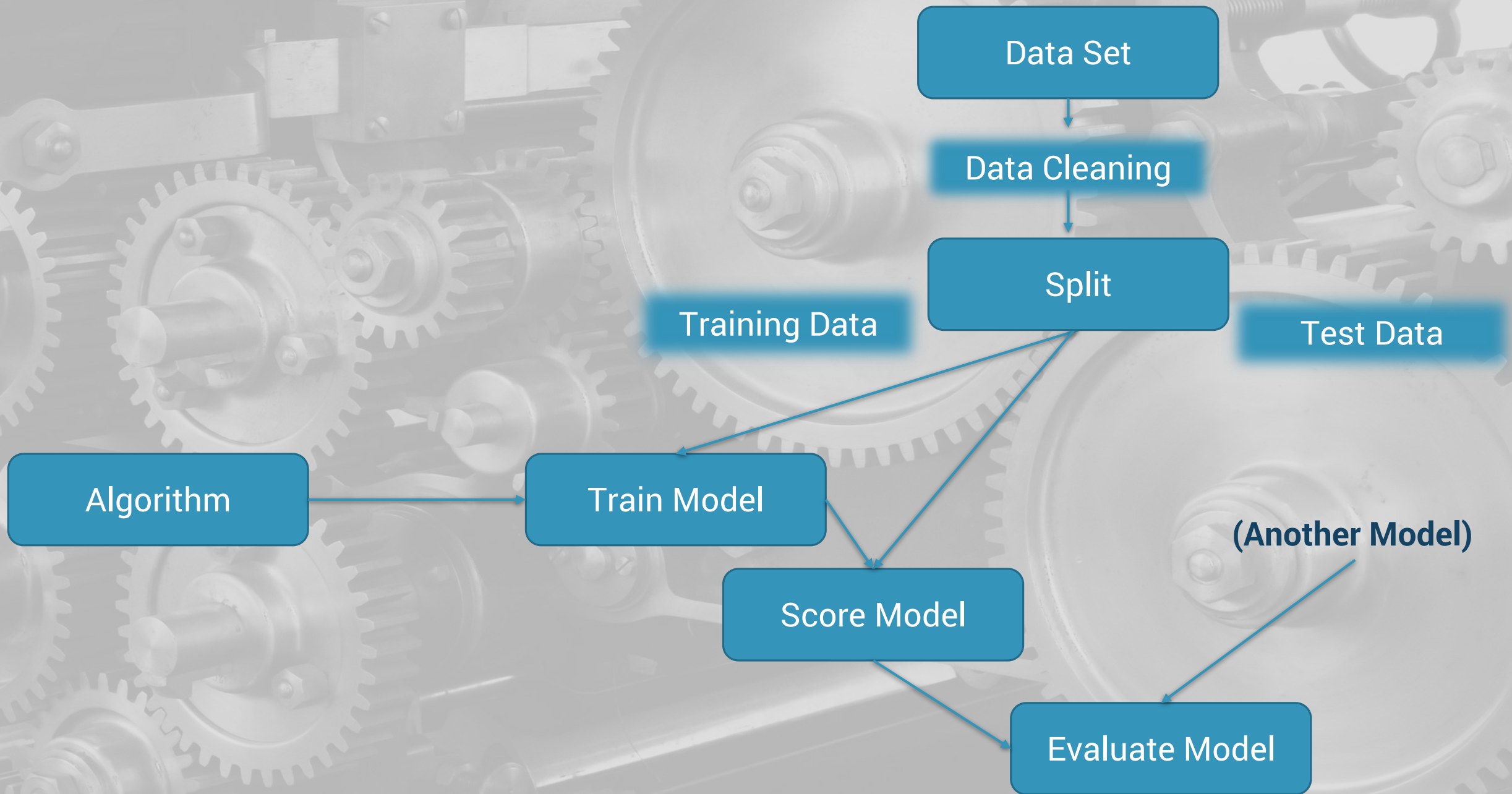| Square Footage	Zip Code	Price
2000	98052	5000 $
3000	75284	4000 $
3500	68164	5500 $
2000	45465	5400 $
7520	98645	10000 $

➡️ **Classification - predicting a discrete category**

➡️ **Regression - predicting a value**

➡️ **Anomaly Detection - identifying unusual data**

➡️ **Clustering - grouping similar data together**

- 25 different machine learning algorithms
- Modules for importing and cleaning data
- Create experiments to train and test models
- Publish models as web service (REST)

- → **Sample data**
- → **Upload dataset from local machine**
- → **Enter data manually**
- → **Import data from external website**

→Modules

→Datasets

→Experiments

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.

ANOMALY DETECTION

One-class SVM — >100 features, aggressive boundary

PCA-based anomaly detection — Fast training

CLUSTERING

K-means

Discovering structure

Finding unusual data points

MULTI-CLASS CLASSIFICATION

Fast training, linear model — **Multiclass logistic regression**

Accuracy, long training times — **Multiclass neural network**

Accuracy, fast training — **Multiclass decision forest**

Accuracy, small memory footprint — **Multiclass decision jungle**

Depends on the two-class classifier, see notes below — **One-v-all multiclass**

Three or more

Predicting categories

START

REGRESSION

Ordinal regression — Data in rank ordered categories

Poisson regression — Predicting event counts

Fast forest quantile regression — Predicting a distribution

Linear regression — Fast training, linear model

Bayesian linear regression — Linear model, small data sets

Neural network regression — Accuracy, long training time

Decision forest regression — Accuracy, fast training

Boosted decision tree regression — Accuracy, fast training, large memory footprint

Predicting values

Two

TWO-CLASS CLASSIFICATION

Two-class SVM — >100 features, linear model

Two-class averaged perceptron — Fast training, linear model

Two-class logistic regression — Fast training, linear model

Two-class Bayes point machine — Fast training, linear model

Accuracy, fast training — **Two-class decision forest**

Accuracy, fast training, large memory footprint — **Two-class boosted decision tree**

Accuracy, small memory footprint — **Two-class decision jungle**

>100 features — **Two-class locally deep SVM**

Accuracy, long training times — **Two-class neural network**