Predictive & Descriptive Data Analytics

Predictive (Data)Analytics

Predictive analytics use-cases

- Predicting whether an email is spam or not

- Predicting the stock price, weather condition

- Predicting whether customer will be defaulter or not

- Classifying an image is animal or not?

- Suggest items to be bought in an online store
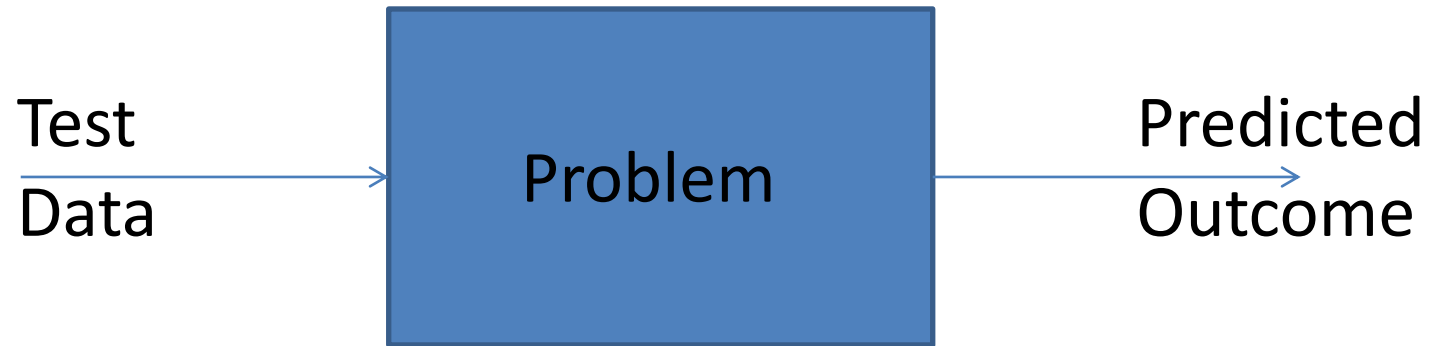
- Predict ratings of un-rated movies

Need more use-cases??

- https://www.kaggle.com/wiki/DataScienceUseCases
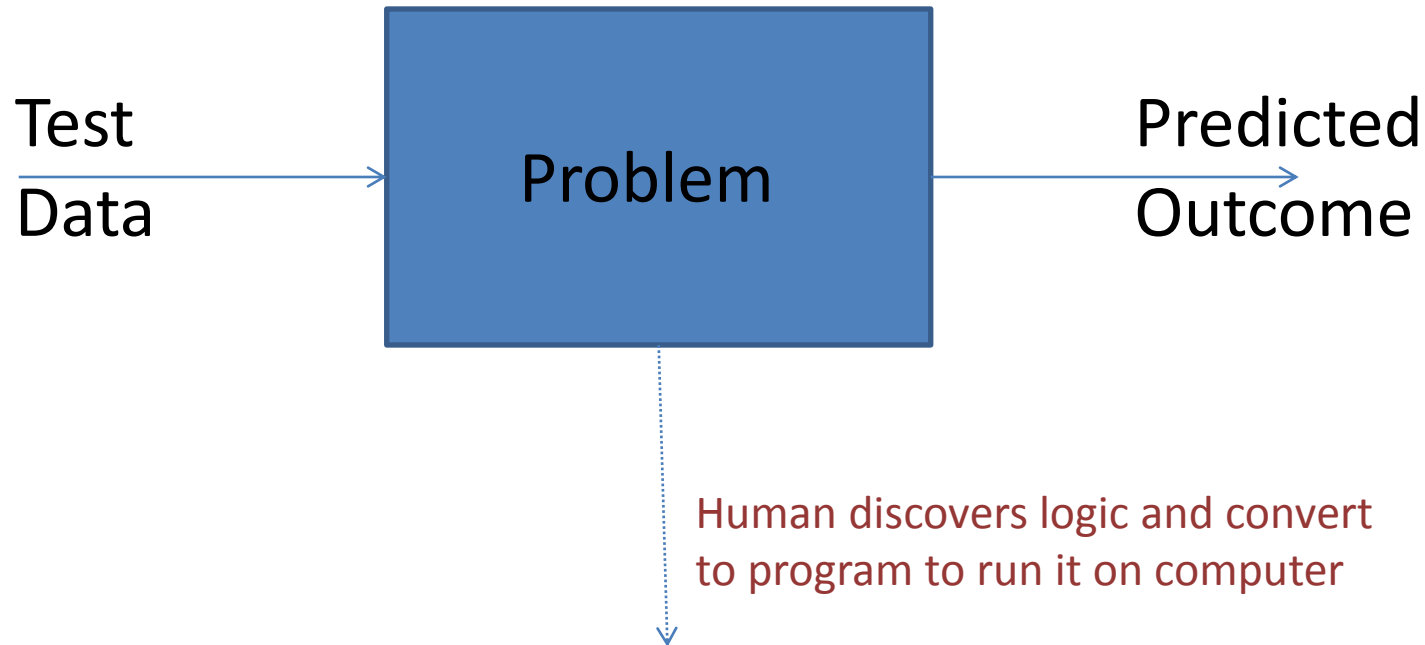
Predictive analytics: Naming Conventions

- Classification: Outcome is discrete value

- Regression: Outcome is continuous value

- Recommenders: Outcome is either rating or discrete value.

Predictive analytics problems

Test Data → | Problem | → Predicted Outcome

Traditional Approach

Test Data → Problem → Predicted Outcome

Human discovers logic and convert to program to run it on computer
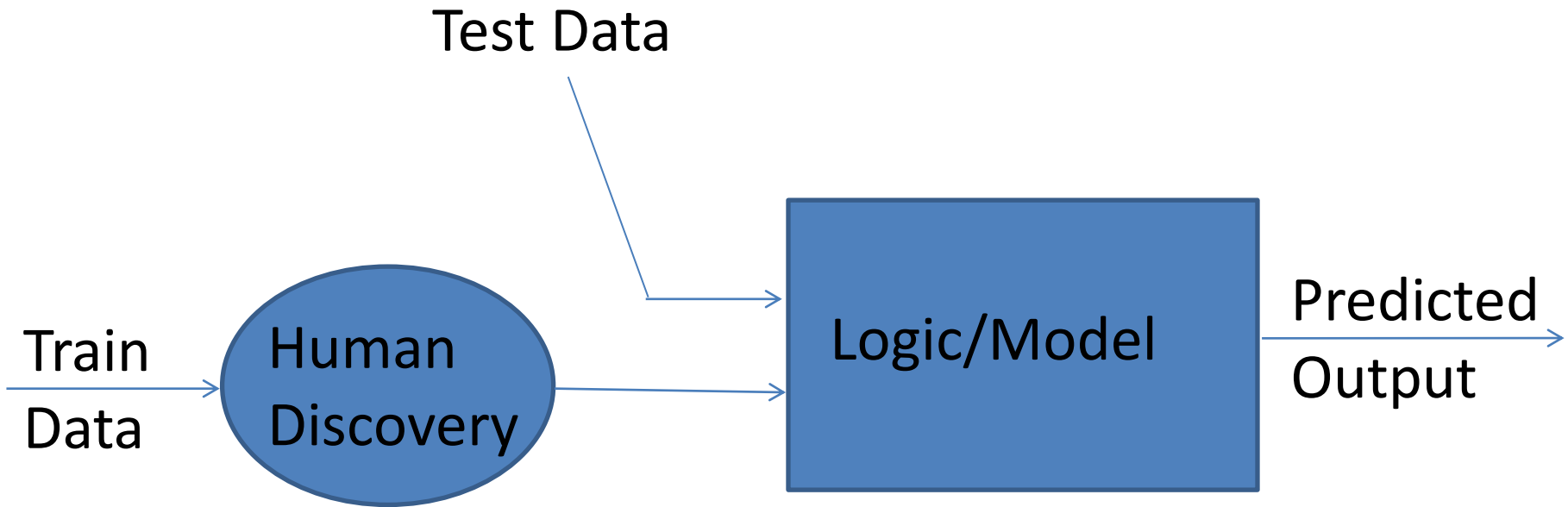
How does human discovers logic with testdata??

Not possible. We need history/train data for discovering logic

Traditional Approach

Test Data

Train Data → Human Discovery → Logic/Model → Predicted Output
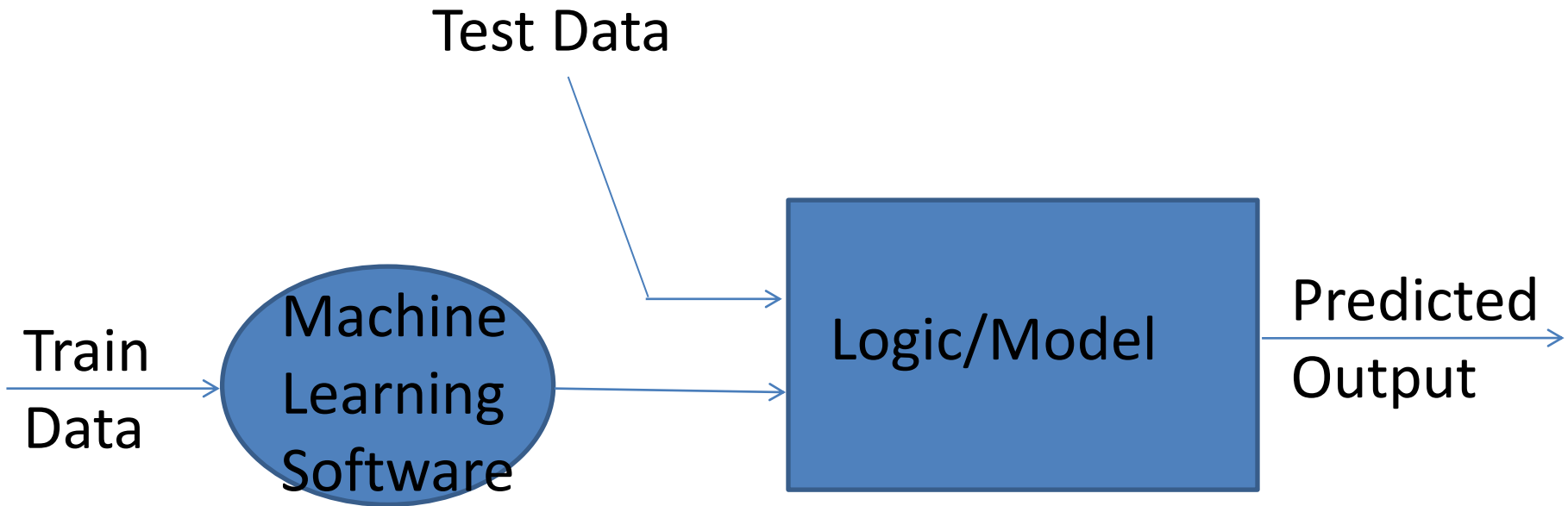
Issues:

- Human has to manually gothrough the train data to discover pattern or logic that can be applicable for test data.
- Hard-coded logic and Logic gets changed whenever new pattern is discovered.
- What if train data is big?

Machine Learning Approach

Test Data

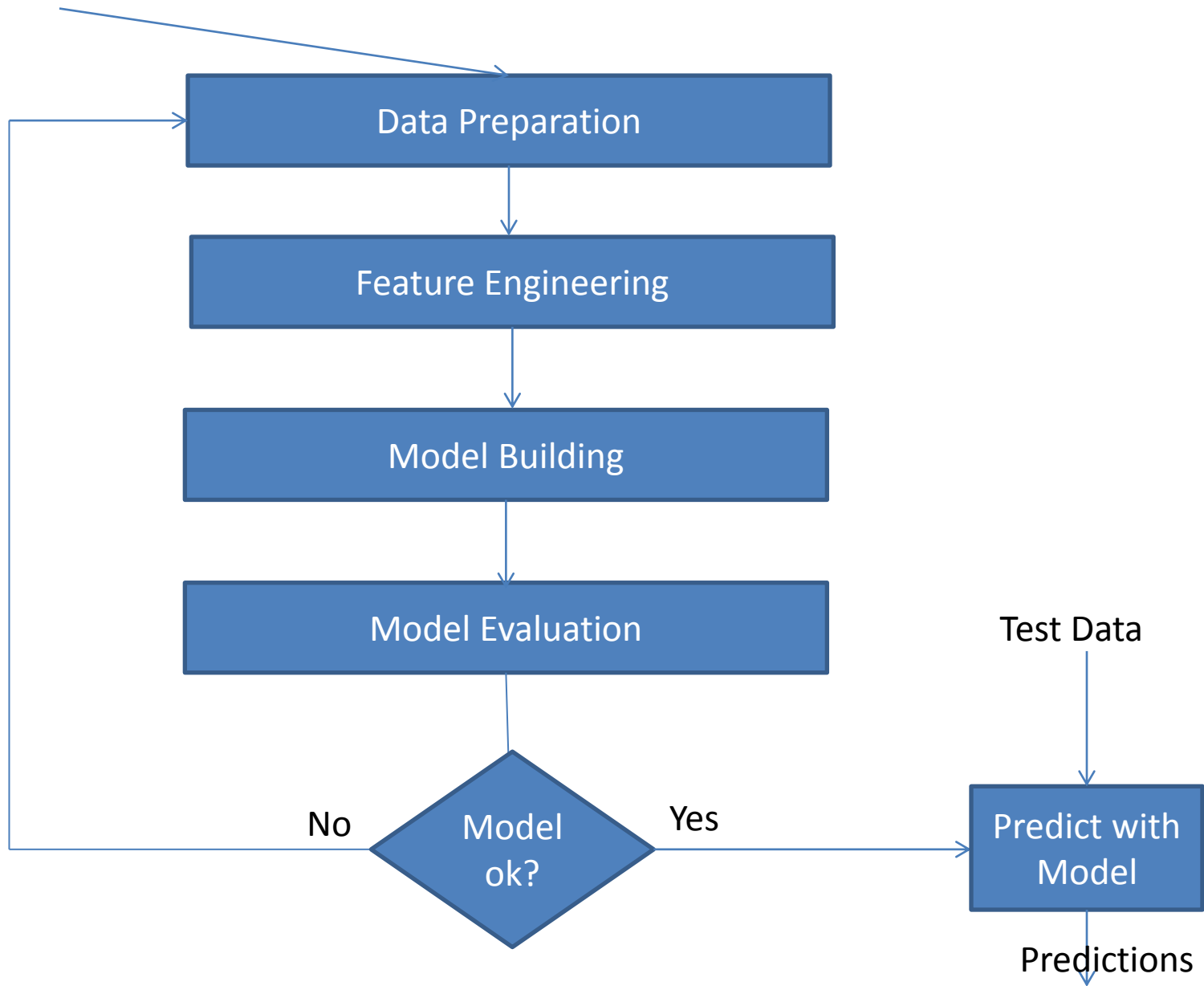Train Data → Machine Learning Software → Logic/Model → Predicted Output

- Any thing in machine happens via software. So, we have to write programs that discovers logic automatically. We call it as supervised ML programming.

- Different Machine learning approaches discover logic/pattern in different ways. We have to find what approach is best for given data.

Supervised ML Approaches

ML Approach	Form of Logic/Model	Predictive Category
Tree Approach	Decision Trees	Classification, Regression
Probabilistic Approach	Probabilities	Classification
Linear, Polynomial Equation Approach	Weights of each variable	Regression
Neural Network Approach	Weights of each neuron input	Classification, Regression
Ensemble Approach	Importance of each model and learning for each model	Classification, Regression
Support Vector Machine Approach	Support Vectors	Classification, Regression
Nearest Neighbor Approach	Remember input data	Classification, Regression, Recommenders
Matrix Factorization Approach	Matrices with latent factors	Recommenders

Structure of (Supervised)ML program

Train Data

```
             Data Preparation
                   │
                   ▼
           Feature Engineering
                   │
                   ▼
             Model Building
                   │
                   ▼
            Model Evaluation
                   │
                   ▼
    No  ◄──── Model ok? ────► Yes
```

Test Data

Predict with Model

Predictions

Descriptive Data Analytics

Descriptive analytics use-cases

- Customer Segmentation

- Reduce the number of features/dimensions

- Find the items that are frequently bought together in transactions

Descriptive analytics: Naming Conventions

- Clustering: Find the groups of related data

- Feature Reduction: Outcome is either rating or discrete value.

- Outlier Detection: Outcome is either rating or discrete value.

- Frequent Pattern Mining: Outcome is continuous value

Descriptive analytics problems

Input
Data → **Problem** → groups/
outliers
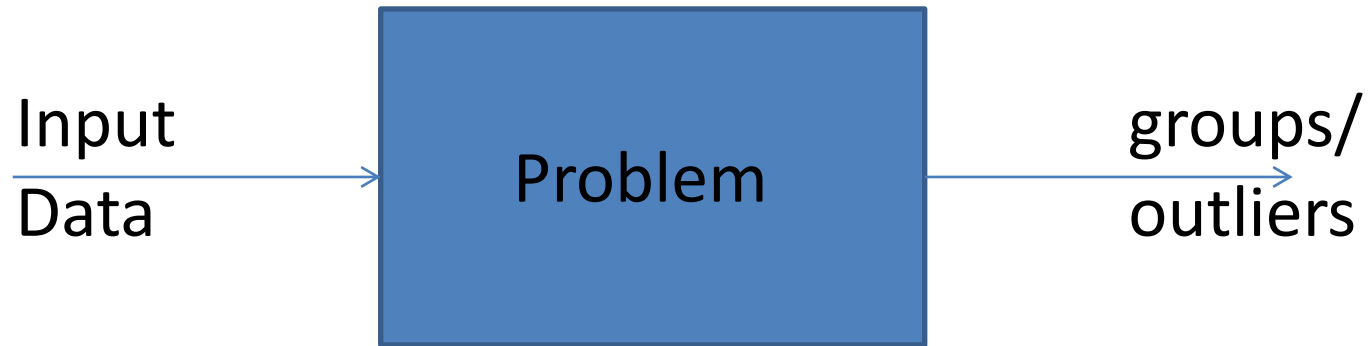
We don't have any train data as such here.

Traditional Approach

Input Data → (Human Discovery) → [Logic/Model] → groups/ outliers
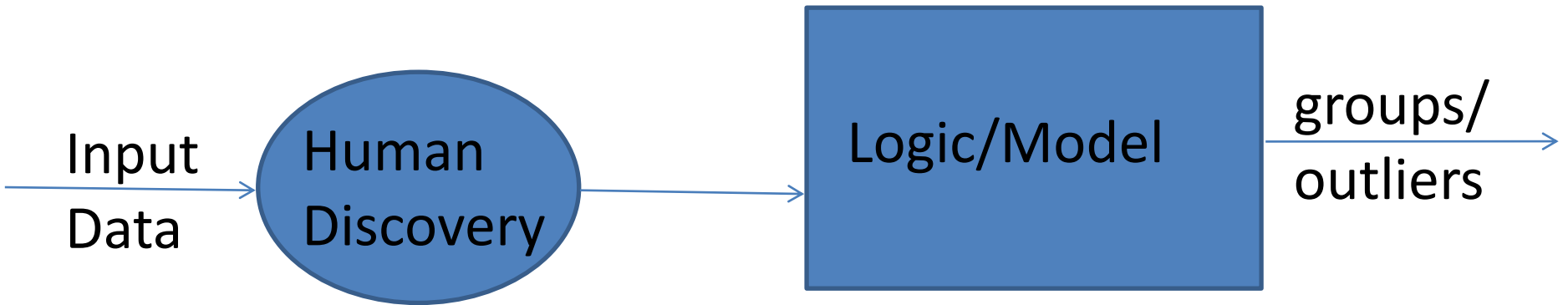
- Human discovery has same issues as before.

Machine Learning Approach
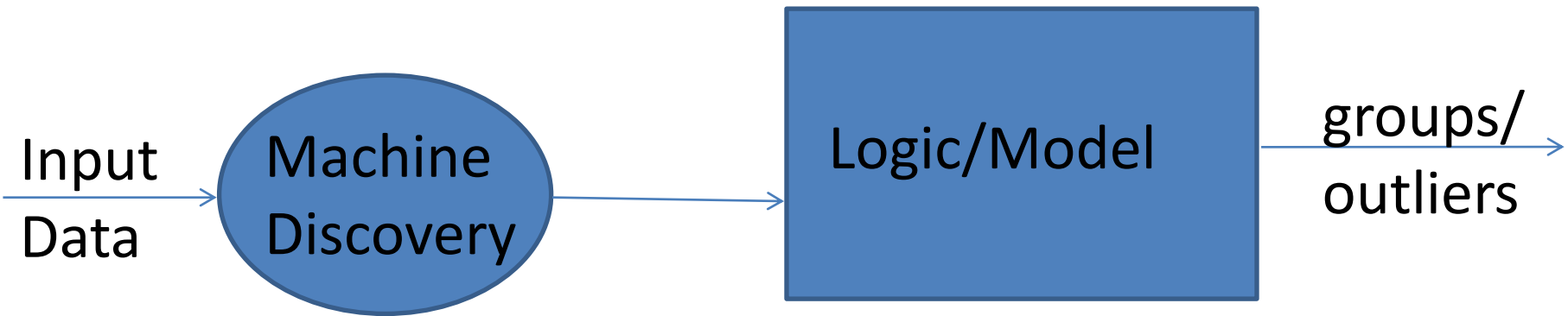
- Let machine learning program discover the pattern/logic from data automated way.

- Also called as Unsupervised ML programming.

Un-supervised ML Approaches

ML Approach	Form of Logic/Model	Predictive Category
Iterative Approach	Groups of data points	Clustering
Agglomerative Approach	Trees	Clustering
Variance based Approach	New basis	Feature Reduction
LOF Approach	Outlier Scores	Outlier Detection
Apriori Approach	Association Rules	Frequent Pattern Mining

Structure of (Un-supervised)ML program

Input Data

```
                    ┌──────────────────────────────┐
                    │      Data Preparation         │
                    └──────────────────────────────┘
                                   │
                                   ▼
                    ┌──────────────────────────────┐
                    │     Feature Engineering       │
                    └──────────────────────────────┘
                                   │
                                   ▼
                    ┌──────────────────────────────┐
                    │       Model Building          │
                    └──────────────────────────────┘
                                   │
                                   ▼
                    ┌──────────────────────────────┐
                    │      Model Evaluation         │
                    └──────────────────────────────┘
                                   │
                                   ▼
          No                 ◇ Model ok? ◇          Yes        ┌──────────────┐
                                                               │ Final Model  │
                                                               └──────────────┘
```