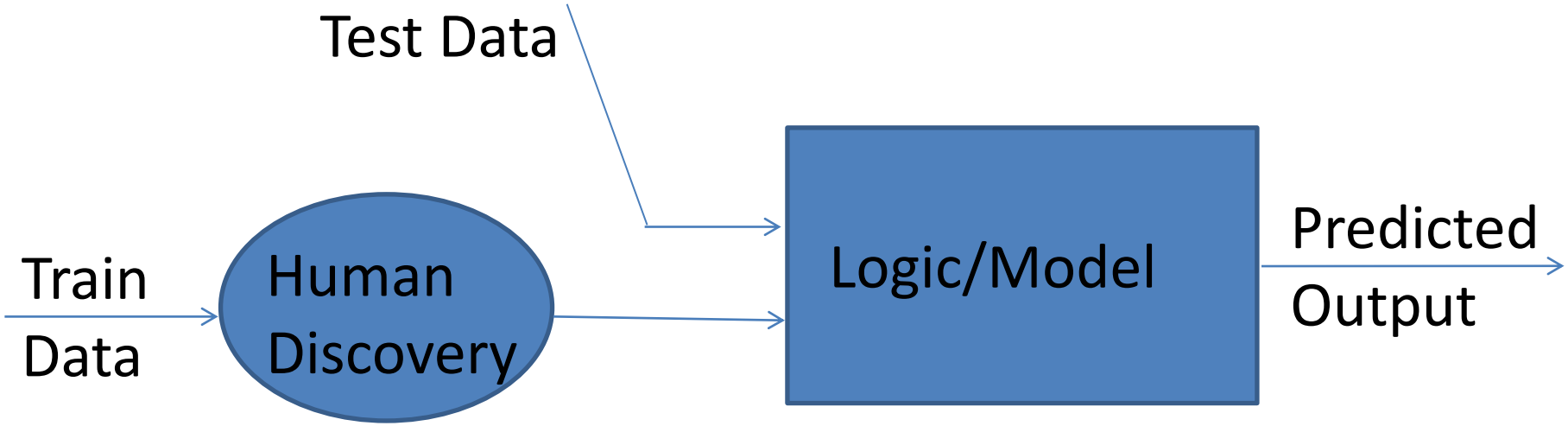
Machine Learning

How do you solve data analytics problems?

- . . . Lets start our journey with predictive analytics problems first

Titanic: Passenger survival prediction

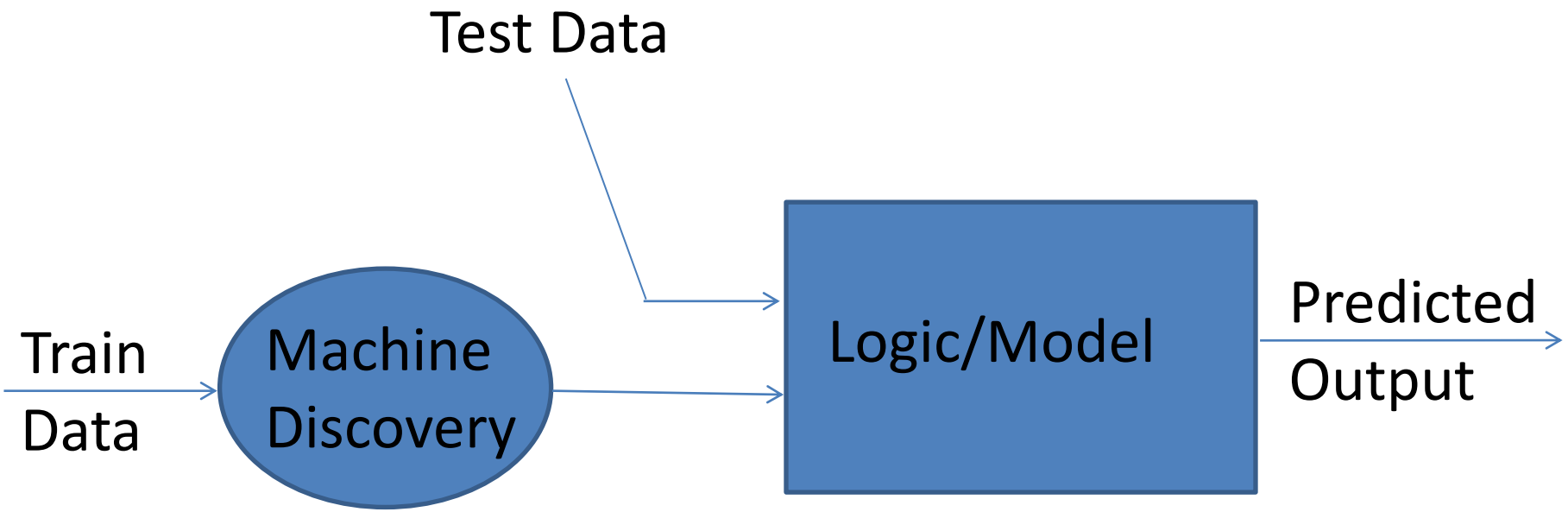
Traditional Approach: Human discovery



Issues:

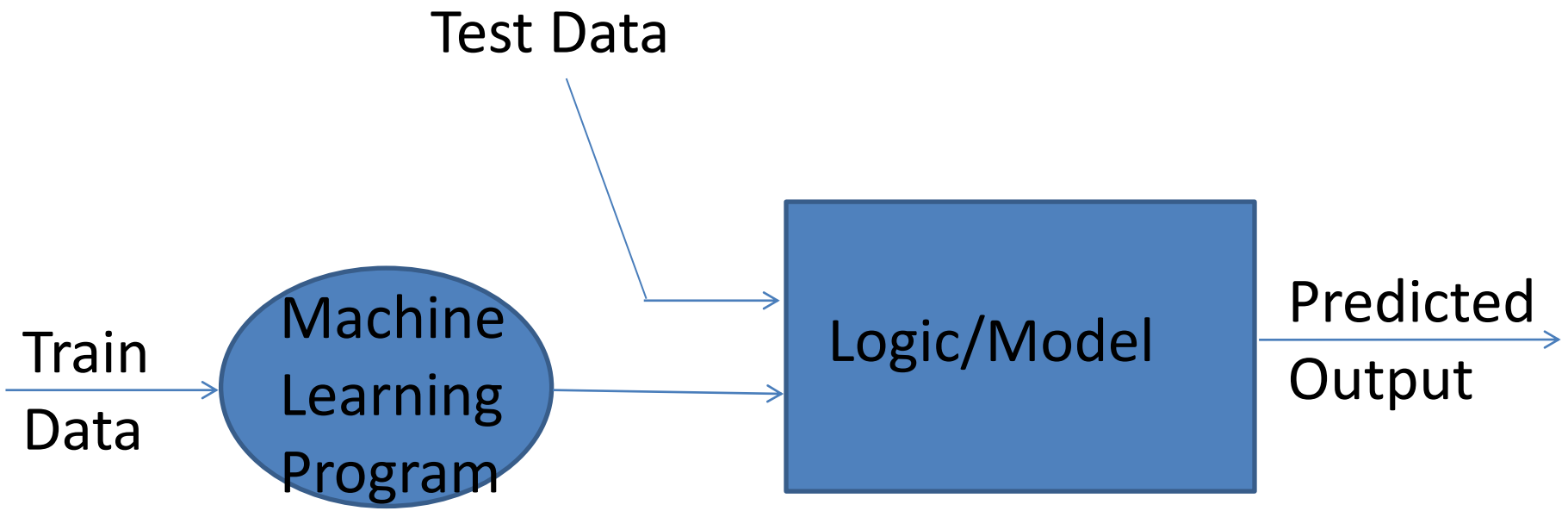
- Human has to manually go through the train data to discover pattern or logic that can be applicable for test data.
- Hard-coded logic and needs to rediscover and update the logic whenever train data gets changed.
- Is it practical to discover the logic/pattern if train data is big?

New Approach: Machine discovery



- If machine could discover logic/pattern, its great!!
  - No hardcoded logic
  - Machine can rediscover logic if train data gets changed
  - Machines can process big data sets(with more peers) as well.
- But How does machine discovers logic/pattern?

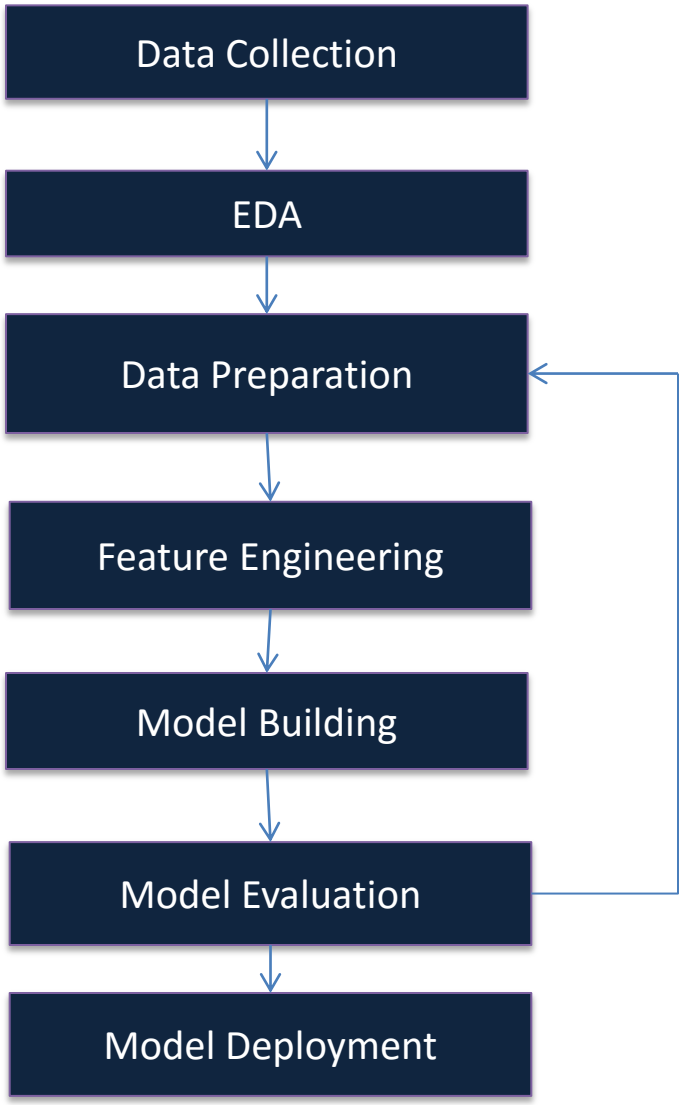
New Approach: Machine discovery



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

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

Data Analytics Lifecycle



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

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