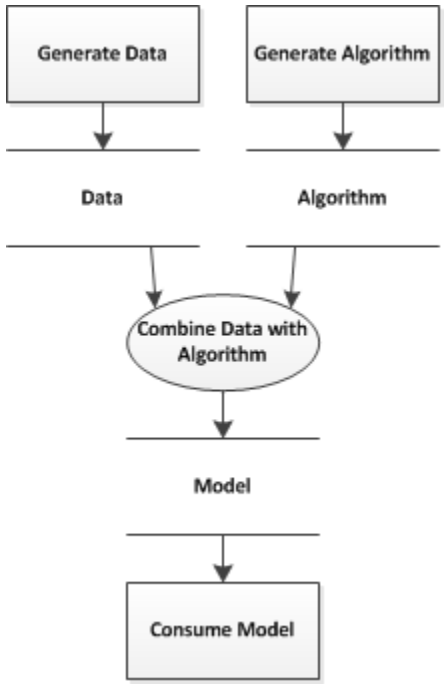
Data and Models in Supervised Learning

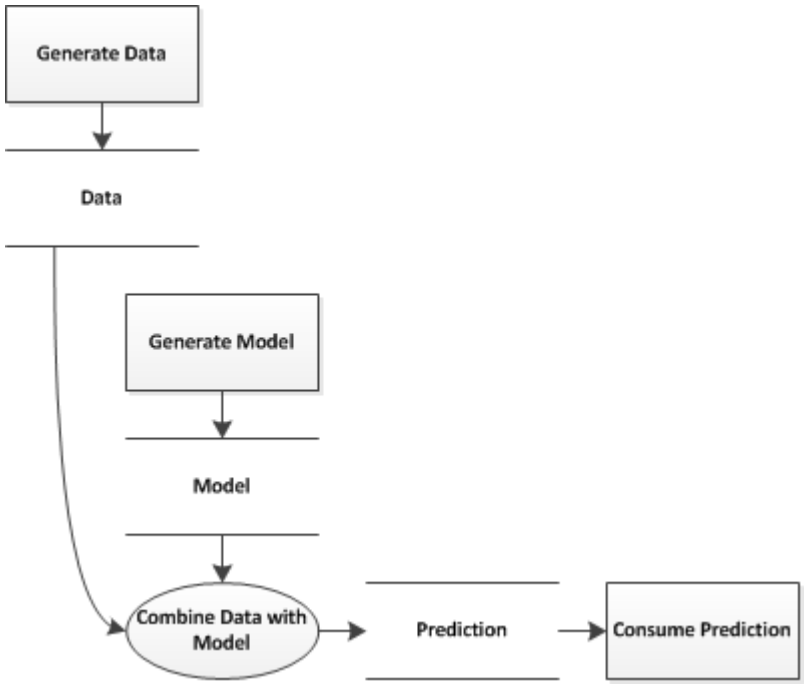
From Data to Predictions (0)

From Data to Predictions (1)



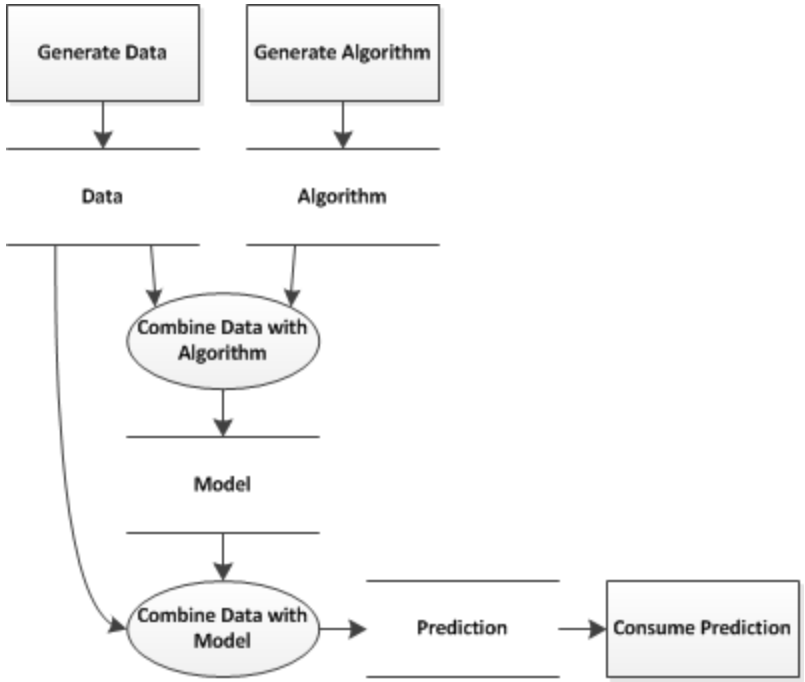
Data + Algorithm → Model

From Data to Predictions (2)



Model + Data → Prediction

From Data to Predictions (3)



Data + Algorithm → Model
Model + Data → Prediction

From Data to Predictions (4)

- Pseudo Assignments (Derivations):
  - Data + Algorithm → Model
  - Model + Data → Prediction

- Create Model from Algorithm and Data
  - Example Algorithm: Logistic Regression
    - Create Model: model <- glm(formula, data=trainSet, family="binomial")
- Predict from Model and Data
    - Predict: prediction <- predict(model, newdata=testSet, type="response")

Data + Algorithm → Model
Model + Data → Prediction

From Data to Predictions (5) Review

- A model or hypothesis is (best response)
  - a combination of test data and training data
  - a predictor based on data and algorithm
  - a falsification of a theory
  - a verified theory as long as the model was not falsified
- A model applied to new data leads to a (best response)
  - Prediction
  - Falsification / Verification
  - Hypothesis
  - errors
- A model applied to test data leads to a (best response)
  - Prediction
  - Falsification / Verification
  - Hypothesis
  - errors
- A hypothesis that cannot be tested
  - is a law if the data are consistent
  - is an untested hypothesis
  - is not a hypothesis
  - is a theory

Break

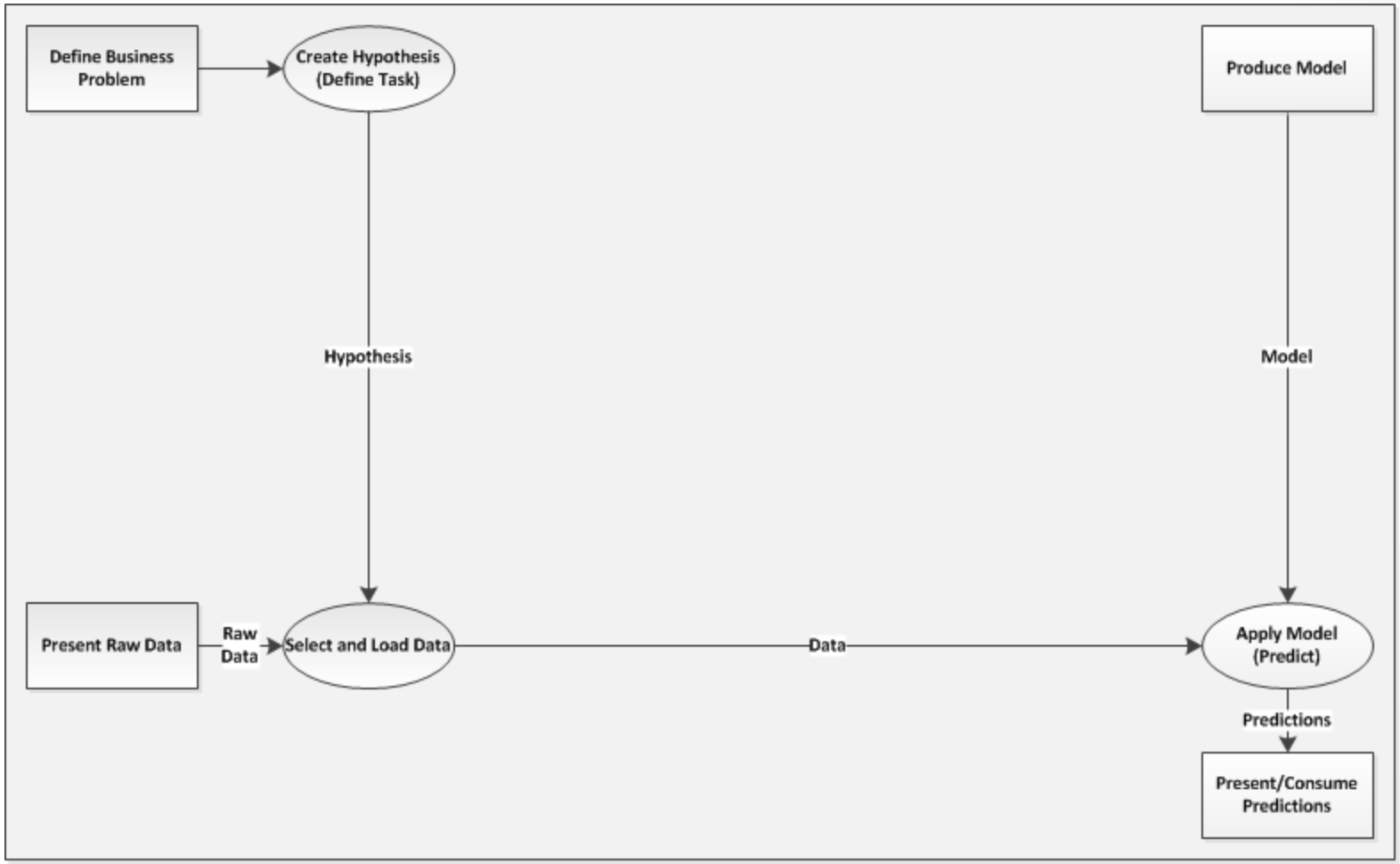
- Colbert on Predictive Analytics
  - http://www.colbertnation.com/the-colbert-report-videos/408981/february-22-2012/the-word---surrender-to-a-buyer-power

(0) DFD of Supervised Learning

(1) Model Acts on Data

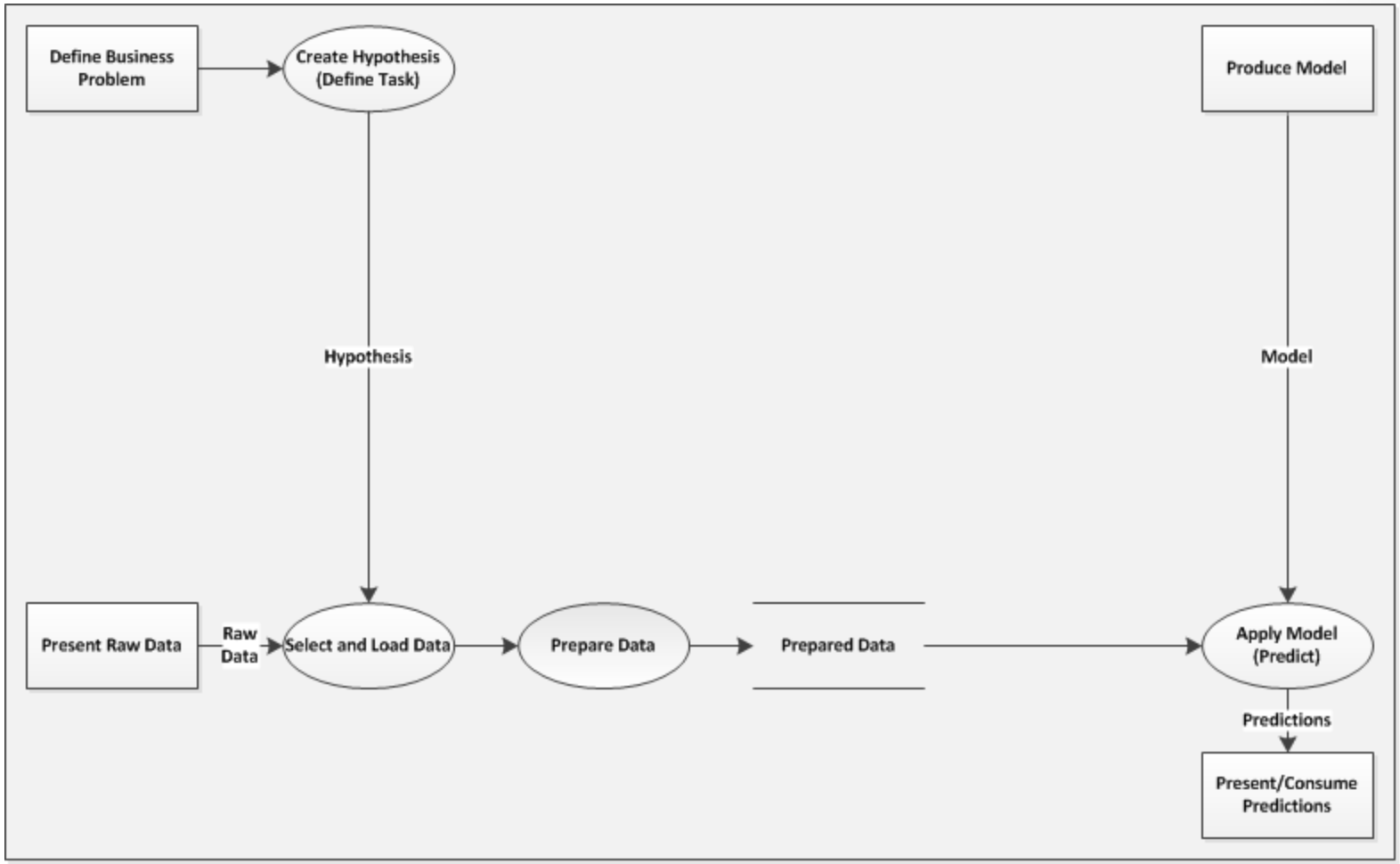
Model + Data → Prediction

(2) Data Selection Reflects Hypothesis / Business Problem



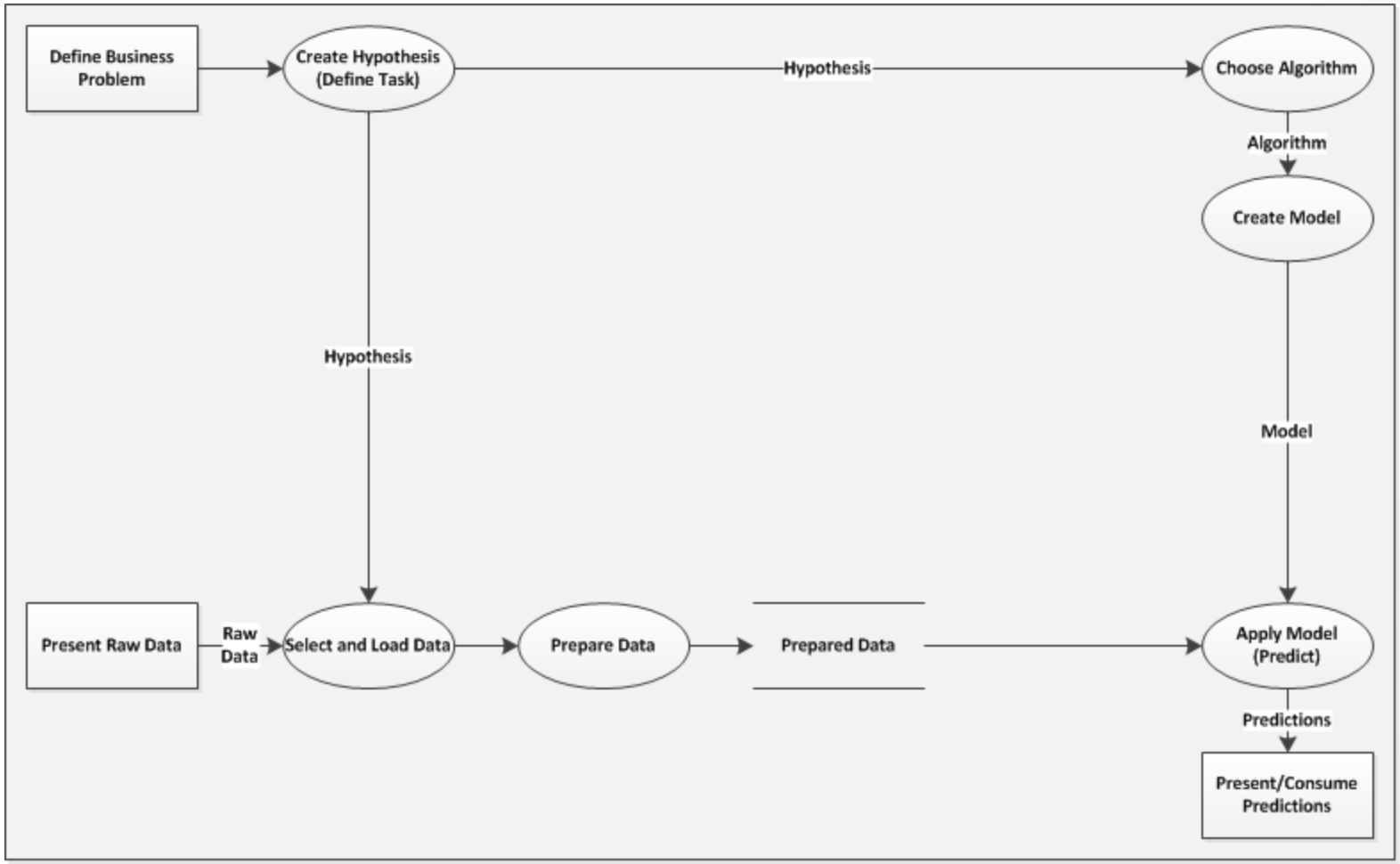
Hypothesis determines what data are loaded

(3) Data Needs Preparation



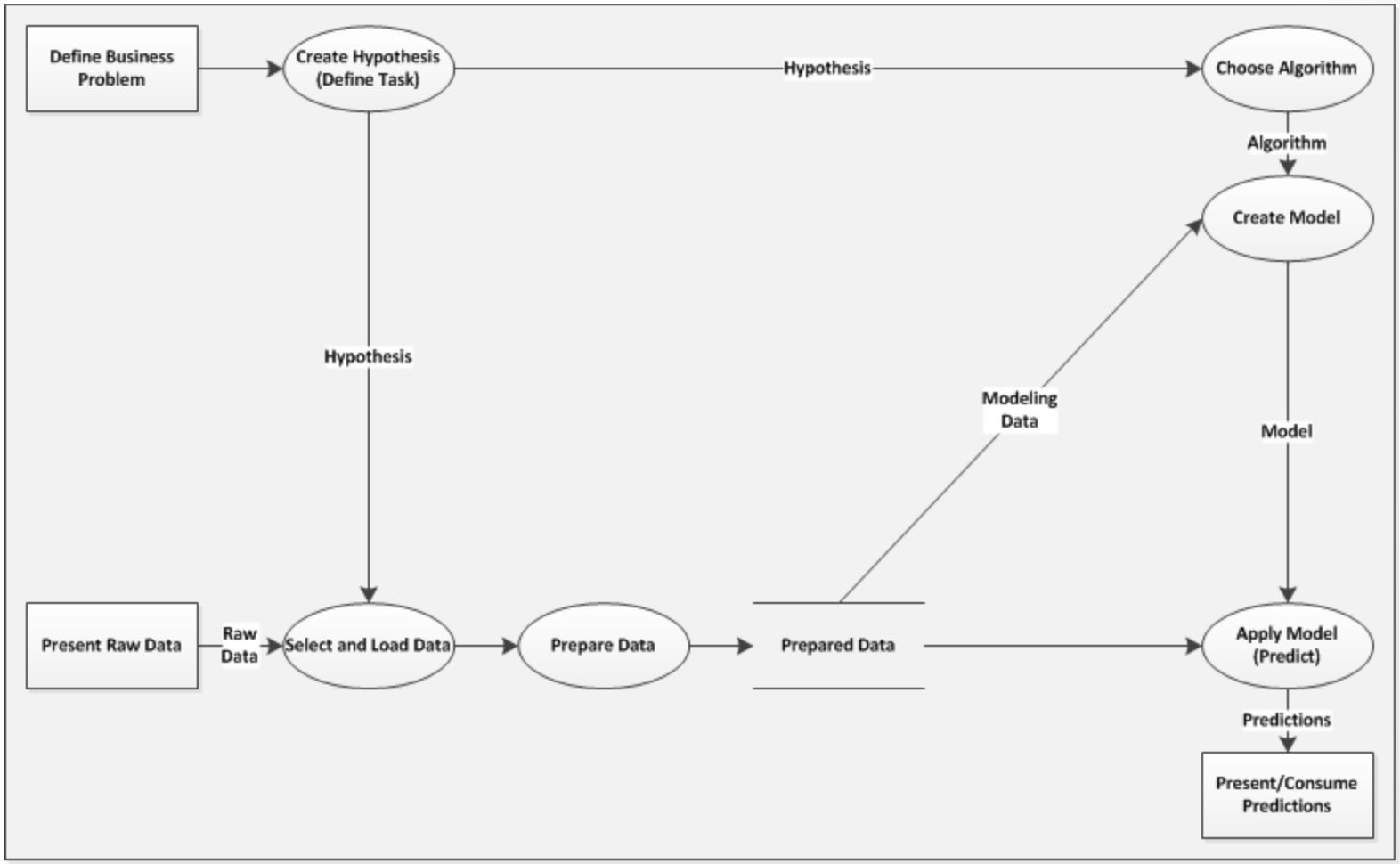
Data need to be prepared for use by a model.

(4) Model Creation Reflects Hypothesis / Business Problem



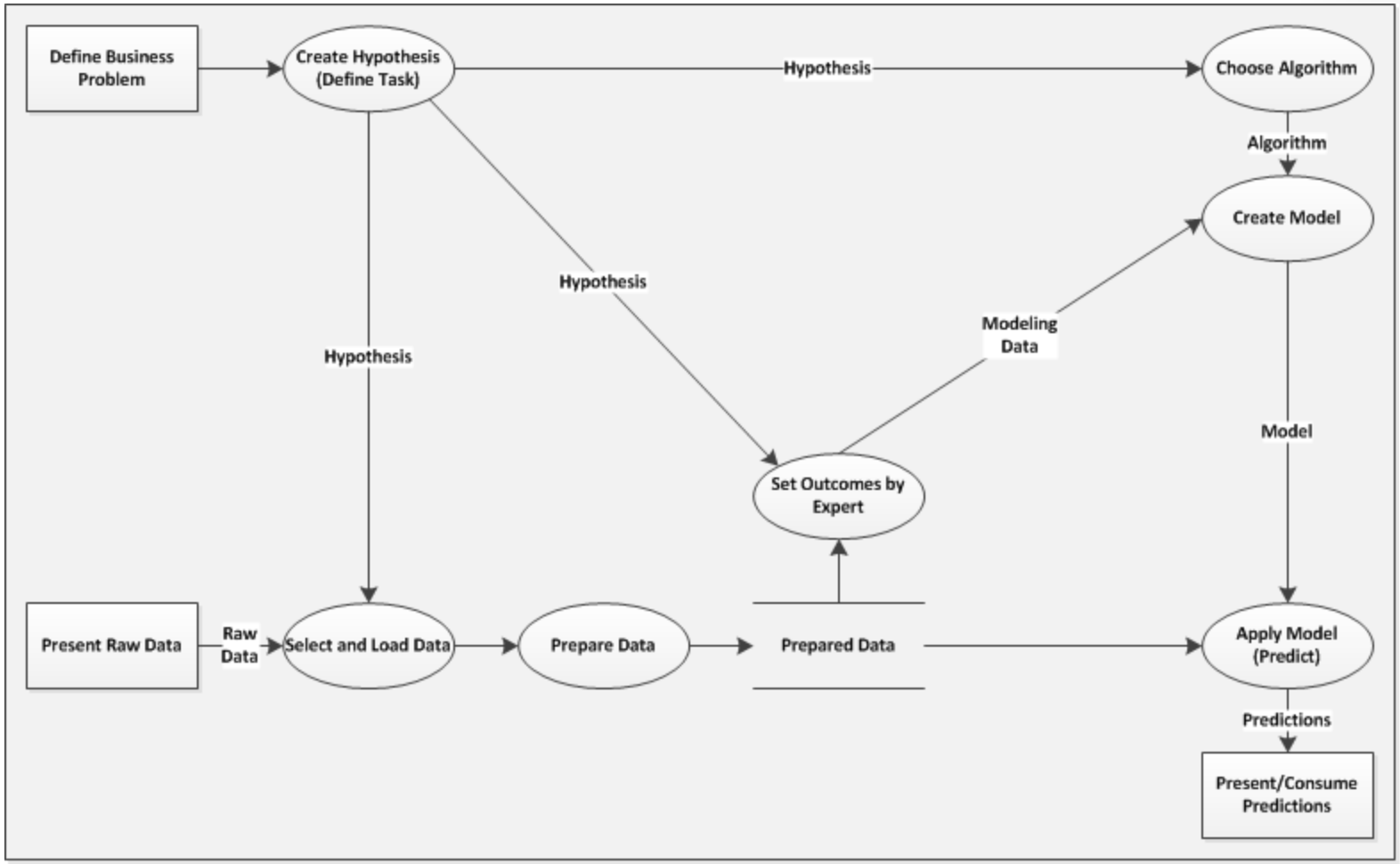
Hypothesis determines the choice of Algorithm.

(5) Model Creation needs Data



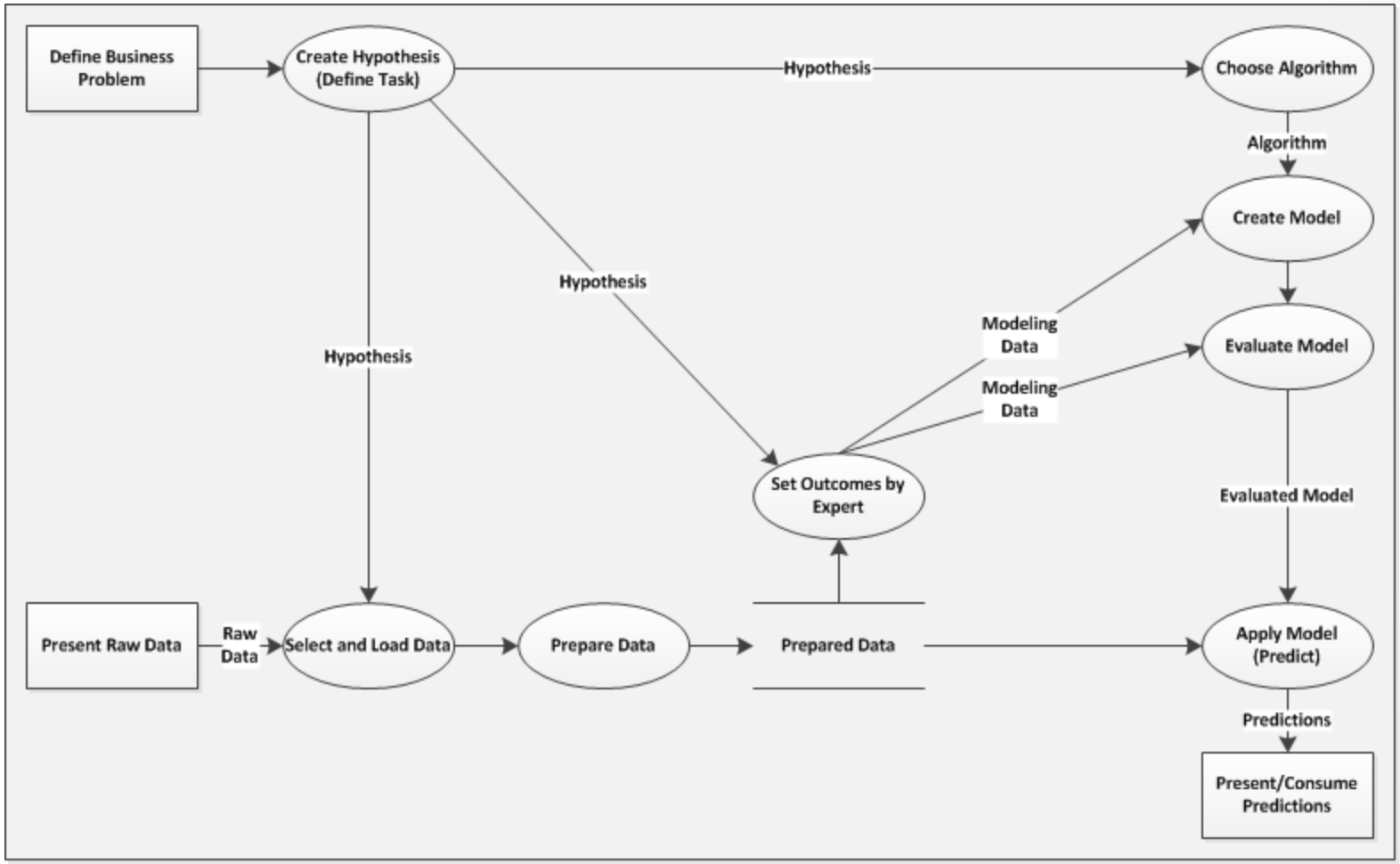
Data + Algorithm → Model

(6) Supervised Training needs Data Labeled with Outcomes



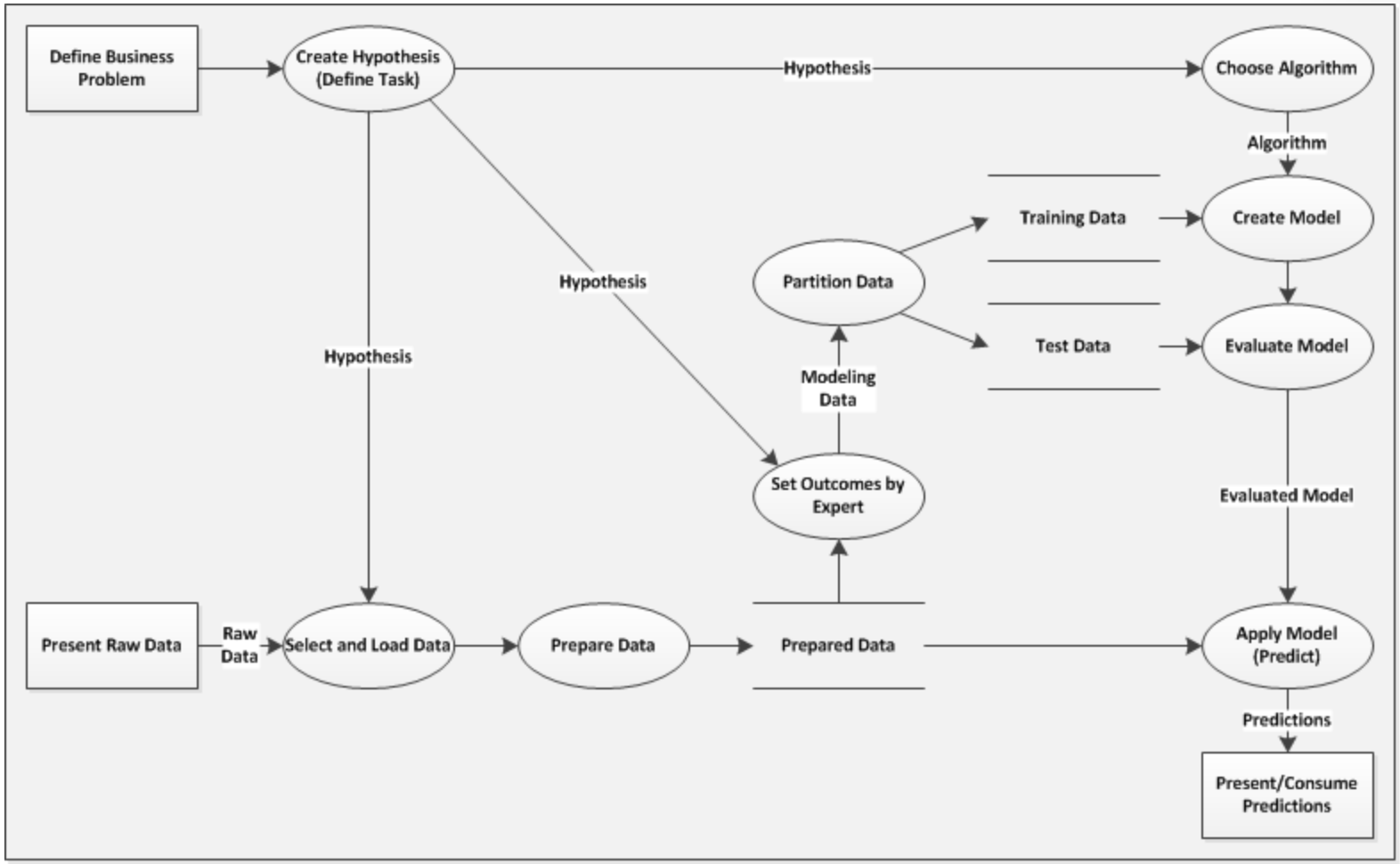
Supervised Learning requires expert labeling of data.

(7) Models need to be Evaluated



Do not trust predictions from an un-tested model!

(8) Creation & Evaluation of Model may not use same Data



Do not test a model using training data!

Data and Models in Supervised Learning