PROFESSIONAL & CONTINUING EDUCATION
UNIVERSITY of WASHINGTON

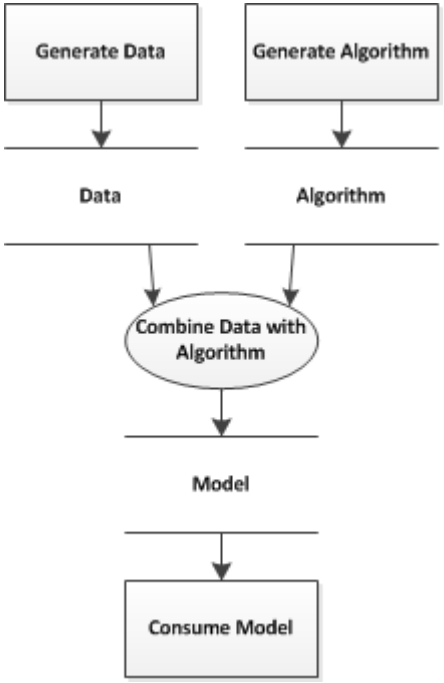
Data Flow in Supervised Learning

FROM DATA TO PREDICTIONS

> How do we get from data to predictions?

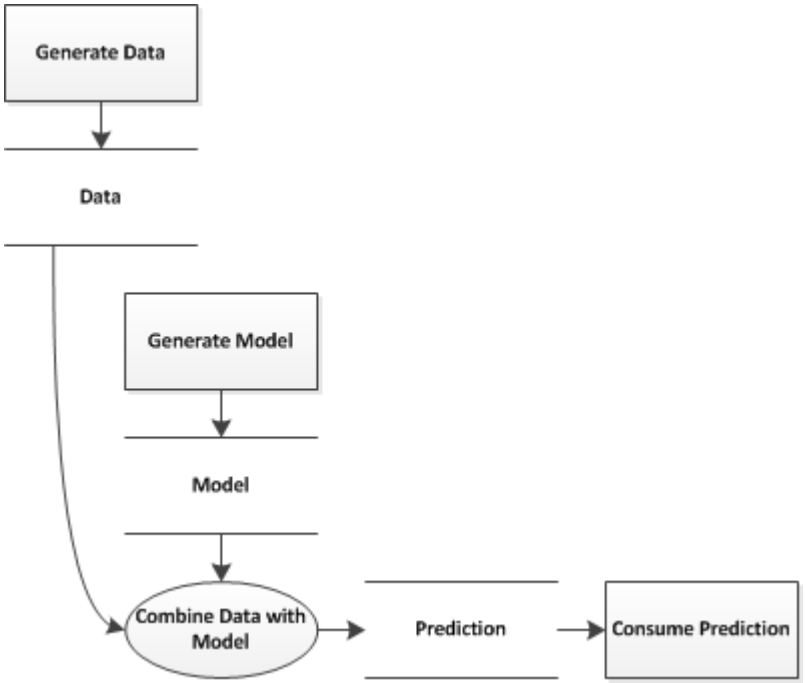
Data → ? → Predictions

FROM DATA TO PREDICTIONS



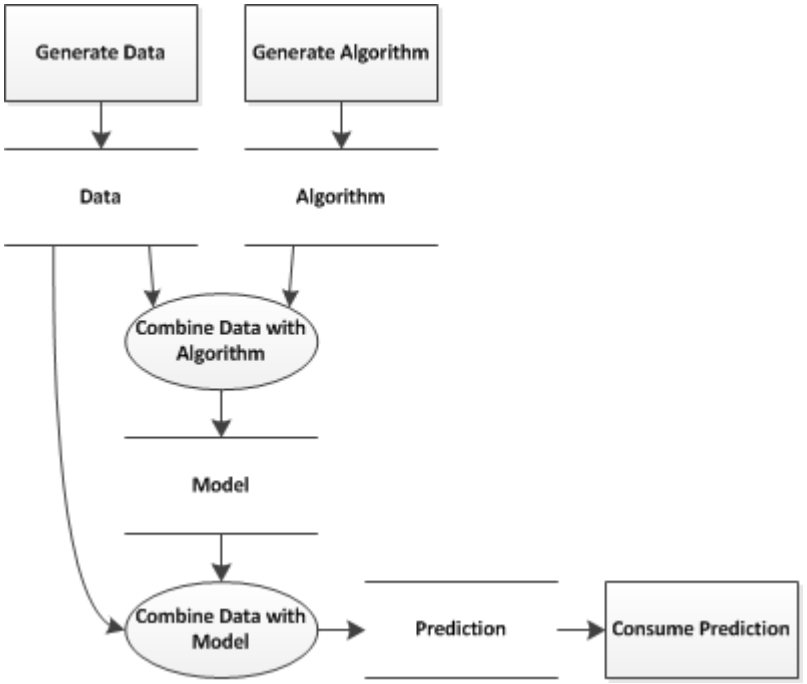
Data + Algorithm → Model

FROM DATA TO PREDICTIONS



Model + Data → Prediction

FROM DATA TO PREDICTIONS



Data + Algorithm → Model
Model + Data → Prediction

FROM DATA TO PREDICTIONS

- Pseudo Assignments (Derivations):
  - Data + Algorithm → Model
  - Model + Data → Prediction
- Create Model from Algorithm and Data
  - Example Create Logistic Regression
    - model = LogisticRegression()
    - model.fit(OldInputs, OldTarget)
- Predict from Model and Data
    - prediction = model.predict(NewInputs)
    - The prediction are for “new” target values

Data + Algorithm → Model
Model + Data → Prediction

FROM DATA TO PREDICTIONS

Some Algorithms for Supervised Learning

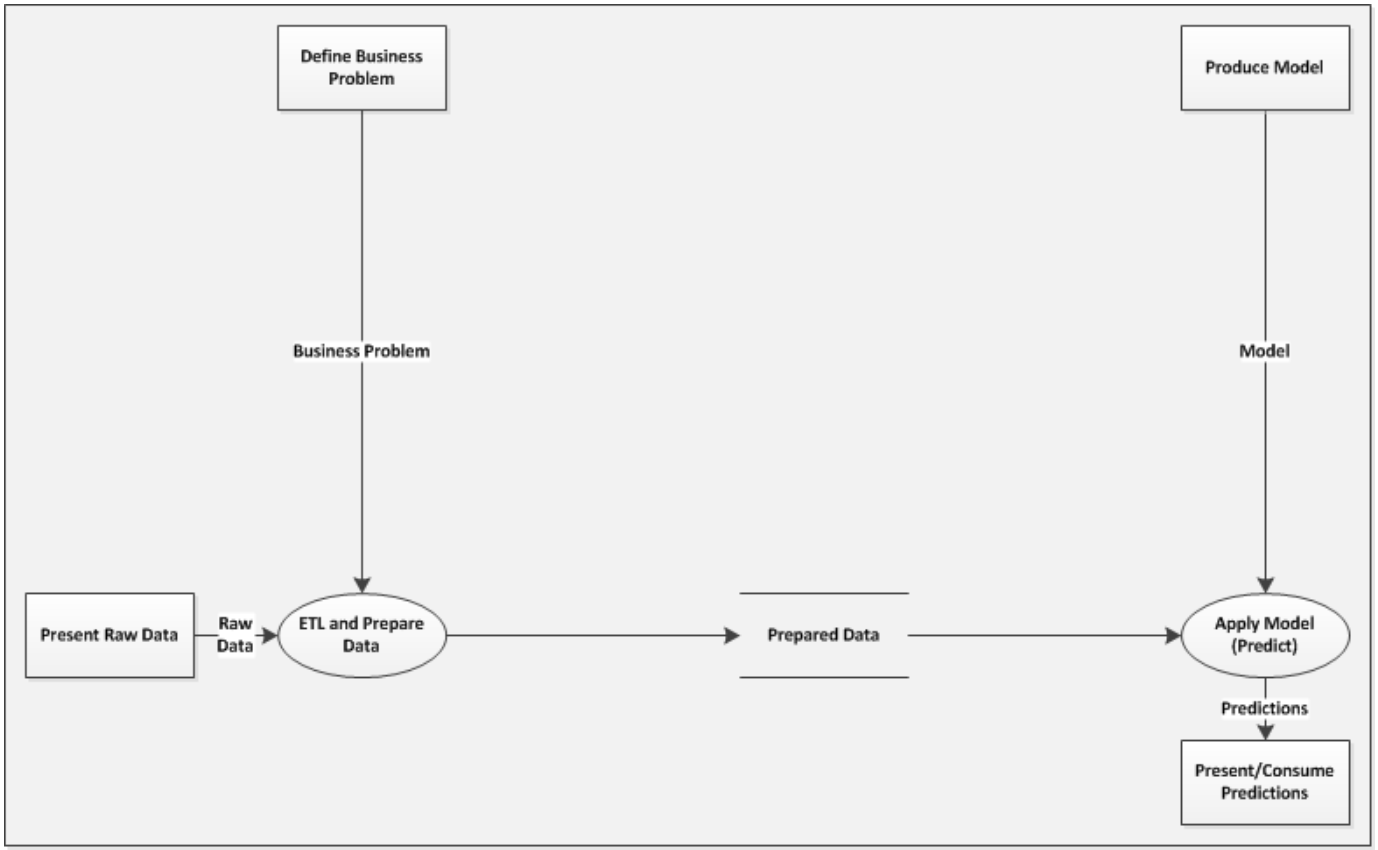
- Classification
  - Logistic Regression
  - Neural Network
  - Decision Tree
  - Naïve Bayes
- Regression
  - Linear Regression
  - Regression Trees
  - Neural Network

DFD OF SUPERVISED LEARNING

MODEL ACTS ON DATA

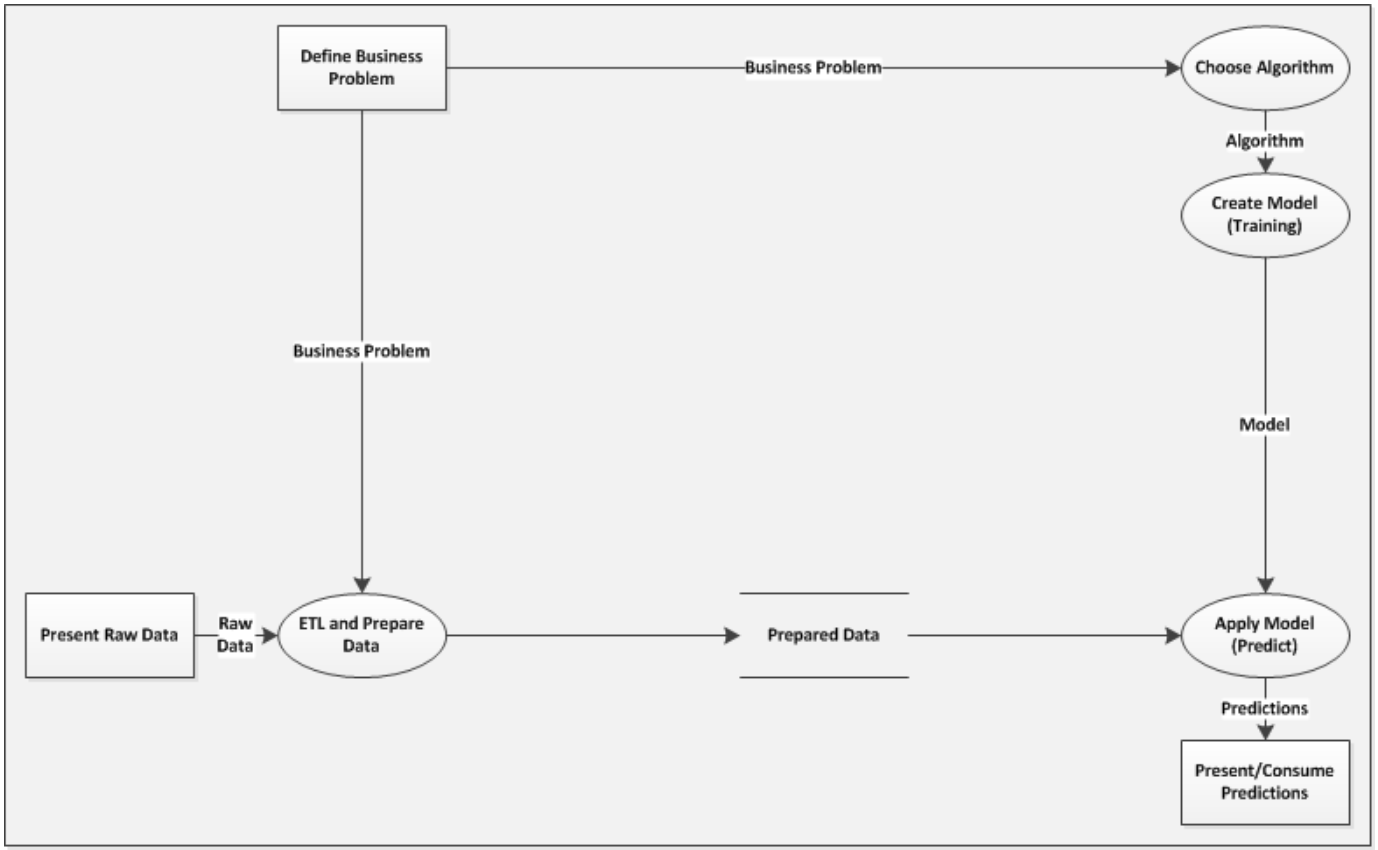
Model + Data → Prediction

DATA ETL AND PREPARATION DRIVEN BY BUSINESS PROBLEM



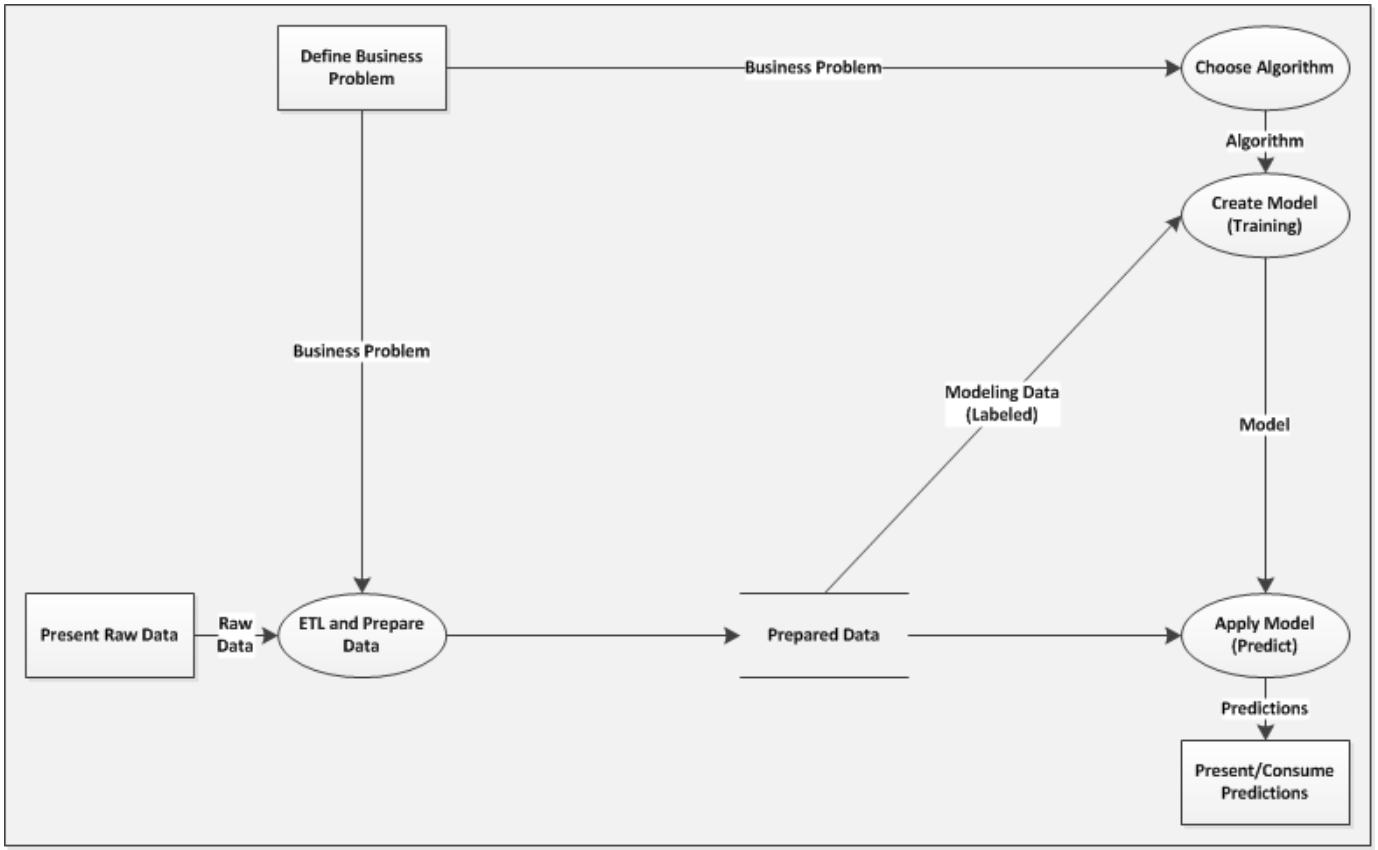
Business Problem determines ETL and Data Prep

ALGORITHM CHOICE DRIVEN BY BUSINESS PROBLEM



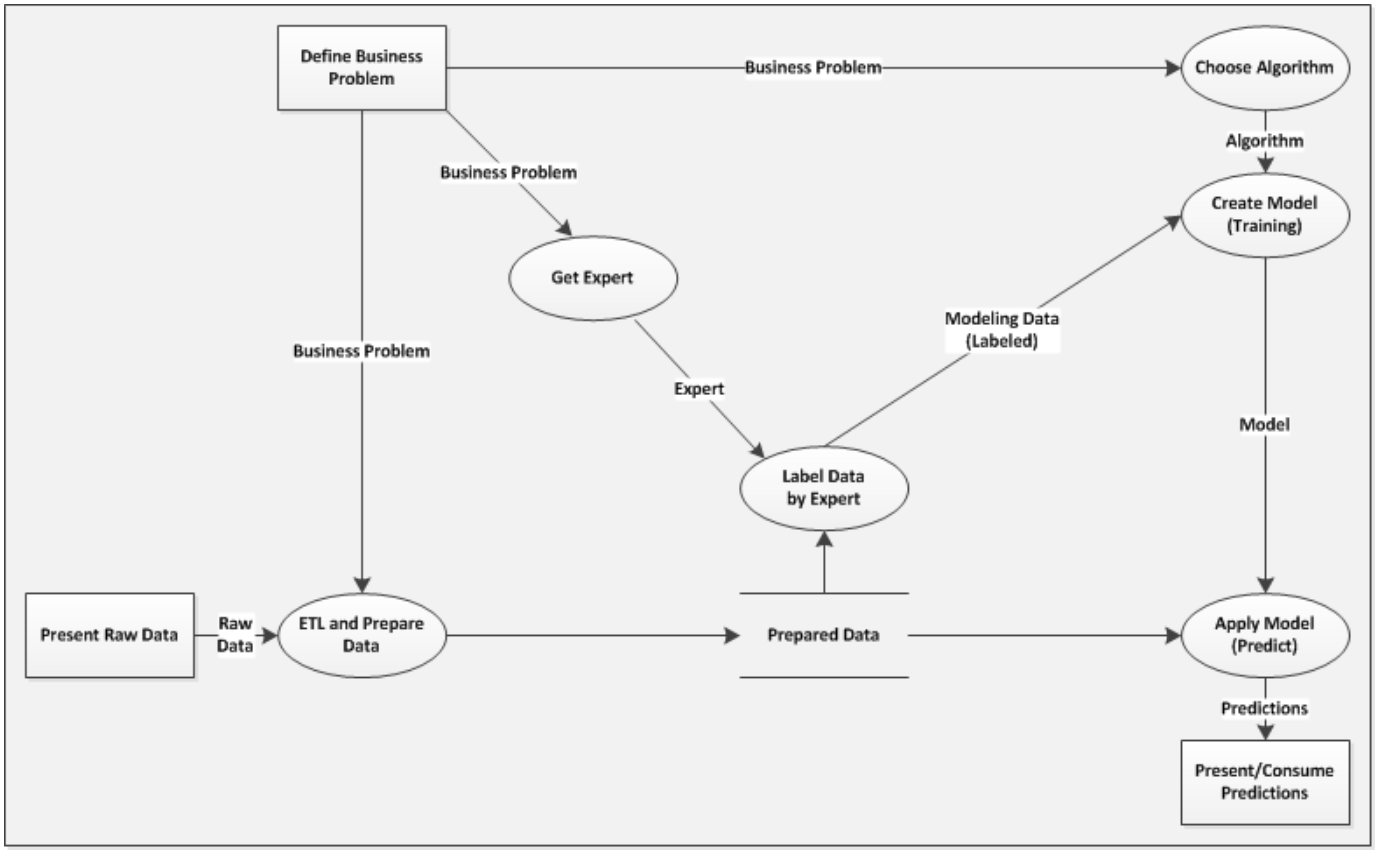
Business Problem determines the choice of Algorithm.

MODEL CREATION NEEDS DATA



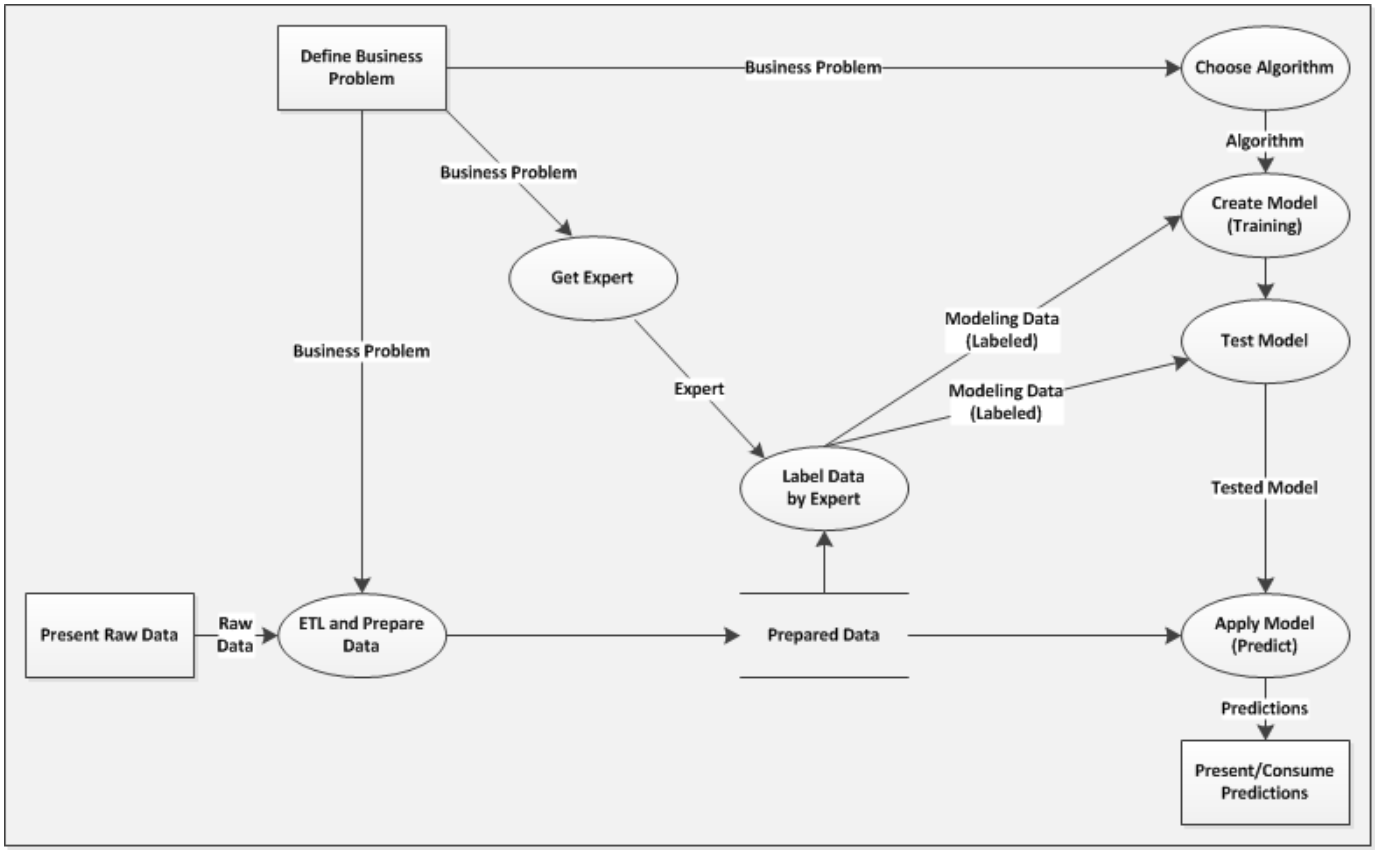
Data + Algorithm → Model

SUPERVISED TRAINING NEEDS DATA LABELED WITH OUTCOMES



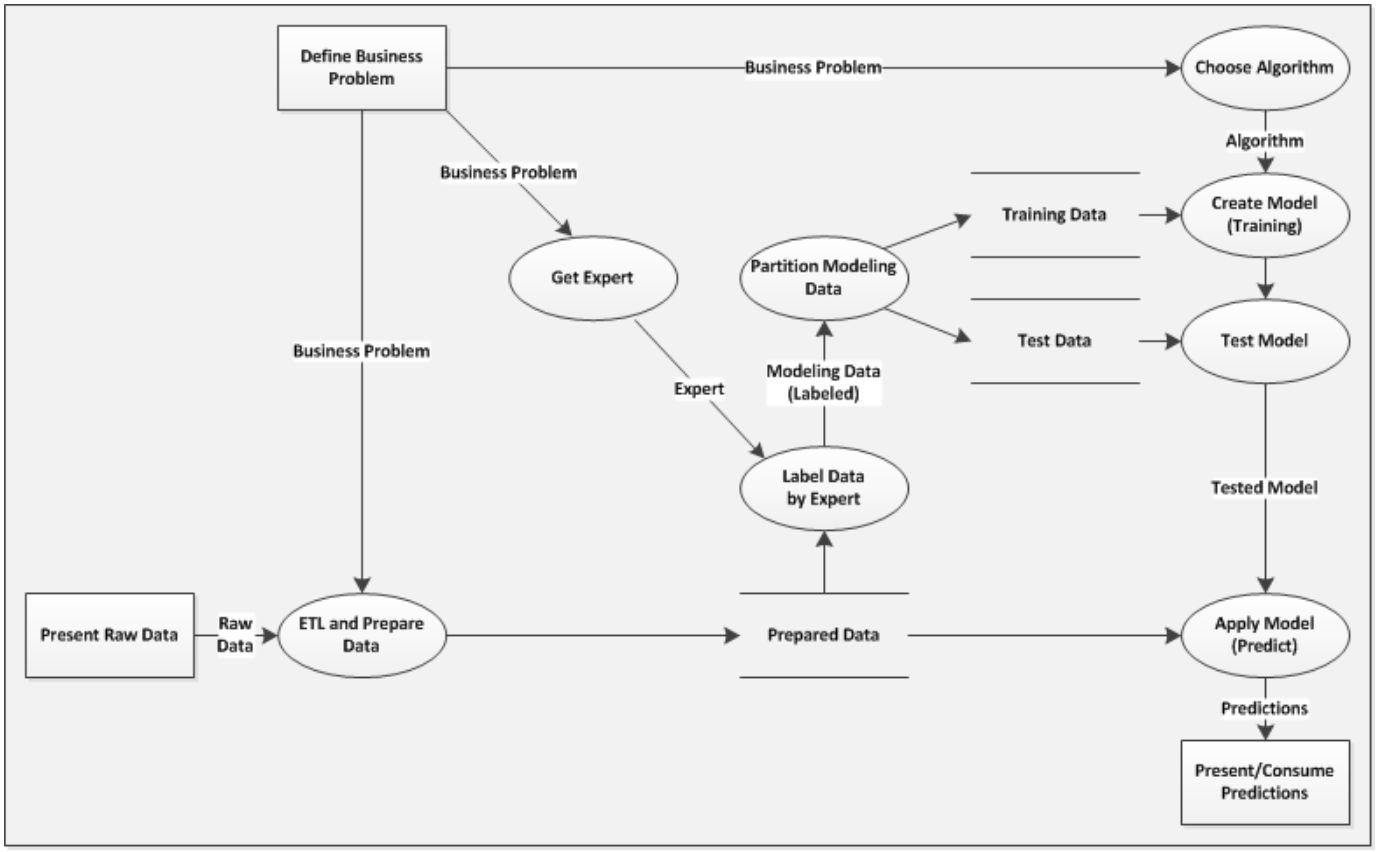
Supervised Learning requires expert labeling of data.

MODELS NEED TO BE TESTED



Do not trust predictions from an un-tested model!

TRAINING & TESTING OF MODEL USE DIFFERENT DATA



Do not test a model using training data!

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

REVIEW, CONT.

- 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