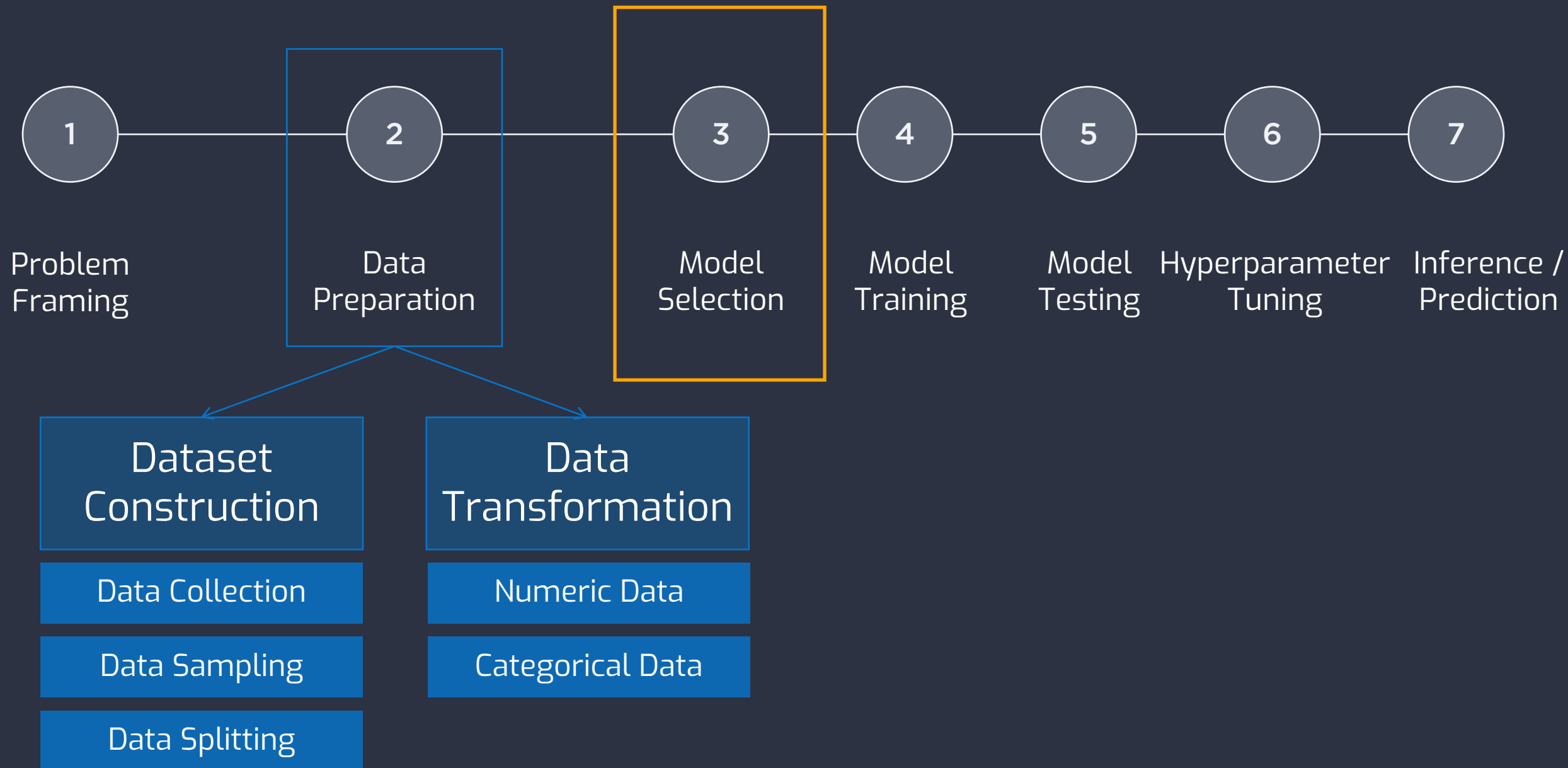


COMP2261 ARTIFICIAL INTELLIGENCE / MACHINE LEARNING

# Model Selection

Dr SHI Lei

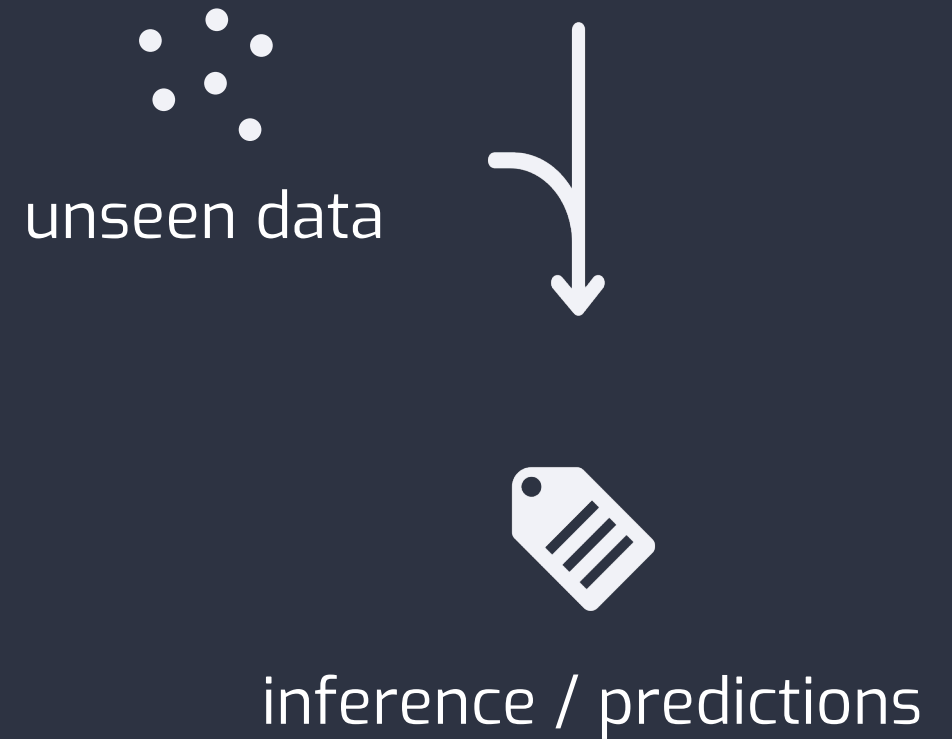


## Learning Objectives

- Understand machine learning algorithms vs models
- Have an overview of model selection



machine learning algorithm  
to train  
machine learning model

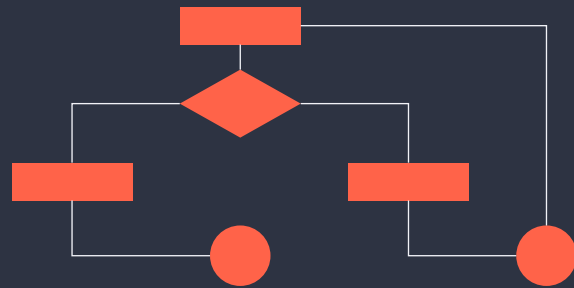


# Model Selection

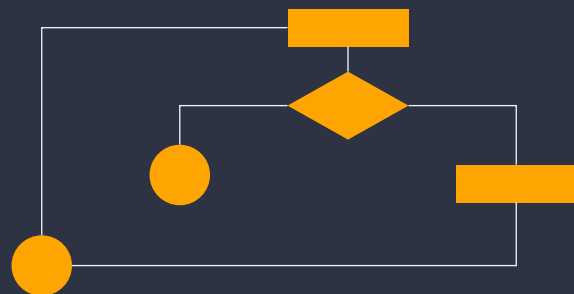


# Machine Learning Algorithm Selection

Which to try? →

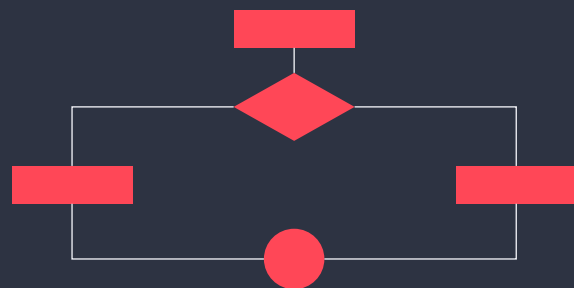


model 1



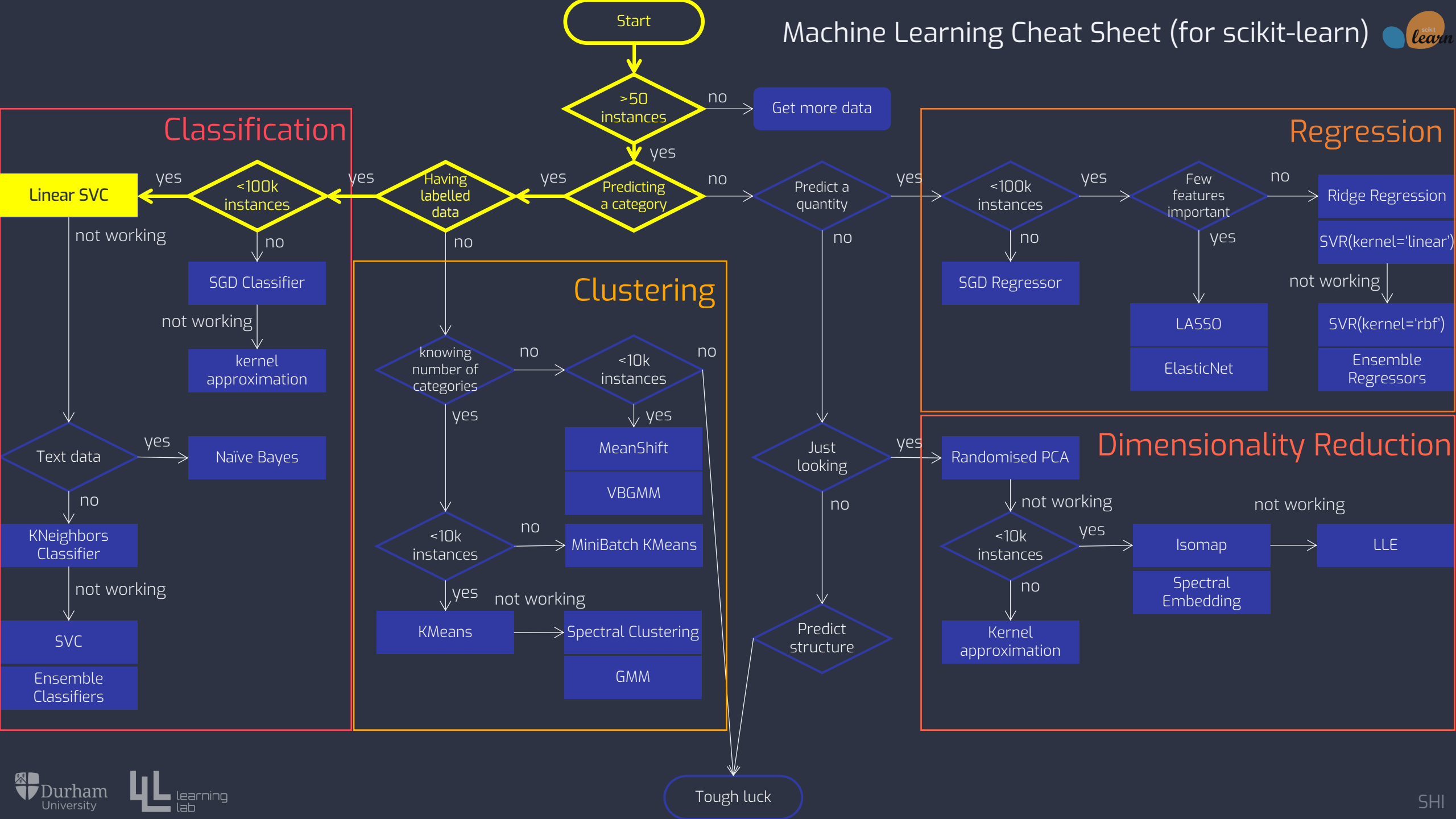
model 2

...



model N

To pick the one that performs the best.



Classification

Regression

Clustering

Dimensionality reduction



# Classification

# Classification

- Classification models are used to predict a class label for instances in the problem domain, e.g. diagnostics, fraud detection.
- Classification algorithms learn from a training set with many instances of inputs and outputs (labels) to create a classification model.
- Supervised learning.

# Regression

# Regression

- Regression models are used to predict a specific value in a continuous distribution, e.g. weather forecasting, population growth prediction.
- Regression algorithms learn from a training set with many instances of inputs and outputs (labels) to create a regression model.
- Supervised learning.

# Clustering

# Clustering

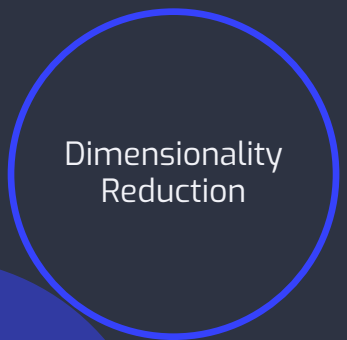
- Clustering models are used to discover interesting patterns in data, e.g., customer segmentation, recommender systems.
- Clustering algorithms learn from a training set with many instances of inputs and outputs (unlabelled) to create a clustering model.
- Unsupervised learning.

# Dimensionality Reduction

# Dimensionality Reduction

- Dimensionality Reduction models are used to decrease the number of input features, transforming data from high to low dimensional space, e.g. pre-processing data, big data visualisation, feature elicitation.
- Dimensionality reduction simplifies the dataset with fewer features, which is very useful, as more input features usually makes a predictive modelling task more challenging and may cause problems e.g. overfitting.
- Unsupervised learning.





Feature elicitation  
Meaningful compression  
Structure discovery



Recommender systems  
Targeted marketing  
Customer segmentation



Image classification  
Diagnostics  
Fraud detection



Population growth prediction  
Weather forecasting  
Market forecasting

## ✓ Takeaway Points

- A machine learning model is the output of a corresponding machine learning algorithm.
- To select appropriate machine learning models, we need to firstly select appropriate machine learning algorithms.
- During model selection, it is necessary to take into consideration the nature of the data and the problem.

