Introduction

Data Mining for Business Analytics in Python
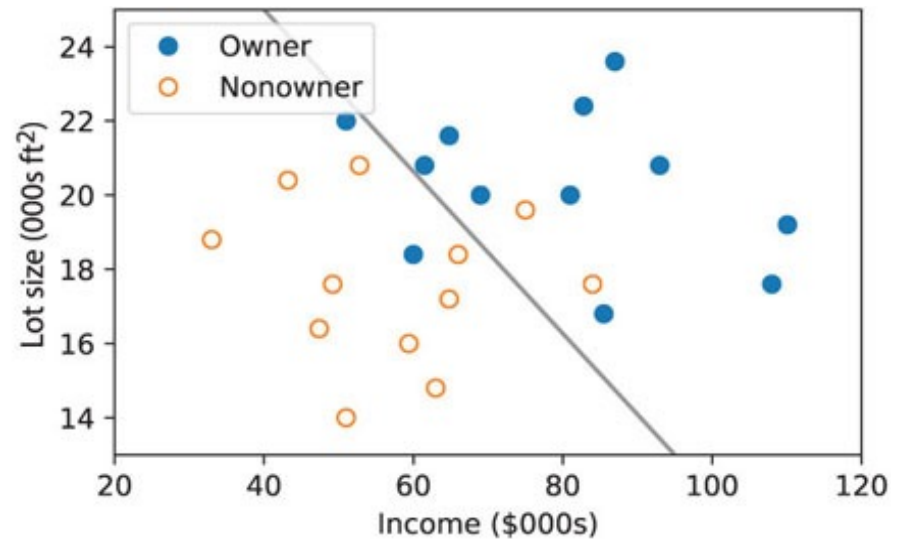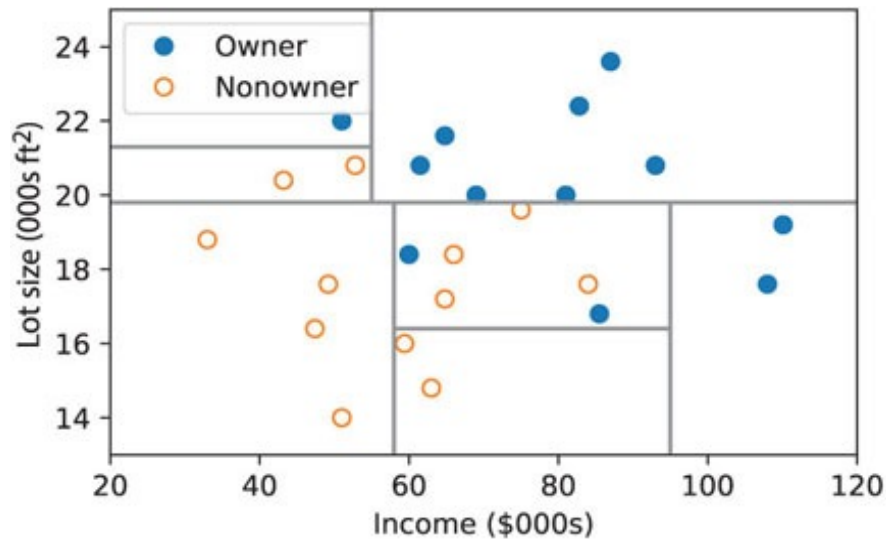
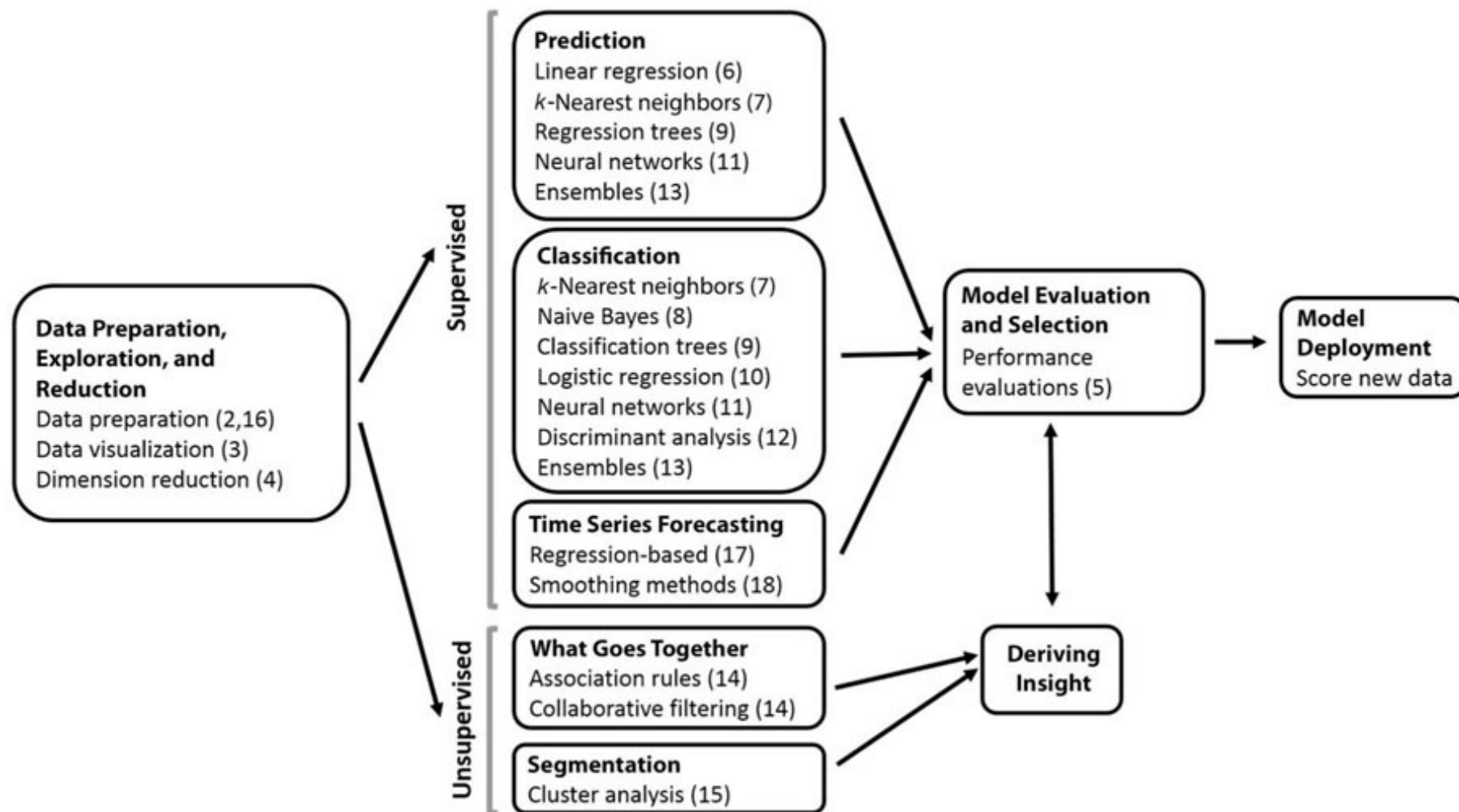**Shmueli, Bruce, Gedeck, & Patel**

Terms

- <u>Business Analytics:</u> The practice and art of brining quantitative data to bear on decision-making.
- <u>Data Mining:</u> Business analytics methods that go beyond counts, descriptive techniques, reporting, and methods based on business rules.
- <u>Big Data:</u> Data mining and big data go hand in hand, a relative term comparing data/methods/devices available today and the past. It presents the four V`s (volume, velocity, variety, veracity-reliability/accuracy/trustworthiness) challenge.
- <u>Data Science:</u> A mix of skills in the areas of statistics, machine learning, math, programming, business, and IT.

Why are there so many different methods?

- The goal is to find a combination of household income level and household lot size that separates buyers (blue solid circles) from nonbuyers (orange hollow circles) of riding mowers.

- The first method (left panel) looks only for horizontal and vertical lines to separate buyers from nonbuyers, whereas the second method (right panel) looks for a single diagonal line.

Data mining from a Process Perspective

Prediction
Linear regression (6)
k-Nearest neighbors (7)
Regression trees (9)
Neural networks (11)
Ensembles (13)

Classification
k-Nearest neighbors (7)
Naive Bayes (8)
Classification trees (9)
Logistic regression (10)
Neural networks (11)
Discriminant analysis (12)
Ensembles (13)

Time Series Forecasting
Regression-based (17)
Smoothing methods (18)

Supervised

Data Preparation, Exploration, and Reduction
Data preparation (2,16)
Data visualization (3)
Dimension reduction (4)

Model Evaluation and Selection
Performance evaluations (5)

Model Deployment
Score new data

Unsupervised

What Goes Together
Association rules (14)
Collaborative filtering (14)

Segmentation
Cluster analysis (15)

Deriving Insight

Numbers in Parentheses indicate chapter numbers