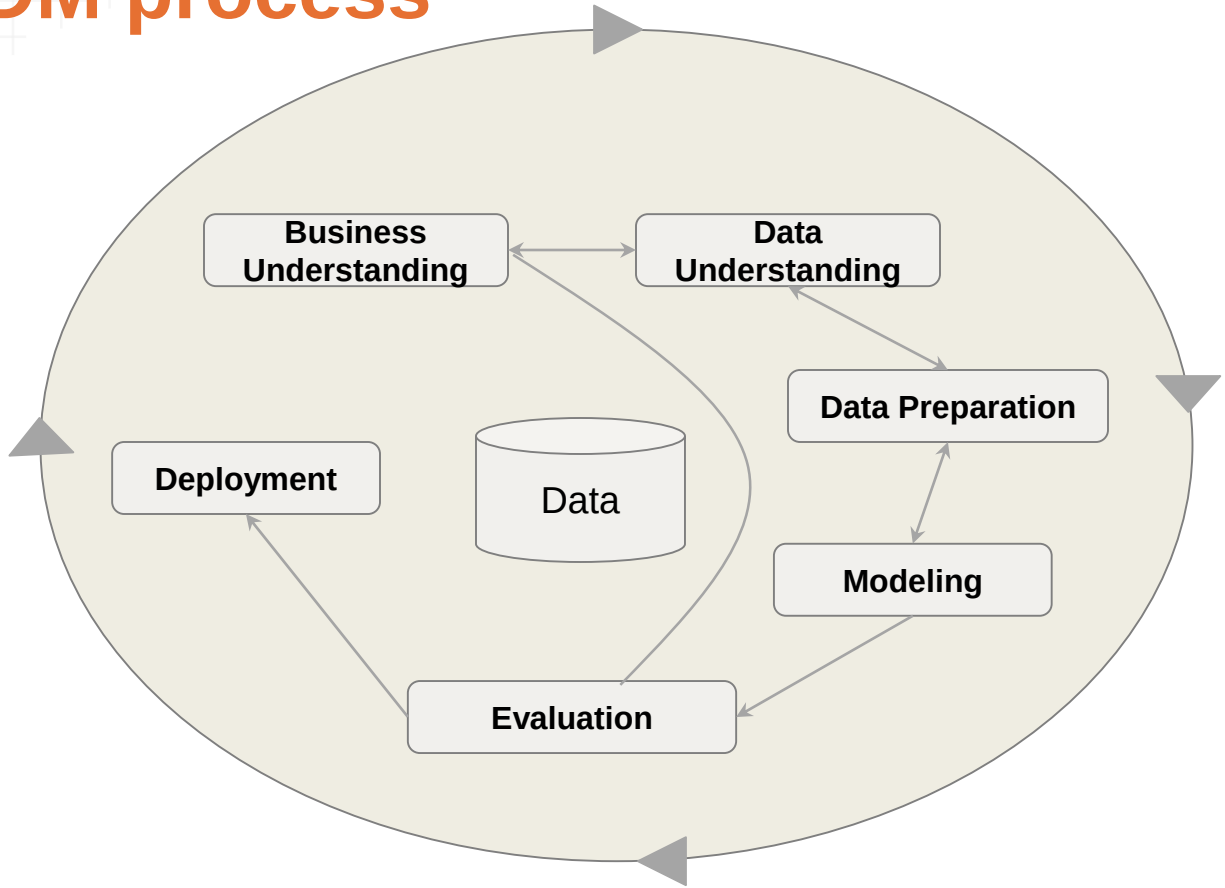
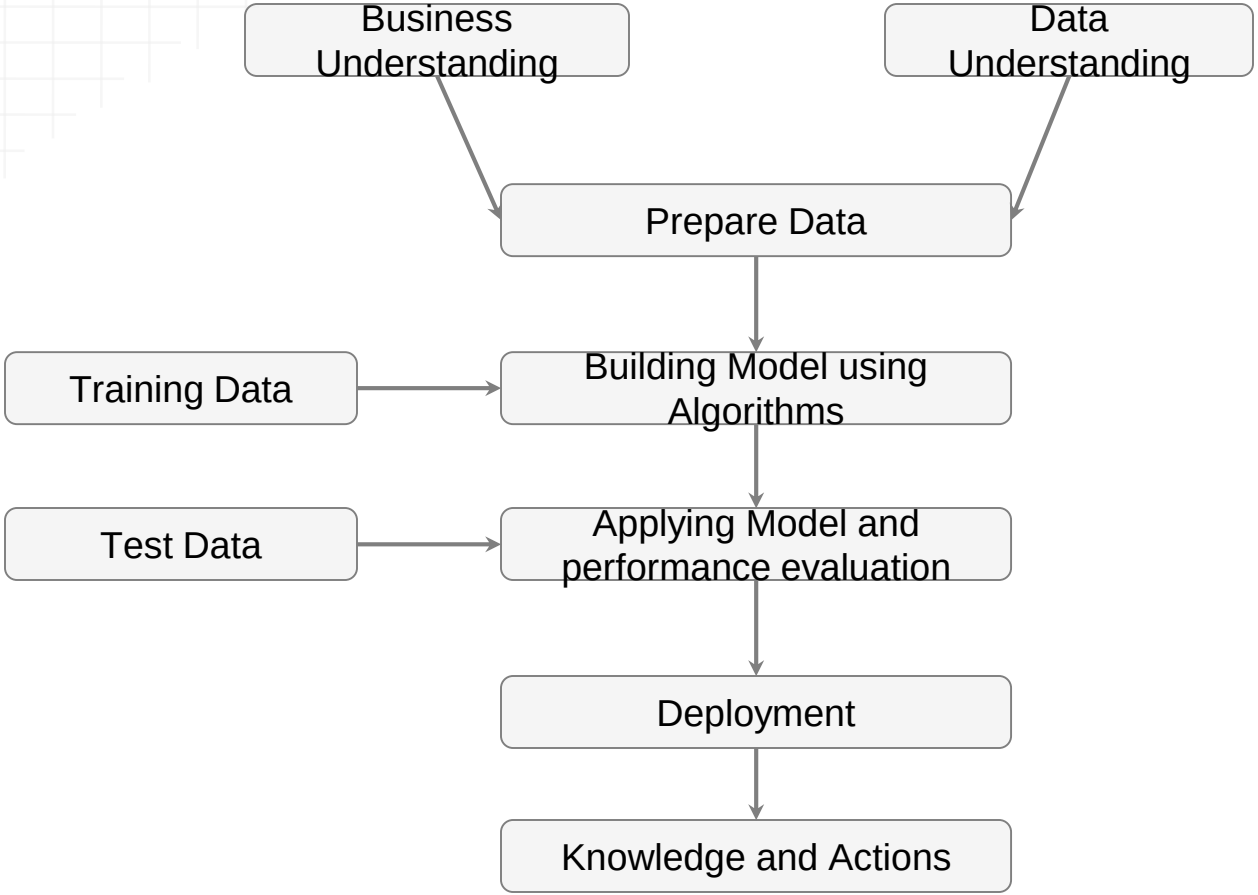
2. Data Mining Process

CRISP DM process



Process



1. Prior Knowledge

2. Preparation

3. Modeling

4. Application

5. Knowledge

1. Prior Knowledge

Gaining information on:

- Objective of the problem
- Subject area of the problem
- Data

Table 2.1 Data Set

Borrower ID	Credit Score	Interest Rate
01	500	7.31%
02	600	6.70%
03	700	5.95%
04	700	6.40%
05	800	5.40%
06	800	5.70%
07	750	5.90%
08	550	7.00%
09	650	6.50%
10	825	5.70%

2. Data Preparation

Data Exploration

Data quality

Handling missing values

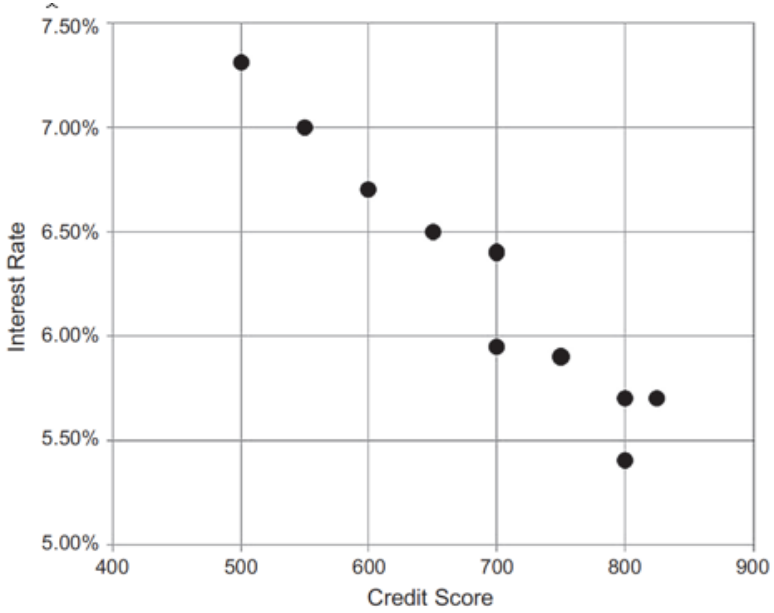
Data type conversion

Transformation

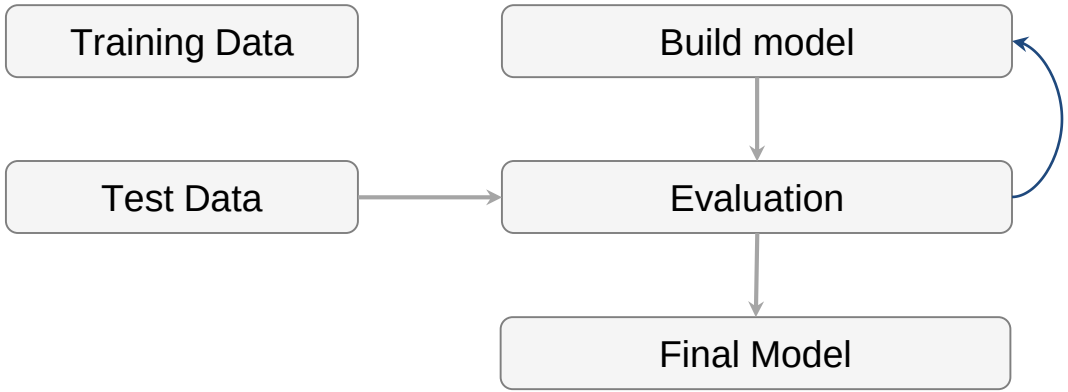
Outliers

Feature selection

Sampling



3. Modeling


Training Data
Build model
Test Data
Evaluation
Final Model

3. Modeling

Spliting training and test data sets

Table 2.3 Training Data Set

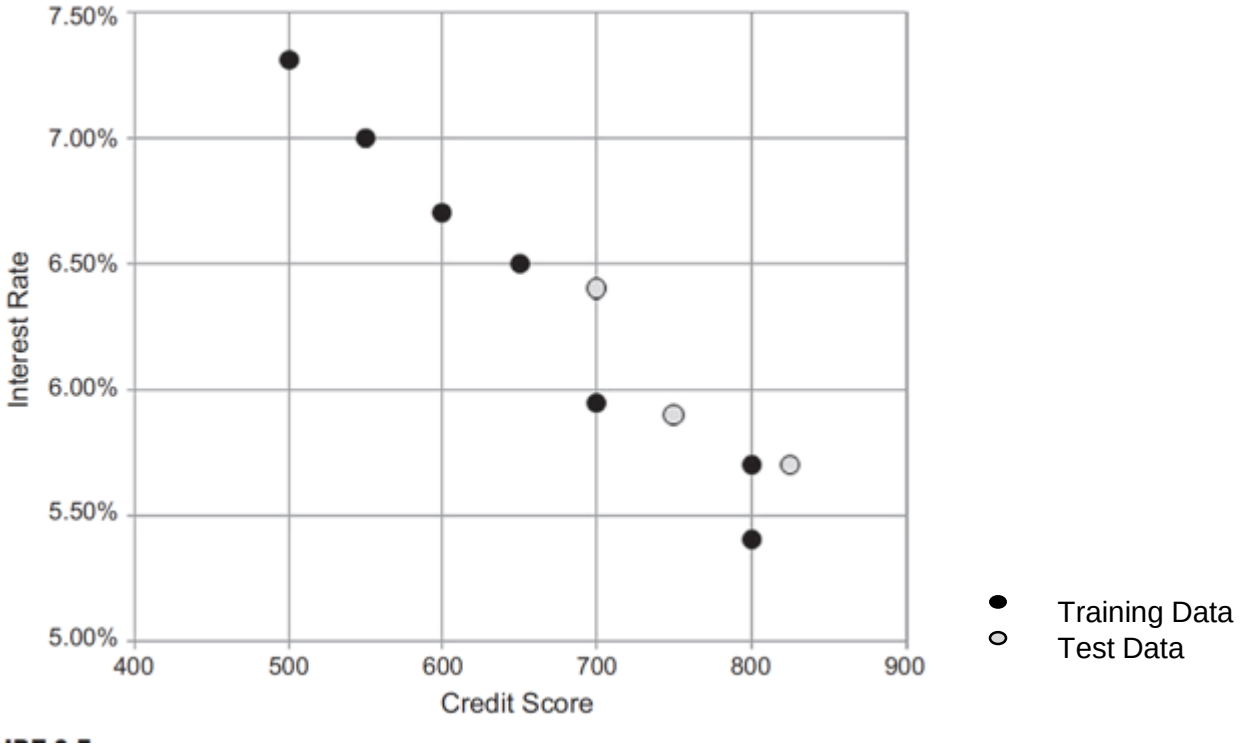
Borrower	Credit Score (X)	Interest Rate (Y)
01	500	7.31%
02	600	6.70%
03	700	5.95%
05	800	5.40%
06	800	5.70%
08	550	7.00%
09	650	6.50%

Table 2.4 Test Data Set

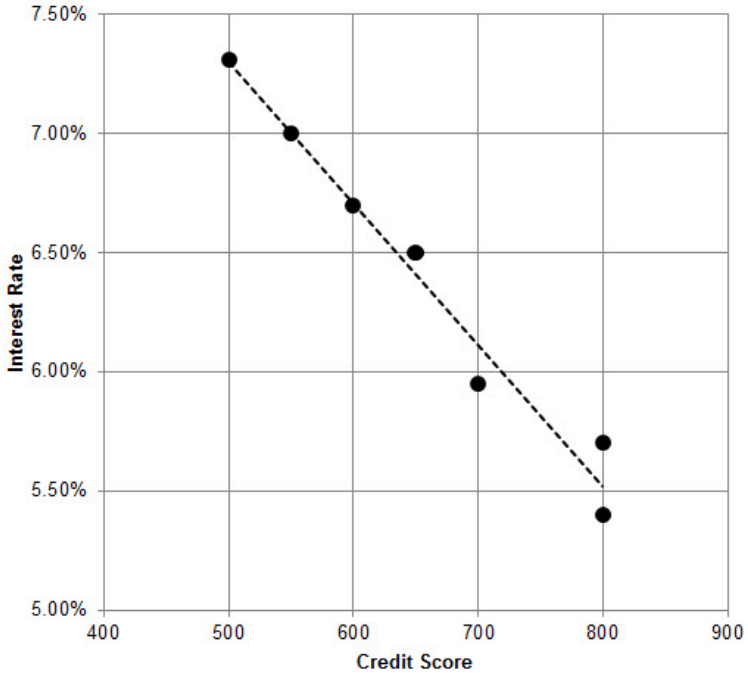
Borrower	Credit Score (X)	Interest Rate (Y)
04	700	6.40%
07	750	5.90%
10	825	5.70%

3. Modeling

Spliting training and test data sets



3. Modeling



$$y = 0.1 + \frac{6}{100,000}x$$

3. Modeling

Evaluation of test dataset

Table 2.5 Evaluation of Test Data Set

Borrower	Credit Score (X)	Interest Rate (Y)	Model Predicted (Y)	Model Error
04	700	6.40%	6.11%	-0.29%
07	750	5.90%	5.81%	-0.09%
10	825	5.70%	5.37%	-0.33%

3. Application

Product readiness

Technical integration

Model response time

Remodeling

Assimilation

5. Knowledge

Posterior knowledge

Kotu, V., & Deshpande, B. (2014). *Predictive analytics and data mining: concepts and practice with rapidminer*. Morgan Kaufmann.