4. Evaluation

4.1 Data Collection

Table 4.1 Companies grouped by lifecycle stage.

Figure 4.1 Distribution of company ages by lifecycle stage (KDE plot).

4.2 Feature Extraction

4.2.1 Missing Values

Figure 4.2 Missing features per observation (histogram).

Figure 4.3 Missing observations per feature (histogram).

4.2.2 Feature Components

Figure 4.4 Eigenvalues extracted from PCA model (line plot).

Figure 4.5 Component loadings on each factor from conceptual framework (matrix).

Figure 4.6 Inter-correlations of each factor from conceptual framework (matrix).

4.3 Classification Algorithms

Table 4.2 Classification accuracy metrics for each algorithm.

Figure 4.7 Receiver operating characteristics for each algorithm (line plot).

4.4 Experimentation

4.4.1 Practical

Figure 4.8 AUC ROC for different training set sizes (line plot).

Table 4.3 Time profiling by system component.

4.4.2 Robust

Table 4.4 Comparison of 2013 slice from 2016 dataset with original 2013 dataset.

Table 4.5 Features ranked by importance for different time slices.

Figure 4.9 ROC curves for models trained on different time slices (line plot).

4.4.3 Predictive

Table 4.6 Features ranked by importance for different prediction windows.

Figure 4.10 ROC curves for models trained on different prediction windows (line plot).

Figure 4.11 Change in company stage over different prediction windows (matrix).

Figure 4.12 ROC curves for target companies at different lifecycle stages (line plot).

Figure 4.13 AUC ROC for models by lifecyle stage and prediction window (matrix)

Table 4.7 Three example company profiles and their predictions.

C. Conceptual Framework

Table C.1 List of included features.

D. Additional Classifier Results

Table D.1 Hyperparameter gridsearch for each algorithm.

Table D.2 Optimal Logistic Regression model.

Figure D.1 Optimal Decision Tree model (tree).