

Draft Thesis Structure

Predicting Startup Exits Using Machine Learning Techniques

- Title Page
- Abstract (300 Words)
- Acknowledgements
- Contents
- List of Tables
- List of Figures
- 1. Introduction (1,500 Words)
 - 1.0 Intro - Background, Rationale
 - 1.1 Criteria - Practical, Robust, Predictive
 - 1.2 Signposting - Literature Review, Design & Implementation, Evaluation
- 2. Literature Review (3,000 Words)
 - 2.1 Technology Startups
 - 2.1.1 Entrepreneurship Behaviours
 - 2.1.2 Determinants of Startup Performance
 - 2.1.3 Startup Development Lifecycle
 - Figure 2.1 Startup development lifecycle (graphic).
 - 2.1.4 Investment Strategy for Startups
 - 2.1.5 Exit Strategy for Startups
 - 2.2 Startup Investment
 - 2.2.1 Venture Capital Firm Strategy
 - Figure 2.2 Venture capital fund lifecycle (graphic)
 - 2.2.2 Investment Opportunity Search
 - 2.2.3 Investment Decision Making
 - 2.3 Proposed Framework
 - Figure 2.3 Proposed conceptual framework for startup investment (tree).
 - 2.4 Research Gap
- 3. Design & Implementation (4,000 Words)
 - 3.0 Introduction
 - Figure 3.1 System architecture overview (flow chart).
 - 3.1 Data Collection
 - Table 3.1 Data sources indicated to be relevant to startup investment.
 - 3.1.1 Source Characteristics - Databases, Social Media, Other
 - 3.1.2 Source Evaluation
 - 3.2 Feature Extraction
 - Table 3.2 Features indicated to be relevant to startup investment.
 - 3.3 Classification Algorithms
 - Table 3.3 Evaluation of machine learning algorithms for startup investment.
 - 3.3.1 Task Characteristics
 - 3.3.2 Algorithm Characteristics
 - 3.3.3 Algorithm Evaluation
 - 3.4 Experimentation
 - 3.4.1 Practical - training_size, time_taken
 - 3.4.2 Robust - n_slices
 - 3.4.3 Predictive - period, stage
- 4. Evaluation (5,000 Words)
 - 4.1 Data Collection
 - Table 4.1 Final test dataset counts 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 Number of missing features per observation (histogram).
- Figure 4.3 Number of 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 - training_size, time_taken
 - Figure 4.8 AUC ROC for different training set sizes (line plot).
 - Table 4.3 Time profiling by system component.
 - 4.4.2 Robust - n_slices
 - 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 - period, stage
 - 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 lifecycle stage and prediction window (matrix)
 - Table 4.7 Three example company profiles and their predictions.
- 4.5 Discussion
 - 4.5.1 Practical - training_size, time_taken
 - 4.5.2 Robust - n_slices
 - 4.5.3 Predictive - period, stage
- 5. Conclusions (1,500 Words)
 - 5.1 Evaluation of Criteria - Practical, Robust, Predictive
 - 5.2 Future research - Network Analysis, Temporal Analysis, Full Automation
 - 5.3 Summary
- Bibliography
- A. Glossary
- B. Statistical Measures
- C. Conceptual Framework
 - Figure C.1 Proposed conceptual framework for startup investment - extended (tree).
 - 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).
- E. Original Honours Proposal
- F. Revised Honours Proposal
- G. Revised Literature Review