Data Science

Paragraphing, Sectioning and Coherence of the Whole Piece

# Activity:

**Read the section from the literature review and answer the following questions:**

* How are the paragraphs structured?
* How does each paragraph achieve the purpose of the section? How are the different paragraphs connected?
* How do the sections fit together?

## **Dissertation Title: Can Machine Learning Accurately Predict the Success of a Start-Up?**

*2.4.2 Data Analysis and Machine Learning in Venture Capital*

In general, current VC investors rely on their network to identify potential deals. VC investors have been slow to incorporate any form of a data-driven approach into their investment frameworks, which is especially ironic given the fact that VCs invest in cutting edge technology themselves (Sun et al., 2019). In a recent study, it was found that only “38% of VC investors use data to source and evaluate all investment opportunities” (PitchBook, 2019). This is even more evident in angel and early-stage VCs who are 20% more likely to use personal networks (PitchBook, 2019). With most VCs (82%) relying primarily on personal networks, the question here is why VCs are yet to incorporate a more data-driven approach into their investment framework (PitchBook, 2019). Deal selection is an integral part of the investment framework and identifying, analysing, and interpreting the key signals which may relate to the future success of a start-up is essential, especially given the high-risk factor and natural uncertainty of start-ups (Hoenig & Henkel, 2015). However, there is an ever-increasing pool of literature which has researched the effectiveness of data driven approaches (through machine learning) in aiding the VC investment process.

*2.4.3 Theoretical use of Machine Learning in Venture Capital Investment Process*

Multiple papers have shown very strong results and findings by creating machine learning models to try and aid the investment process of VCs. Zbikowksi & Antosiuk (2021) aimed to create a binary classification model to predict the success of company, with the success of a company being defined as the completion of a series B funding round. Using a very large training data set from Crunchbase (N = 213,171), and a large testing data set (N = 10,000), they found an average test set accuracy of 85% using three different machine learning methods: 1) Logistic regression (86%); 2) Support Vector Machine (84%); and 3) XGBoost (85%). Furthermore, looking at individual variables within the models, they found that the most important variables were ones related to geographical features, such as country code, region size, and city size (Zbikowksi & Antosiuk, 2021).

In a similar paper, Arroyo et al. (2019) created multi-label classification models to predict the next event of a start-up. These labels were: 1) Acquired, 2) Closed, 3) IPO, 4) Funding Round, 5) No event. Five different machine learning methods were used and subsequently compared for their global accuracy in correctly classifying the next event: decision trees, random forest, extremely randomized trees, gradient tree boosting, and support vector machines. All models showed high accuracy above 74.6% (decision trees) with gradient tree boosting showing an accuracy of 82.2%. They concluded that “machine learning can help in the baseline screening to early-stage investors that are looking at potential investments with no relevant quantitative data or track record” and is therefore conducive to very early (seed) stage focused VCs (Arroyo et al., 2019). They note that the model “cannot and should not be the only tool used to evaluate the company, but is definitely the first step an investor should undertake in order to move forward a conversation or not”, and furthermore that it “augments the ability to process information and assess companies in absence of more traditional financial data” (Arroyo et al., 2019). Of course, these research papers have not directly applied these techniques in a practical setting and the success they have shown remains theoretical. However, there are some notable examples of VC firms using a data driven approach to their advantage in a practical investment setting.

*2.4.4 Practical use of Machine Learning in Venture Capital Investment Process*

Even with the incorporation of simpler machine learning methods into the VC investment process such as a logit framework, it has been demonstrated that VC firms are more likely to lead their companies to successful exits (Nahata, 2008).