A logo for college computing

Description automatically generated

**Predicting Start-up Success**

|  |  |
| --- | --- |
| *Student Full Name:* | Karen Montserrat Andraca Guzman |
| *Student Number:* | 2024345 |
| *Module Title:* | Strategic Thinking |
| *Assessment Title:* | CA 1 – Capstone Project Proposal |
| *Assessment Due Date: Data of Submission* | 27-10-2024 |
|  |  |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Abstract

The goal of this capstone project is to create a prediction model that will evaluate the probability that a start-up company will become a unicorn, which is defined as having a valuation of more than $1 billion. We will examine a number of factors that affect start-up success, such as industry sector, funding history, and financial performance, using publicly accessible data from sources including Crunchbase and Kaggle.

After identifying important trends and behaviours through exploratory data analysis (EDA), pertinent features will be chosen for modelling. To develop a strong prediction model, we will use machine learning techniques learning in class like logistic regression and random forests.

The outcomes, which highlight significant features that support a start-ups potential for quick development and success, will offer useful information to investors and entrepreneurs. In addition to improving knowledge about unicorn scene, this study intends to provide practical suggestions for encouraging investment and creativity in up-and-coming businesses.

Contents

[Introduction 1](#_Toc180511970)

[Important definitions - Chapter 1 2](#_Toc180511971)

[What is a start-up - Chapter 1.1 2](#_Toc180511972)

[What is a “Unicorn Company” Chapter 1.1.1. 2](#_Toc180511973)

[Scope and Methodology - Chapter 2 3](#_Toc180511974)

# Introduction

The last ten years have seen a dramatic change in the start-up scene, with an increase in the number of businesses becoming unicorns—those valued at more than $1 billion. These unicorns have become significant players in the world economy, fostering innovation, generating employment, and boosting Economy. Strong investment, talent access, and encouraging government policies all contribute to each country successful startup ecosystem. But even with the enthusiasm and promise, most entrepreneurs fail to reach this level of success.

Investors, founders, and politicians must all have an excellent knowledge of the characteristics that set successful enterprises apart from those that fail. The goal of this capstone project is to create a prediction model that identifies the crucial qualities and factors influencing a start-ups chances of becoming a unicorn. Through the examination of past data from many sources, we aim to find patterns that help direct strategic planning and decision-making.

# Important definitions - Chapter 1

## What is a start-up - Chapter 1.1

I want to begin by examining the idea of a start-up and how, although it is a very basic description, it also fits into the modern world and its business practices. [[1]](#footnote-1)Eric Ries defines a startup as an organization dedicated to creating something new under conditions of extreme uncertainty.

To put it another way, a start-up is a business that is still in its beginning and aims to provide an innovative service or product that will lead to success and the positioning required to make money. However, the founders are inexperienced and frequently lack the resources needed to establish a traditional company, requiring apply with institutions or investors, known as "angels" in this industry, which suggests a significant challenge given the risks and uncertainty of the business.

### What is a “Unicorn Company” Chapter 1.1.1.

We begin by defining a unicorn company, which is essentially a start-up that has been established in the market and has a valuation of at least $1,000 million dollars without being public companies (listed on the stock exchange). As excellent examples, we have the following at the top of the global rankings:A screenshot of a phone

Description automatically generated

Source: CB Insights.

The Chinese media and entertainment company ByteDance, which owns the social media platform TikTok, is the world’s most valuable unicorn company at $225 billion.

Meanwhile, the largest unicorn company in the United States is SpaceX, led and founded by Elon Musk, valued at $150 billion. SpaceX is the only company on the list in the industrial category.

OpenAI, founded in 2015, is the world’s highest-valued artificial intelligence pure-play company. Best known for its groundbreaking program ChatGPT, it is worth $80 billion.[[2]](#footnote-2)

# Scope and Methodology - Chapter 2

Develop a model that predicts the likelihood of a startup becoming a unicorn. By examining data and patterns of unicorn companies, we want to forecast how a budding business can become a unicorn.

Analysing the business sectors in the and the ranking of unicorn companies around the world to identify which one is most likely to produce a favourable environment for new start-ups and prosperous expansion is our next goal.

And lastly, apply the models we learnt in class to evaluate and interpret the data and confirm whether it makes sense for an investor to place an investment on an start-up.

As part of the planned methodology, data will be gathered from a variety of websites, including Crunchbase, Enterprise Ireland, Pinchbook, Kaggle, and a few others to supplement if I am unable to gather all the information I need. Once I have the information I need, I have chosen to test the following models: To assess my versions' success and failure in light of the binary classification, I have selected the logistic regression model in order to increase prediction accuracy and view successful and unsuccessful examples while accounting for independent variables. In order to assess the significance of each variable in the prediction and to have enough data for the prediction, I have chosen to employ the random forest model, which will examine subsets of features to add diversity and lower the chance of overfitting.

To determine whether a startup has the potential to become a unicorn company, it will be required to collect important data, such as the profitability of products and services, and the sources of revenue. In addition, it is meant to analyse the market's size, trends, and direct and indirect competitors. We could go over the financial area's projections, expenses, income growth rates, and investor engagement and investment levels.

According to the findings we anticipate, investors and entrepreneurs will have a better idea of what steps to take to make their start-up businesses stand out and, in the best case scenario, become unicorn companies. Given that some of the required data is sensitive information specific to each company and that there are likely to be fewer examples in Ireland, the capacity and volume of information that will be gathered will determine the best and most accurate panoramas.

If we use the "Unicorn Companies Overview" database, which is available for free and in csv format on the Kaggle website, as a source of information, we can look at information like each company's valuation, industry they are in, city of origin, year of foundation, amount of money raised, and investors involved. This will allow us to determine the most likely scenario for a start-up to succeed in becoming a unicorn by looking at typical examples of the biggest unicorns in the world.

The project is dedicated to being transparent in its approach, clear information will always be provided on the origin of the data and its analysis, essential ethical characteristics are taken into consideration that are important to guarantee the integrity and responsibility of the use of the data when carrying out this project to predict the success of a start-up, and the school and teachers are exempt from any improper usage of data, pointing out that this project is merely educational and for use only by the subject. This project will continue maintain its ethical standards and be receptive to criticism and a review of its applied methodologies.

**Ethical Machine Learning Research Student Researcher Capstone Project Ethics Form**

As a student researcher in the capstone project, you are vital in ensuring the project is conducted ethically and without bias. Before you embark on the capstone project, please carefully read and understand the following information.

By conducting the capstone project, you will be actively engaged in the following research activities:

1. Data analysis and preprocessing.
2. Training and evaluating machine learning models.
3. Drawing conclusions on the machine learning models and the dataset, nothing else.

The capstone project will comprehensively examine machine learning models without introducing biases or negatively affecting vulnerable populations in society. The focus is on improving the fairness and accuracy of the results obtained from the machine learning models. The focus is on the models themselves and the dataset without delving into social work, bias or anthropology. Use deductive reasoning on the model outcomes.

|  |  |  |
| --- | --- | --- |
| Yes | No | As a student researcher, your responsibilities include: |
|  |  | Adhering to ethical guidelines provided for data handling and analysis. |
|  |  | Data is anonymised, and personal identifiers will be removed to protect the privacy of individuals. |
|  |  | Ensure that the research does not introduce bias and avoids adverse effects on the vulnerable in society and minors. |
|  |  | Focus on evaluating and improving machine learning model accuracy and results. |

I have read and understood the information provided in this ethics form, and I agree as a student researcher to abide by the ethics outlined in this form in the capstone project. I understand that my role will focus on machine learning models and dataset analysis and will not involve social work, health and care, vulnerable social groups or anthropology (no personal interviews of human subjects).

If you have any questions or concerns or require clarification on the capstone project's ethics, please contact James Garza at jgarza@cct.ie.

Student Name (printed): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Number: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Signature: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Ries, E. (2011). The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. United Kingdom: Crown. [↑](#footnote-ref-1)
2. Visual Capitalist. (2021, March 18). Ranked: The top 10 most valuable unicorn companies. <https://www.visualcapitalist.com/cp/ranked-the-top-10-most-valuable-unicorn-companies/> [↑](#footnote-ref-2)