Abstract

CIND820: Big Data Analytics Project

Iuliia Sharafutdinova

D1H & 501188055

Tamer Abdou, PhD
Assistant Program Director, Data Science

September 25, 2022

Abstract

Over the past decade, billions of dollars have been invested by funds, government and financial institutions in entrepreneurial ventures—what is often referred to as venture capital. This project aims to analyze attributes of venture capital investments in start-up industry by using a panel data set of 101 countries over the period 1983 to 2014 in order to build predictive model for success and failure raising funds.

This study analyzes critical factors of success and failure of startups. The main goal, therefore, is to identify which crucial characteristics found in the sample had the greatest impact on the success and failure of the startup to receive funding.

The dataset about start-up companies is extracted from CrunchBase dated December 02, 2014 and downloaded from Data.World website: https://data.world/datanerd/startup-venture-funding

The dataset includes amount invested by funds, acquisitions, stages, the total amount of venture capital, funding rounds, regions, categories of the companies and year of foundation.
The dataset is embracing 114,506 number of rows and 23 attributes including both 5 quantitative and 18 qualitative target attributes.

Four machine learning methods will be used in this project – Logistic regression, Decision Tree, Random-Forest and KNN.

The Decision Tree and Random-Forest will be created using Python and R, while the KNN and Logistic regression method will be applied through Python. Finally, Tableau will be used for designing visualization of analysis results.