

ABSTRACT

Aim for this research is to find targeted startups to fund, based on startup's success variable using data on Crunchbase. This study uses quantitative research with descriptive approach, where the writer will use predictive analysis alongside Crunchbase as the data source. This study will test which variable affects a startup's success (IPO, acquired, unicorn, late stage)

On this study, the writer will use some variable data, some of it are: company's name as key variable, startup status (whether the startup already IPO, acquired, unicorn, late stage) as predictable variable, startup characteristic variable, and fund activity variable. 3.000.000+ startup data by the end of 2023 on Crunchbase will use as the population for this study. Sample sorting methods used in this study are: licensed startup or startup that operate in South East Asia, startups whose business fields are limited to several priority sectors, startup founded in 2008-2023, and startups that received their last funding in the period 2018-2023.

Using defined input variable and using logistic regression algorithm to find which variable has strong connection to a success startup. After success startup variable found then using clustering method and K-means algorithm, targeted fund startup will be determined. Using clustering will determine which startup have the same success profile in the future in one cluster.

Keywords : *Venture capital, startup, logistic regression, K-Means, clustering*