Unveiling The Key Attributes of Leading Crowdfunding Projects

Research Project

Master of Science in FinTech

Kanishka Dhyani

X22232745

School of Computing

National College of Ireland

Supervisor: Victor Del Rosal

National College of Ireland

MSc Project Submission Sheet

Student Name: Kanishka Dhyani

Student ID: X22232745

Programme: Master of Science in FinTech **Year:** 2024

Module: MSc. Research project

Lecture: Victor Del Rosal

Submission Due Date: 12th August, 2024

Project Title: Unveiling The Key Attributes of Leading Crowdfunding Projects

Word Count: 5661 **Page Count:** 20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature: Kanishka Dhyani

Date: 12th Aug 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties**.
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

UNVEILING THE KEY ATTRIBUTES OF LEADING CROWDFUNDING PROJECTS

Kanishka Dhyani
X22232745

Abstract

In the era of finance, crowdfunding is considered as uncontrolled way of financing upcoming projects and it facilitates the financial needs of recently developed projects. Understanding the importance and growth of crowdfunding several platforms are providing favourable opportunities to founders, who can furnish their ideas and ask for funds from investors. Kickstarter is the most popular platform of crowdfunding, offering numerous categories for funding and among all, one of the highly funded projects are under games, design and technology category, indicating the advancement as well as interest of people in this industry. To evaluate and identify the key attributes of these crowdfunding projects, which are influencing the outcome of the process either success or failure, the study utilizes the secondary data gathered from private scrapping website and previous related researches for greater insights. Machine Learning models: Logistic Regression, Random Forest Classifier, Gradient Boosting with XG boost and K-Nearest Neighbours are applied to analyse the factors affecting the outcome of the project. The evaluation with accuracy score, AUC-ROC, Mean Squared Error and confusion matrix, the study finds that backers count and goal are the key attributes effecting the state of project, whereas duration have a bare minimum impact. Further, tokenization and distribution are used to examine the impact of words in the funding decisions of the backers.

Keywords: Crowdfunding, Kickstarter, Influential Factors, Predictive modelling

1. INTRODUCTION

At the time of commencing the business, the new ventures majorly face the external financing difficulties due to lack of pervious performance reports, larger funds generally provided by the angel investors and venture capitalists but the entrepreneurs who are seeking small funds, some gets from the family or friends but some still remain unfinanced due to insufficient security value to investor and not able to convince

investors (Belleflamme et al.,2014). To overcome this hindrance, innovative creator came with the idea of crowdfunding which taps the "crowd" rather than high profile investor. The crowdfunding platforms permits the efficient blending of funders and seekers, accumulating small chunks of funds, transforming to large pools and reducing the geographic hurdles (Pomeroy et al.,2019).

In 2020, the transaction value was $1.15 billion worldwide with average funding of $7.57K per project, and forecasted $1.27 bn in 2028 indicating the steady growth in the crowdfunding market due to believing crowdfunding as legal means of raising funds and enhance accessibility according to the publicly available data at Statista. Currently, as per the Kickstarter statistics available on their website $8,124,409,601 amount is pledged to projects with 23,285,530 backers. The platform has 15 major categories includes games, design, technology, film & video, publishing and others, out of all category's games, design and technology category gets the highest funding by the backers amounting to $5.75bn out of total $8.12bn contributing to 71% in total funds. Lending-based crowdfunding has the largest market share among the all categories of crowdfunding (Böckel, Hörisch & Tenner, 2021) Observing the constructive growth of reward-based crowdfunding, will definitely see more categories and participation of investors, which will ultimately contribute towards to growth of economy.

As per Belleflamme et al. (2014), tremendous increase is noticed in the targeted amount through crowdfunding, more than £1 million is targeting with *Trampoline systems* for new software. Successful campaigns are vital for founders, backers, operators and other interested people but where there is online involvement of non-professional backers makes it little difficult to make investment decision as compared to traditional investment ways (Koch and Siering 2019). Companies looking for the solutions in relation to the new product launch, crowdfunding can be useful to get valuable signs on the potential of market (Schwienbacher and Larralde, 2010). It is not only for raising funds from the crowd but assists the founder and platform about the preferences of the public. The innovation in the technology industry is unstoppable and not only tech giants are involved but numerous tech startups and sole proprietor with creative ideas are there in the field.

The significant financed amount through crowdfunding in the games, design and technology campaigns indicates the interest and knowledge of backers in the industry. The large-scale technology or fintech companies are able to seed funds through venture capital and angel investor but the medium and small-scale founders face challenges and due to this these days small scale founders are seeking funds through crowdfunding. It is prominent to improve the existing models to enhance their chances to get the funds from investor. There are several types of crowdfunding that are: equity based, reward based, donation based, profit sharing, peer-to-peer, debt securities and hybrid models having certain process and considerations (Forbes and Schaefer, 2017). In the study, only reward based crowdfunding projects and its success factor are analysed with the help of machine learning models and other techniques.

1.1. Motivation and Objective

The requirement of the contribution in the present studies is to furnish the extensive understanding about the functions of crowdfunding for games, design and technology projects. The research aims to elucidate the factors deciding the positive and negative outcomes of the crowdfunding projects. The need of analysing the driving force of the successful crowdfunding project and facilitate the entrepreneurs perform better for attaining crowd funding effectively as their alternate financing option. To display the combination of finance and technology via machine learning accuracy in determining the expected results of successful projects

1.2. Research Question

The purpose of the study is to verify the research question listed below:

What are the key drivers contributing to the successful outcome of leading crowdfunding projects and how significantly does the duration of the campaign influence the state of project?

1.3. Contribution

The paper will offer valuable insights regarding to effectuality of crowdfunding as substitute financing option and implication of venture success. Systematic analysis of the crowdfunding trends will assist the investor as well as capital seeker to measure the interest of market and improve the business function before the launch. For fund

seekers, it will be beneficial to recognize the key attributes of the projects that are highly influential for the success of their project, so that while launching the project they can set the features accordingly for attaining better funds. Similarly for backers and hosting bodies as they get multiple project proposals and to select the best project for hosting as well as investing, they require proper screening, so rather than going for manual scrutiny, this machine learning model will assist the process.

1.4. Limitation of the Study

The study is concentrated on the single data source from crowdfunding platform that is Kickstarter and historical secondary data of crowdfunding used for the study, might fail to provide current trends and patterns. Further, only three categories of projects are taken into consideration making the study specific for those categorical funding. The market conditions and external economic factors affecting the crowdfunding projects are ignored.

1.5. Structure of the Study

There are diverse section and sub sections in the paper with primarily 5 sections, which includes introduction describing the context of work, then literature review section briefly mentioning and critically analysing the decisive points of the previous related researches. Following that, a methodology explaining the structured procedure, as well as tools and data insights for assessing the entire approach. The full report of the research procedure, including the method used, the results obtained, the scope, and other necessary information, is then supplied in the design and implementation part. In conclusion, the findings and recommendations for further research are described, along with future work.

2. RELATED WORK

Unavailability of initial funding acts as strong hindrance to innovation (Cosh et al.,2009). Crowdfunding is an emerging process for financing new businesses innovative ideas, helping founders and investors to connect as well as compensate each other, ranging from low to high rewards and having broad range of goals which makes it distinct from traditional funding method (Mollick, 2012). Further, Kuppuswamy and Bayus (2018) states that another reason creating difference between traditional methods

and crowdfunding is that potential backer can monitor the other supportive investor and it's timing prior deciding their funding showing the social details which will ultimately contribute in the success of campaign. Crowdsourcing and micro finance concepts are the inspiration behind the of workflow of crowdfunding but stills it stands out uniquely in the fundraising category which is assisted by multiple internet sites (Poetz & Schreier, 2012).

2.1. Ecosystem of Crowdfunding

A relatively big number of investors are supporting a campaign in crowdfunding, by providing respective funding to the creator of project and online crowdfunding platform are acting as intermediator to bring these parties together, where campaign creator develops webpage for project furnish all the relevant information which includes attractive pictures, texts, videos or formats and contributors connect with the creator via comments and then the founder aims to convince the contributor about their project worthiness for funding (Koch and Cheng 2016). The creator set a goal for funding and between the launch period and deadline, if the goal is reached then the state of project is considered as successful otherwise failed. The outcome of the projects is dependent on variable qualitative as well as quantitative factors. As per Koch and Siering (2015), there are high possibilities of not achieving the funding target if goals are set high.

Different parties involved creates diverse impacts in the crowdfunding process by forming an ecosystem, which controls the enabled functions and practices, identifying the stakeholders and their potential influence is the effective way understanding the ecosystem (Beaulieu et al., 2015). Looking at the root of the crowdfunding, it is technology enabled, making the platform provider the main character in the whole process. The website designer connects multiple sources of social media where reach of content is highly speedy and viewership is more, so that they can arrange maximum backers for the funding. Through the websites legal requirements and finalizing the deal structure are enforced making the crowdfunding website providers crucial for the whole ecosystem (Gelfond and Foti, 2012).

Founders are those who post their projects and ideas on the crowdfunding platform having variety of background and are known by numerous names such as "start-ups",

"entrepreneur", "creators", "owners" and "firms". The unfulfilled founder's desire of capital requirement derives the crowdfunding phenomenon, talking about the funds comes the equally important participant 'backers' who not just only provide the funds but also facilitate the market testing and provide opinion about the worthiness of campaign they generally fund the projects but also through their virtual network, they promote the project in the crowd for more awareness and similar to founders they also called by many names in the literature such as "investor", "funder", "contributors" and "lenders" (Beaulieu et al., 2015).

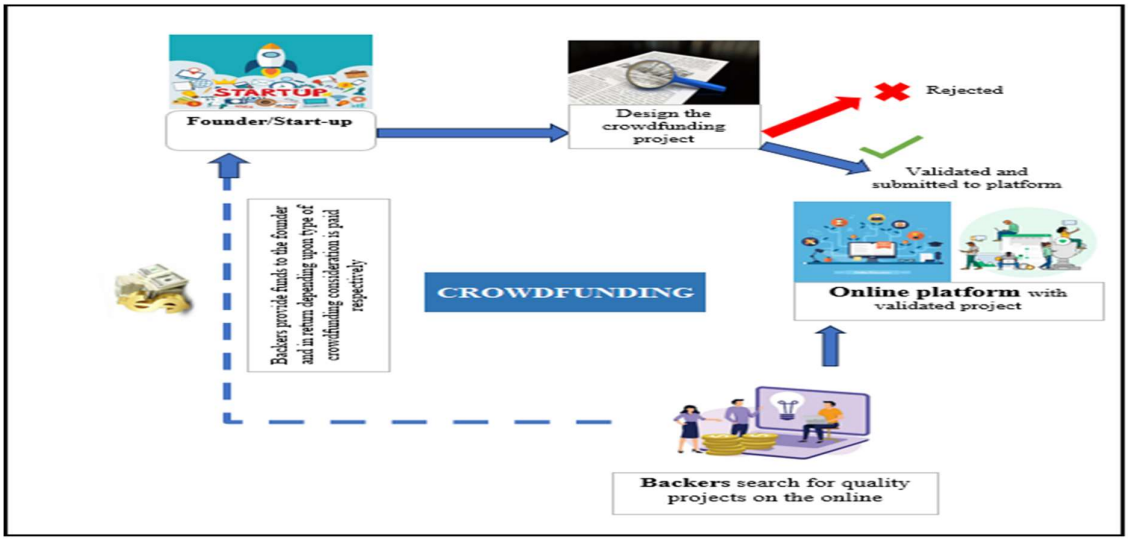

Fig 1: Crowdfunding Mechanism

2.2. Influential Factors of Crowdfunding

Desired funding level successfully reached by some projects but still some campaigns fail to get funds, resulting to founders interestingly seeking for the influential factors of successful funded projects. As per Mollick (2014), the success margin of crowdfunding project is very low and otherwise the project failed to reach the goal, the success of a project depends on the founder's network, quality of project and geographic component (in case of cultural product). However, the previous investing experience or behaviour of the backer also plays vital role in funding the upcoming projects as earlier performed activity indicates the higher knowledge and reliability in comparison with the newly formed backers but the current literature is insufficient about the influence of past funding behaviour on success of the project (Koch and Siering 2015).

As per the study conducted by Zvilichovky et al. (2013) it is evident that previous backing experience has a positive impact on the outcome of the project, they examined two factors that are number of projects create by founder and number of projects backed by the contributors but the study doesn't consider other relevant features which can be impactful for judging the outcome leading to the paper's limitation. While making funding decision information asymmetry is the main problem, backers cares about the final outcome of the project leading to the rewards but ultimately it is the backer's interest which can make the project successful even though backers don't have the in depth information and facts about the project as compared to the founder because of that limited knowledge about the project funders assess the creditability and quality of project but at the end, the funding decision lies in the hand of backer and not exactly on the facts (Koch and Cheng 2016).

The creative way to represent the ideas and information also affects the decision making of the backer and overall outcome of the project, generally the visual presentation have a powerful psychological impact, project founders who provides sufficient information in an innovative way tend to get more funding (Kranz et al., 2015). Xu et al. (2016) states that, crowdfunding mainly have two phases: raising capital and project implementation, therefore the outcome addresses two primary dimensions that are project goal reach and successful implementation of project by founders and after investigating the performance in the project implementation, it was found that sponsor involvement is highly crucial for the successful funding. The backers evaluates the estimated economic value and then takes the decision (Cordova et al,. 2015). Both the qualitative as well as quantitative aspects needs to be taken into consideration while analyzing the outcome, key factors vary with the research depending upon the quality and quantum of their collected data.

Highlighting the transformative impacts of crowdfunding on entrepreneurial finance Agarwal, Catalini & Goldfarb (2011) states that, in music industry crowdfunding has benefits but for economic transactions in early-stage projects, it diminishes the structural presence. Allison et al., (2017), while analyzing the influence effective influence in crowdfunding indicates that the quantum of funding depends on the

accounts and network of the platform operator and states that not every crowdfunding is successful with limited signaling and social capital. Early-stage projects which are considered as creative are generally funded locally, but without conducting due diligence projects which gets funds through online platform is quite surprising in terms of economics of crowdfunding (Agarwal, Catalini & Goldfarb, 2014)

In the technology related projects specifically in the reward-based crowdfunding, backers tend to fund more to the project offering the finished technology product, in comparison for services and more attracted towards the advancement and mobility in technology but in terms of small device and connecting device backers didn't show much interest (Westerlund et al,. 2021). Whereas Cordova et al. (2015) states that duration, goal and contribution frequency are key factors for successful project as well as for the project overfunding and agrees with the other researchers that there are high chances of project failure with high goals and duration of project increases the success possibilities. However, Mollick (2014) the success of project is not guaranteed by the duration. Each author has their distinct finding, somewhat contradicting as well making the study more explorable. The influential factor impacting the project's outcome can be both quantitative as well as qualitative, merely depending on the obvious factors sometimes the other qualitative economic factors are disregard.

2.3. Kickstarter

Kickstarter is one of the popular crowdfunding platforms launched in 2009 based in New York, which acts as intermediary between founders and backers, it helps developing businesses by facilitating a good chunks of cash infusion and possibility of getting wider reach, with systematic coordination of demands for distinct groups. Kickstarter mainly operates with fifteen distinct categories and operates on *all-or-nothing* model. The founders are only paid when the pre-defined goal has been exceeded, otherwise the project will be treated as failed. The online platform is well designed, each category has different tabs and project page with all the required information for better navigation. In this reward-based crowdfunding model, rewards depend on the project offering and generally ranging from $1-$10,000, which can't be

shared or offered as equity and doesn't allow any charity or other non-monetary initiatives.

The maximum duration of project is 90 days but the platform recommends 30 days for effective funding. To help the founders bring innovative projects to life is the mission of the platform and resulting to that some creators like TLC, Critical Role earned a huge fanbase. People over 23 million around the world have funded the Kickstarter campaigns. As of now, out of 15 categories, "games" has the highest funding of $2.49 billion solely generating funds 31.60%, following by "design" $1.71 billion and then "technology" successful funding of $1.40 billion out of total funding of $7.47 billion contributing 19% approx. to the total funding generated via Kickstarter platform. The state of top three categories sub-category wise are shown below:

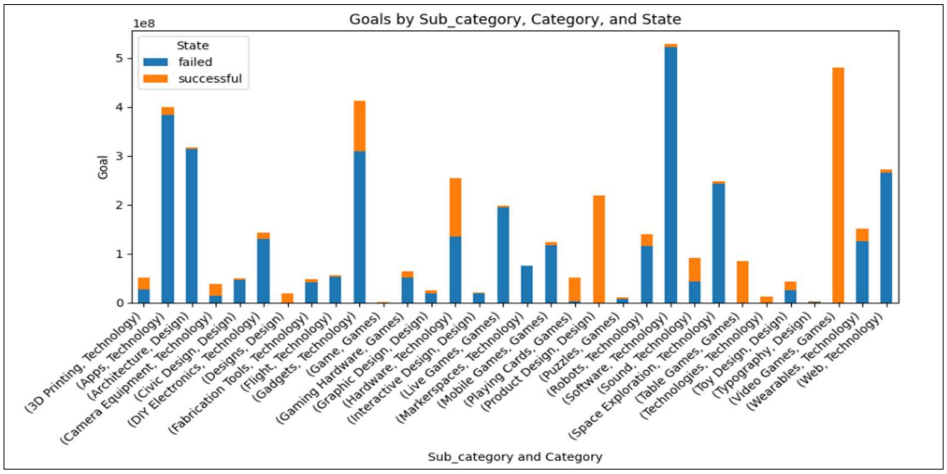

Fig 2: Leading projects on Kickstarter

2.4. Reward Based Crowdfunding

Hemer (2011) states that distinct forms can be taken by crowdfunding depending upon the contribution received by backers, commonly it is divided in four categories which are: reward-based, donation-based, lending-based and equity-based crowdfunding. Kickstarter operates on reward-based model of crowdfunding, where the creator of project has to compensate the backer with non-financial rewards, either by the finished product or service. In reward-based crowdfunding early internal and external social capital plays influence the success of the projects, and positively associated with early backers and funds (Colombo, Franzoni & Lamastra, 2015). Kunz et al., (2016) states

that project quality significantly attracts the reward-based crowdfunding, but in case of product it is undefined until the finished product is produced.

3. METHODOLOGY

Detailed description of the methods and data collected for utilization in the study, is provided in this section. The study applies Knowledge Discovery in Databases (KDD) for extraction of insightful information and involves the steps shown in the figure:

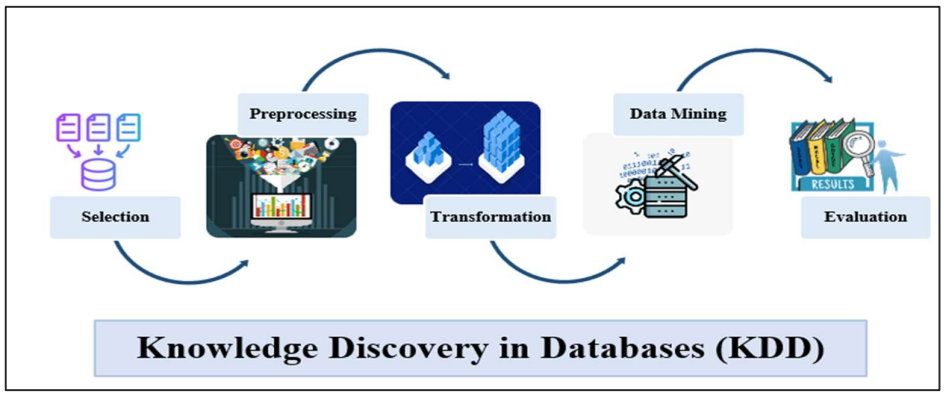

Fig 3: Flowchart-KDD

3.1. Data Selection

Data is obtained from the website called "Web Robots" supporting web crawling technologies, covering the period from 01.07.2009-15.06.2024. The initial dataset has the key information of 58446 projects and Some of the required important information is gathered from the its website. Kickstarter have 15 major project categories, and then segregated into numerous sub-categories depending on the characteristics of respective project, out of the 15 categories, the top three leading categories are selected which are: "Games", "Design" and "Technology". Further "Games" and "Design" category have 8 sub categories each and "Technology" category have 16 sub categories.

3.1.1. Variables

Majorly variables are categorized into three class: dependent variable, independent variables and control variables. The dependent variables are the variable which the model analyse or predict, whereas independent variable is the cause of the effect on dependent variables, keeping unrelated and independent from other variables. Variable

which are held control are considered as control variable in the study. For the study, there is one dependent variable and six independent key features which are briefly explained in section 4 “Design Specification”.

3.2. Data preprocessing

For aligning the data for training the machine learning models, the data needs to be modified accordingly. In the initial stage approx. 110 files of Kickstarter projects are extracted for the period 01.07.2009 to 15.06.2024 from the Web Robot website and then collated all the files to get the Masterfile with all the projects, then filtered the category and separated the datapoints for three categories (technology, game and design). Data cleaning like checking for duplicates, missing values, correlated variables, creation of new variables will be done in excel as well as in Python accordingly the requirement. To avoid any noise and biasness, normalization of data will be performed to get the balanced dataset, in case of text, specific textual analysis will be done and detailed explanation will be given in Implementation section

3.3. Predictive Modelling

For understanding the variables relationships and capturing their patterns, numerous authors have used machine learning algorithm in predictive modelling to predict outcome of crowdfunding projects, in line with earlier researches and considering the similarity of motivation, the study will be applying three machine learning techniques to predict the state of the crowdfunding project mentioned as follows:

Random Forest Classifier: It is a machine learning technique in which numerous decision trees are developed in the random forest, having distinct opinions about the data points and then examining the bunch of decision trees, the most occurrence outcome is selected and then final prediction is made (Pal, 2005). The model works on two basic principles that are :1) decision making question & answer flowchart building, 2) perception of random crowd. Division of multiple nodes gives wide variety and add randomness. The advantage of the technique is it requires only two parameters and accurately provides relative importance of variables.

Gradient Boosting and XG Boost: Machine learning problem solver and reliable ensemble technique, these algorithms fuse weak learners into powerful learner in a repetitive way and estimate the residual errors and to minimise the errors it fits the models (Bentejac et al., 2019). Further XG boost is the advanced gradient boosting enabling formation of new trees after correcting previous errors. Gradient boosting trees are considered to be more accurate and precise than random forests and capable of apprehend the complex data patterns

Logistic Regression: Logistic Regression are suitable for the dataset, where there is binary dependent variable and facilitates the data outlining and notice the relationship dependency among variables, Sigmoid function is used having curve of S- shaped with value ranging from 0-1. The weights and input combination for prediction of output (0 or 1) explained by the given equation: $y = e^{(b0 + b1*x) / (1 + e^{(b0 + b1*x))}}$ (LaValley, 2008). When the class distribution is imbalanced, there is significant impact in the performance of logistic regression

K-Nearest Neighbours- The requirement of performing discriminant analysis K-nearest-neighbour classifier was developed, generally works on the basis of Euclidean distance between test and specific training samples and should be considered as first preference for classification study, (Peterson, 2009).

Tokenization and Word Cloud: For analysing the text in the dataset, tokenization cut down the text into small chunks for smooth machine analysis, those smaller chunks are called tokens, the major reason of performing tokenization is to assist machine for understanding human language and multiple types of tokenization are used such as word tokenization, character tokenization, sub word tokenization. Here, in the study word tokenization will be used, which will break down the text into separate words and then form the word cloud with the words which are frequently present in the failed project name.

3.4. Evaluation

The study will evaluate the performance of trained machine learning models which are Logistic Regression, Gradient Boosting and Random Forest Classifier with prediction of

outcome of crowdfunding project either failed or successful. The accuracy of the models will be evaluated by confusion matrix, Mean Squared Error, Accuracy score and providing classification report. The tokenization of text will be presented with word cloud.

4. DESIGN SPECIFICATION

The structural diagram of work flow indicates the primary steps involved while conducting the study, starting with the collection and extraction of the project details from the website, preprocessing the data including cleaning duplicates value, checking for missing values, correlation checks and tuning the data as per requirement. Then, transforming the features and for text analysis tokenization is used to breaking down the text into smaller words and formed a word cloud. After fine tuning the datasets, it is divided into train-test and then predictive models of machine learning are applied. Lastly, to evaluate the performance of the applied models Confusion Matrix with Mean Squared Error, Accuracy score and AUC-ROC will be presented.

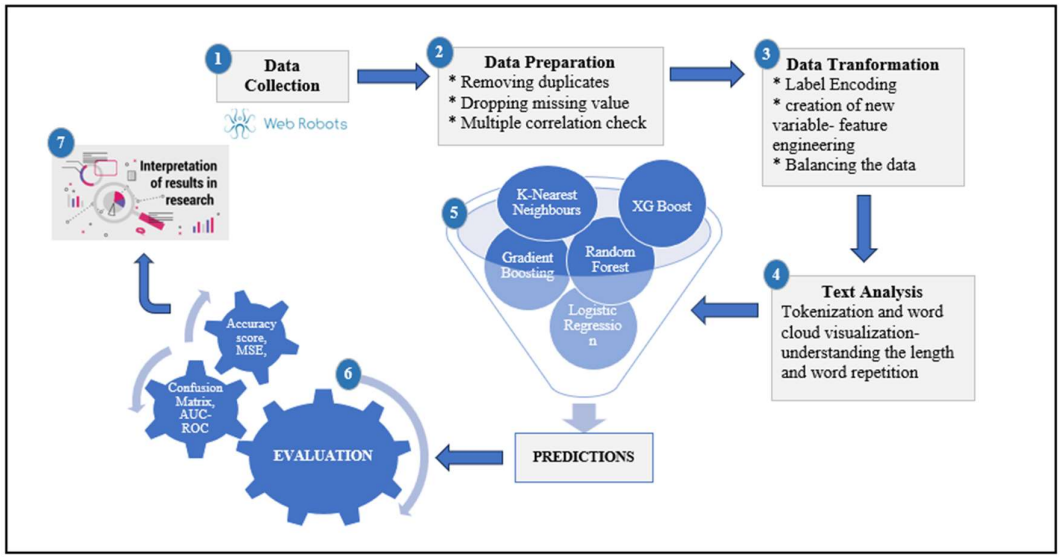

Fig 4: **Architectural diagram**

Variables Categorization: The 'state' of the project is the dependent variable, indicating the actual outcome of the project. The data have 6 types of states namely: cancelled (withdrawal by creator or rejected by platform operator), live (funding is ongoing), submitted (waiting for approval by the platform operator), successful (reached the funding goal), failed (fail to reach the goal) and suspended (project on

hold), but for the study only those projects having the state either successful (reach the defined goal) or failed (unable to get funds as per the defined goal) are considered. In the Fig 3. It is clearly seen that 54.1% project's state is successful and 45.9% contributes to the failed project in the dataset.

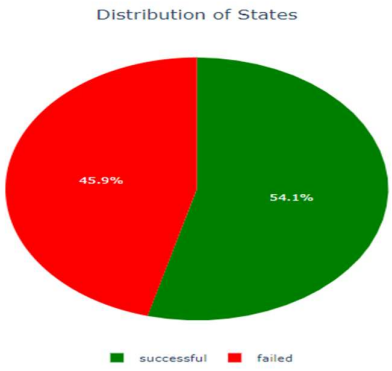

Fig 5. Distribution of dependent variable

Whereas, dependent variable is the effect or the predicted value, but to verify the effect independent variables are used. The cause-&-effect relationship is represented by these independent variables and the effect of selected independent variable on state of project (whether it is successful or failed) is the primary focus of the study. Here, the below mentioned variables are taken, after taking into consideration previous researches on success factor prediction of crowdfunding:

S. No	Variable name	Description of variable
1	**Backers Count**	The total number of capitalists providing funds to projects on the Kickstarter platform
2	**Goal**	The minimum amount which the founders of project are seeking to commence their projects
3	**Blurb and name**	Attractive taglines and names describing the project
4	**Duration**	The time period between launch and deadline of the project or the time period for which project is open for funding
5	**Staff Pick**	Recommendation of selected project by the personnel of Kickstarter
6	**Launched Month**	The month in which project is launched during the year

Table 1. Independent variables

5. IMPLEMENTATION

Data is gathered from the data scrapping website called 'Web Robots" for the period 2009 to 2024 (till 15.06.2024) with only three main categories naming 'Technology', 'Design' and 'Games' because of highest contribution of funds on the Kickstarter platform. The initial raw dataset had all the information about projects under each category, having 39 distinct features. All the extracted files were in CSV format and compilation is done in Microsoft excel of approx. 110 files. The semi-cleaned data is then uploaded in the Google Collab, where rest of analysis is performed using Python language.

Data Preprocessing and transformation: To enhance the data quality and suitability for model fitting, data preprocessing and transformation plays a vital role in the whole analysis process as if the initial data lacks quality, then the models will possibly produce inaccurate results. After filtering the data for specific three categories, only projects having more than $5000 pledged are considered, duplicates and missing values are removed, also only successful and failed projects are taken into account, resulting to reduction in the project counts from 58,446 to 33,531 projects. The data have 39 features, some of them are not relevant for the study such as 'is_starred', 'URL', 'photo', 'fx_rate', 'usd_type' and others, so after considering the usefulness of the variables only 16 variables are selected for further analysis. Further the launched and deadline dates in the raw dataset are in unix time which then converted to the normal date format.

The dataset contains both categorical as well as numerical values. Machine learning models best works with the numerical values therefore, the independent and dependent categorical variables are converted to numerical format using label encoding assigning each unique category, a number starting from '0' alphabetically, for example in case of state (dependent variable) failed and successful are assigned the numerical value '0' and '1' respectively. To simplify and effective machine learning, new variables are created such as day of launch, month of launch, year of launch, name length, duration and is_weekend breaking down the 'launched' variable which is string having timestamp in it. Then, key statistical measures are calculated including mean, median, standard deviation indicating the occurrence and variability of features as shown in Table 2.

	Count	Mean	Min	25%	Median	75%	Max	S.D
backers_count	33531	572.659748	0	5	70	377	105857	2224.22
goal	33531	138111.3	5014	10000	20000	50000	100000000	1780017.00
Duration	33531	35.5427	1	30	30	40.04	97.78	11.81
pledged	33531	235156	0	256	12101	59644	481621800	4140484.00
staff_pick	33531	0.171274	0	0	0	0	1	0.38
state	33531	0.541409	0	0	1	1	1	0.50
month_of_launcl	33531	6.553816	1	4	7	10	12	3.38
year_of_launch	33531	2018.77689	2009	2016	2019	2021	2024	2.94
launched_day	33531	2.120456	0	1	2	3	6	1.65
is_weekend	33531	0.100206	0	0	0	0	1	0.30
name_length	33531	38.434464	1	25	41	54	83	16.55

Table 2. Descriptive Statistics

Highly correlated variables create hindrances in evaluating the single coefficients because of difficulty in effect determination on dependent variable. To assess the relation and check any multiple correlation between data variables, correlation matrix is plotted for better visualization, as it can be seen in Fig 6, only currency and country are highly correlated and other variables have moderate to minimal correlation like backers count is positively correlated with state (dependent variable), pledged (0.1989), staff pick (0.2590), and duration is negatively correlated to state (-0.21) but it is acceptable.

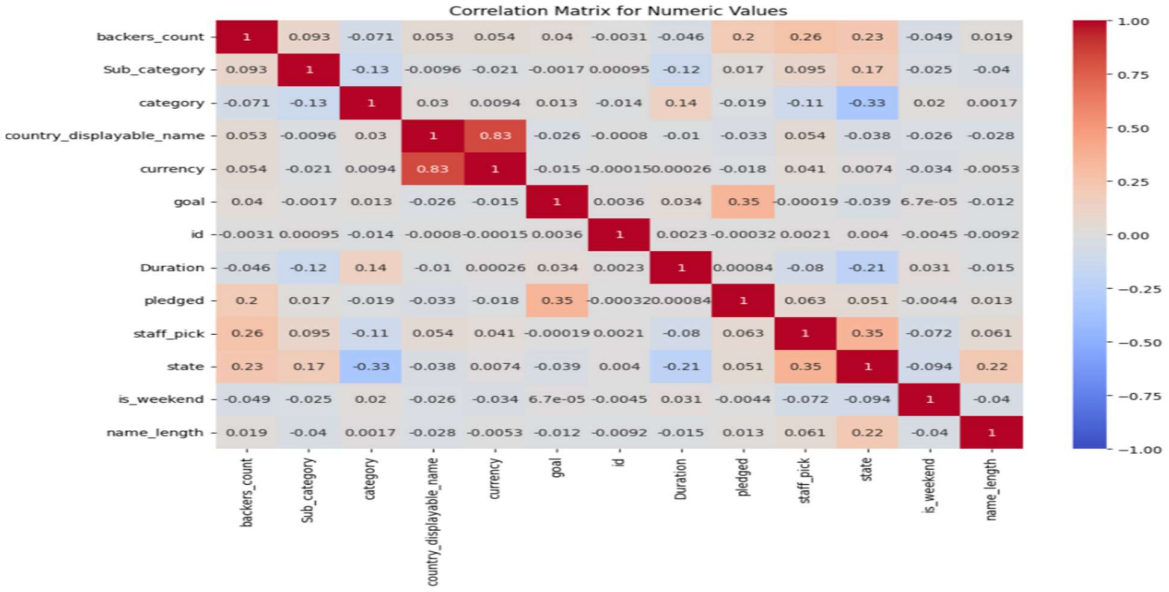

Fig 6. Correlation Heatmap

The important feature is verified and selected with the help of Mutual Information Gain one of the filter-based method indicating the materiality of variables in relation to the

dependent variable. As per the method, significant features are backers count, staff pick, duration, category and goal, feature mutual information is presented below in fig 7.

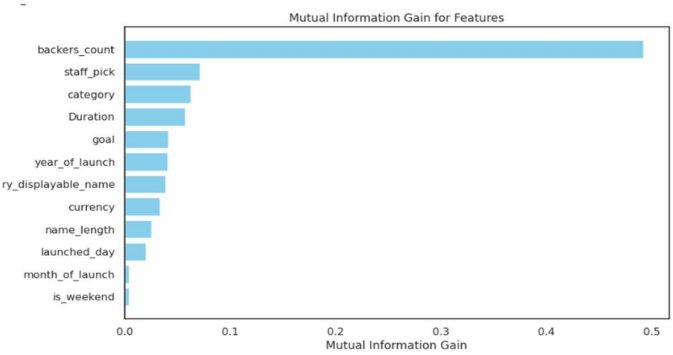

Fig 7. Feature Mutual Information

Textual Analysis: The name variable indicates the name of the project launched and each project has its distinct name. The name variable type is string and for preprocessing of text, tokenization is done which includes breaking down the textual data stream into chunks of informative words (tokens) and then the occurrence of those word is examined in failed or successful project and better presentation word cloud is formed.

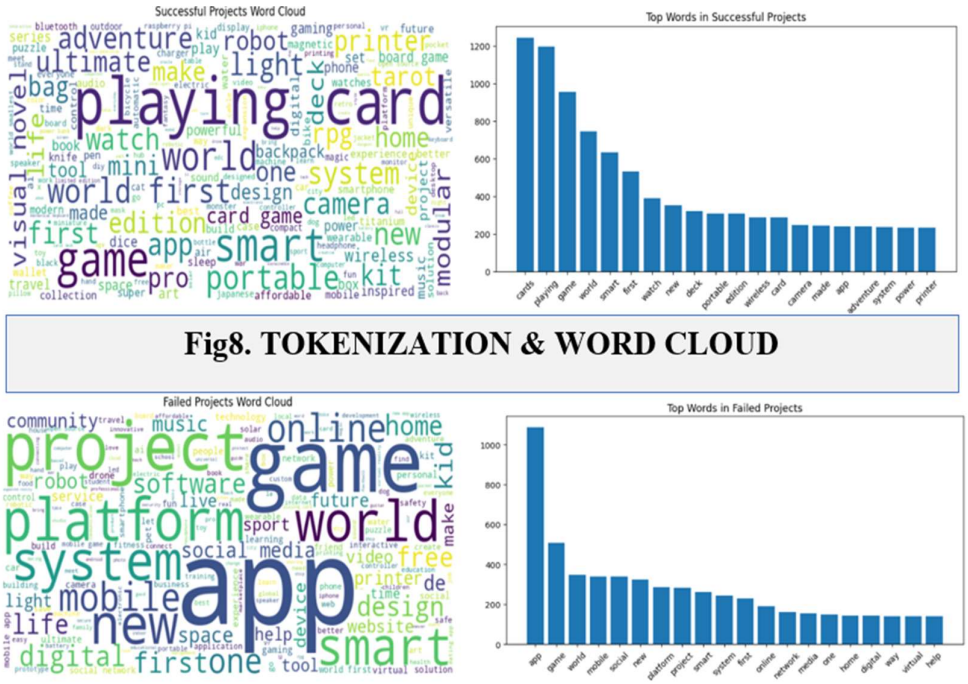

Fig8. TOKENIZATION & WORD CLOUD

Data partition and implemented models: The machine learning models are trained on 70% of the data and 30% of the data is used for testing the trained models (70:30 training-test split), the models are trained to predict the state (dependent variable) of the crowdfunding project using machine learning algorithms like RF classifier, Logistic

Regression, K-Nearest Neighbours, Gradient Boosting with XG Boost. The predicted values are then measured with the evaluation matrix contains Accuracy Score, Mean Squared Errors, AUC-ROC and confusion matrix. For analysing the name length distribution by dependent variable is considered. Detailed description of evaluation matrix of trained models is explained in the next section.

6. EVALUATION

With the predictive modelling application in the study, the primary motive is to evaluate the model performance and its ability to solve the research problem. The models are evaluated on the basis of 1) Accuracy Score: which is the ratio of accurately predicted values to the total value, 2) Mean Squared Error (MSE): average of squares of errors, 3) AUC-ROC: illustrating the binary classifier's performance, plotting the true positive and false positive rate and 4) Confusion Matrix: presenting prediction summary. The below table represents evaluation matrix of all five machine learning models:

		Least accurate		Most accurate		
Function	**Random Forest**	**Logistic Regression**	**K-Nearest Neighbour**	**Gradient Boosting**	**XG Boost**	**Remarks**
Accuracy Score	0.9267	0.9019	0.91	**0.9628**	0.9244	Higher accuracy indicates, a higher proportion of correct predictions. Gradient boosting have the highest score of 0.9628
MSE	0.0733	0.0981	0.09	**0.0732**	0.0756	Lower MSE indicates better predictive accuracy. Gradient Boosting have lowest MSE of 0.0732
AUC-ROC	0.9783	0.9615	0.9576	**0.9811**	0.9795	Higher AUC-ROC values indicate better model performance in differentiating between the positive and negative classes. Gradient Boosting have the highest (0.9811)

Table 3. Performance Matrix

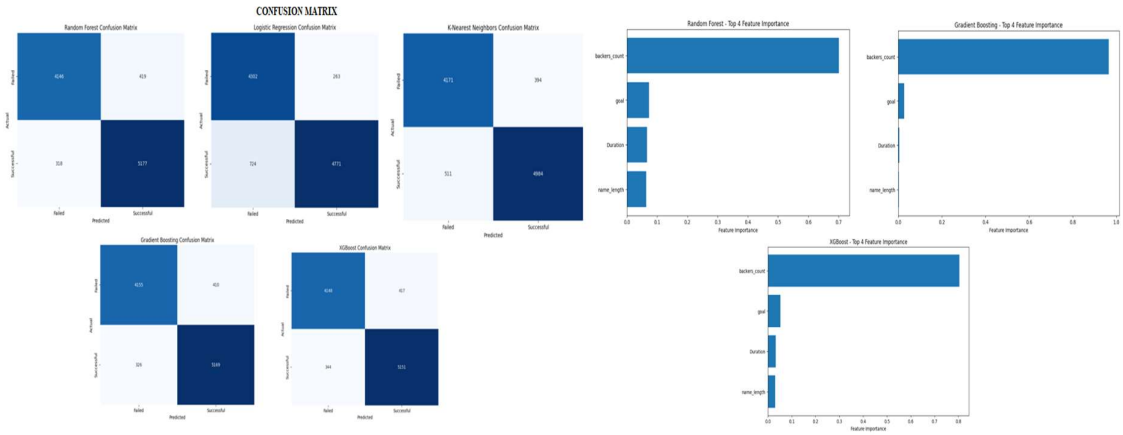

Fig 9. Confusion Matrix and Feature Importance

According to the performance matrix, it is clearly observed that out of all the models Gradient Boosting is the highest performing models having accuracy score of 0.9628 and mean squared error of 0.0732. However Logistic regression has the least accurate model predictions having high MSE indicating more errors. But irrespective of accuracy, backers count and goal feature is the most highlighting for the state of project as seen in Fig 9. Duration is the least impactful variable, and have negligible effect on dependent variable except in some exception cases. Substantial increase in probability of success with each additional backer as a greater number of backers will provide huge capital as can be observed from fig 10 the backers_count have the high importance percentage (93.25%) making it as significant out of all variable. However, goal have a negative impact on success, as the higher the goal, lesser chances to achieve, founders whiles fixing the goal should set it at a lower range, so it can be funded by backers easily rather going for higher goal and not receiving a single penny from it as Kickstarter follows *all-or-nothing* model. The creator of projects should focus to build network of backers and indulging into engagement will be highly effective. On the other hand, duration have negligible effect. In the analysis, it is noticed that there is significant difference in the backers count between successful and failed projects, some projects have good number of backers and lower goals but still the project failed indicating that although backer count is influential factor but not the ultimate deciding factor, the decision of single backer can also impact the state of project.

PERMUTATION IMPORTANCE			
Feature	**importance_mean**	**importance_std**	**importance_percentage**
Backers_count	0.404	0.0045	93.25%
Goal	0.015	0.0019	3.53%
Duration	**0.005**	0.0013	1.24%
name_length	0.004	0.0015	1.03%
staff_pick	**0.002**	0	0.53%
month_of_launch	0.001	0.0012	0.44%
is_weekend	0	0.0003	-0.03%

Fig 10. Permutation Importance

Discussing about the text in the name variable, with the help of tokenization full text is broken into individual words and those words are then analysed leading to conclusion that word ‘app’, ‘game’, ‘help’, ‘social’ and ‘mobile’ are more likely to be used in failed projects. This doesn’t mean that using these words led to failure but considering other factors like market competition and saturation in app category because if we see the fig 1 majority app category projects failed. Considering about the name length successful

project have long name with average 45-50 characters. The T-statistics is 40.1019 indicating the relative significant difference between the name length of failed and successful projects, p is 0 which is less than 0.5. Projects with having character ranging between 12-24 have higher chance of failure and project name lengths with 45-60 character tend to be successful. Psychologically, when backers want to fund a project, they want clear description and some key details in the project name.

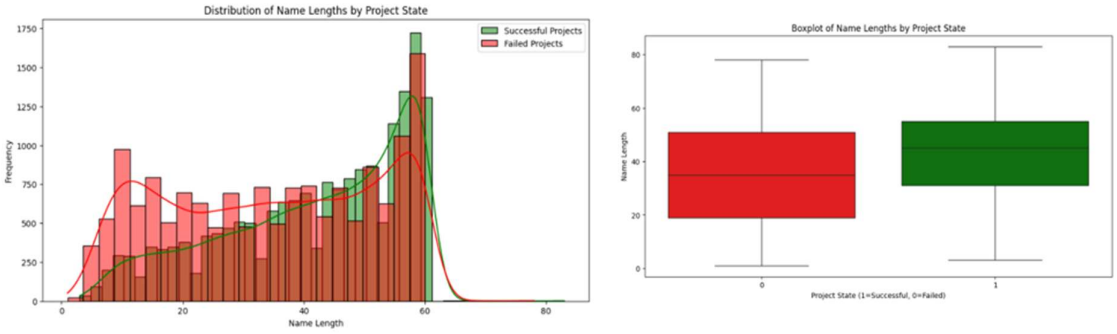

Fig 10. Distribution of Name Length of Project

Only 17% of total projects are staff picked and out of those 5743 projects selected by staff 5336 projects are successful giving the success rate of 93%, individually but wholly there is minimal impact of staff picked due to less proportion of staff picked projects. Other factors like currency, and launch month are not as impactful as other defined above, there no certain pattern is observed for them.

7. CONCLUSION & FUTURE WORK

In the study, key factors impacting the state of project are analysed and predicted, demonstrating the backers count and goal as the key driving force for the possibility of failure and success of a crowdfunding project and duration is not significantly influencing the state, only bare minimal impact is observed, agreeing with the results of other researchers (Mollick 2013, Kuppuswamy & Bayus 2013) stating that higher goals leading to the failure of the projects. However, duration typically not showed any systematic pattern to analyse the effect on the state but in some cases, it is noticed that longer duration of projects slightly reduces the probability of success. Further longer name length with average 45-50 characters, also helps the founders achieve their funding goals.

Crowdfunding builds a functional window for the research of developing ventures, as they have plenty of reason for failed or successful projects. The additional research can be done to analyse the relationship between state under different categories of projects with external market factors like funding mobile apps availability, freelance work quantum, accessibility or others, influencing the funding capacity and decision of backers, as ultimately not only core factors but economic factors also impact the decision making and capital funding of a person.

References

Agrawal, A.K., Catalini, C. and Goldfarb, A., 2011. *The geography of crowdfunding* (No. w16820). National bureau of economic research.

Agrawal, A., Catalini, C. and Goldfarb, A., 2014. Some simple economics of crowdfunding. *Innovation policy and the economy*, *14*(1), pp.63-97.

Allison, T.H., Davis, B.C., Webb, J.W. and Short, J.C., 2017. Persuasion in crowdfunding: An elaboration likelihood model of crowdfunding performance. *Journal of business venturing*, *32*(6), pp.707-725.

Beaulieu, T., Sarker, S. and Sarker, S., 2015. A conceptual framework for understanding crowdfunding. *Communications of the Association for information systems*, *37*(1), p.1.

Belleflamme, P., Lambert, T. and Schwienbacher, A., 2010, June. Crowdfunding: An industrial organization perspective. In *Prepared for the workshop Digital Business Models: Understanding Strategies', held in Paris on June* (pp. 25-26).

Belleflamme, P., Lambert, T. and Schwienbacher, A., 2014. Crowdfunding: Tapping the right crowd. *Journal of business venturing*, *29*(5), pp.585-609.

Bentéjac, C., Csörgő, A. and Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*, pp.1937-1967.

Böckel, A., Hörisch, J. and Tenner, I., 2021. A systematic literature review of crowdfunding and sustainability: highlighting what really matters. *Management review quarterly*, *71*, pp.433-453.

Colombo, M.G., Franzoni, C. and Rossi–Lamastra, C., 2015. Internal social capital and the attraction of early contributions in crowdfunding. *Entrepreneurship theory and practice*, *39*(1), pp.75-100.

Cordova, A., Dolci, J. and Gianfrate, G., 2015. The determinants of crowdfunding success: evidence from technology projects. *Procedia-Social and Behavioral Sciences*, *181*, pp.115-124.

Cosh, A., D. Cumming, and A. Hughes (2009), "Outside Entrepreneurial Capital," *The Economic Journal*, 119(540), 1494-1533.

Forbes, H. and Schaefer, D., 2017. Guidelines for successful crowdfunding. *Procedia cirp*, *60*, pp.398-403.

Gelfond, S.H. and Foti, A.D., 2012. US $500 and a click: Investing the "crowdfunding" way. *Journal of Investment Compliance*, *13*(4), pp.9-13.

Hemer, J., 2011. *A snapshot on crowdfunding* (No. R2/2011). Arbeitspapiere Unternehmen und Region.

Kickstarter. (2024). Kickstarter Statistics. Retrieved from https://www.kickstarter.com/help/stats. Accessed 15 Jul.24

Koch, J.A. and Cheng, Q., 2016. The role of qualitative success factors in the analysis of crowdfunding success: Evidence from Kickstarter.

Koch, J.A. and Siering, M., 2015. Crowdfunding success factors: The characteristics of successfully funded projects on crowdfunding platforms.

Koch, J.A. and Siering, M., 2019. The recipe of successful crowdfunding campaigns: an analysis of crowdfunding success factors and their interrelations. *Electronic Markets*, *29*(4), pp.661-679.

Kranz, T. T., Teschner, F., & Weinhardt, C. (2015). Beware of Performance Indicators. *Business & Information Systems Engineering*, 57(6), 349‑361.

Kunz, M., Englisch, O., Beck, J. and Bretschneider, U., 2016, March. Sometimes You Win, Sometimes You Learn-Success Factors in Reward-Based Crowdfunding. In *MKWI* (pp. 467-478).

Kuppuswamy, V. and Bayus, B.L., 2018. *Crowdfunding creative ideas: The dynamics of project backers* (pp. 151-182). Springer International Publishing.

LaValley, M.P., 2008. Logistic regression. *Circulation*, *117*(18), pp.2395-2399.

Mollick, E., 2012. The dynamics of crowdfunding: Determinants of success and failure. *Journal of business Venturing*, *29*(1), pp.1-16.

Pal, M., 2005. Random forest classifier for remote sensing classification. *International journal of remote sensing*, *26*(1), pp.217-222.

Peterson, L.E., 2009. K-nearest neighbor. *Scholarpedia*, *4*(2), p.1883.

Poetz, M.K. and Schreier, M., 2012. The value of crowdsourcing: can users really compete with professionals in generating new product ideas?. *Journal of product innovation management*, *29*(2), pp.245-256.

Pomeroy, C., Calzada-Diaz, A. and Bielicki, D., 2019. Fund me to the moon: Crowdfunding and the new space economy. *Space Policy*, *47*, pp.44-50.

Schwienbacher, A. and Larralde, B., 2010. Crowdfunding of small entrepreneurial ventures. *Handbook of entrepreneurial finance, Oxford University Press, Forthcoming*.

Statista. (2024) "Crowdfunding-Worldwide" Retrieved from: https://www.statista.com/outlook/dmo/fintech/digital-capital-raising/crowdfunding/worldwide Accessed 15 Jul.24

Westerlund, M., Singh, I., Rajahonka, M. and Leminen, S., 2021. Technology project summaries as a predictor of crowdfunding success. *Technology Innovation Management Review*, *11*(11-12), pp.33-44.

Xu, B., Zheng, H., Xu, Y., & Wang, T. 2016. Configurational Paths to Sponsor Satisfaction in Crowdfunding. *Journal of Business Research*, 69(2): 915-927. DOI: https://doi.org/10.1016/j.jbusres.2015.06.040