Analysis and Outcome Prediction of Crowdfunding Campaigns

Parmeet Kaur, Jaypee Institute of Information Technology, India

Sanya Deshmukh, Jaypee Institute of Information Technology, India

Pranjal Apoorva, Jaypee Institute of Information Technology, India

Simar Batra, Jaypee Institute of Information Technology, India

ABSTRACT

Humongous volumes of data are being generated every minute by individual users as well as organizations. This data can be turned into a valuable asset only if it is analyzed, interpreted, and used for improving processes or for benefiting users. One such source that is contributing huge data every year is a large number of web-based crowd-funding projects. These projects and related campaigns help ventures to raise money by acquiring small amounts of funding from different small organizations and people. The funds raised for crowdfunded projects and hence their success depends on multiple elements of the project. The current work predicts the success of a new venture by analysis and visualization of the existing data and determining the parameters on which success of a project depends. The prediction of a project's outcome is performed by application of machine learning algorithms on crowd-funding data stored in the NoSQL database, MongoDB. The results of this work can prove beneficial for the investors to have an estimate about the success of a project before investing in it.

KEYWORDS

Crowdfunding, Data Analysis, Machine Learning Algorithms, Prediction, Visualization

1. INTRODUCTION

Crowdfunding allows founders of profit, cultural and artistic ventures to raise money by acquiring small amounts of funding from many different small organizations and people. Crowdfunding is especially lucrative for young entrepreneurs aiming to materialize their novel ideas or set-up start-ups that are unable to acquire funding either because of lack of platform or lack of interest of Venture Capitalists (because of inexperience and lack of trust). Crowdfunding provides another but an unconventional method to accumulate funding, in contrast to the age-old methods of relying on government banks for loans or giving equity in the start-up. The projects open for crowdfunding belong to varied domains and categories such as theatre, comics, art, film, and music. Crowdfunding websites like Kickstarter (https://www.kickstarter.com) and Indiegogo (https://www.indiegogo.com) attract a large number of people to invest and in return get rewards or equity (Agrawal et al., 2015). This not only helps the

entrepreneurs in raising money but also helps them to market their products. All this happens online without the involvement of any middlemen, thus distinguishing it even more from other traditional platforms. Here, the investors become a part of the journey of an entrepreneur or inventor.

Crowdfunding is preferred by those who do not want to indulge in traditional methods to fund their projects or do not wish to involve the middlemen (Gerber et al., 2012). However, out of thousands of projects, it is very difficult for an investor to determine which project will be successful. Hence, it becomes imperative to identify the most influential parameters on which a project's outcome depends and subsequently determine or predict the success/failure of a project (Cordovaa et al., 2015). This work utilizes machine learning algorithms using predictive modeling techniques to predict the likely outcome of a crowdfunding venture based on whether it acquires the goal amount it needed initially. Two datasets that belong to two different types of popular crowdfunding websites have been utilized in the work. One of the websites is Kickstarter that is a reward-based crowdfunding platform while the other set of data is from the Indiegogo website, an equity-based crowdfunding platform. The collected data has been stored in multiple fragments, i.e., shards in a distributed database, MongoDB (n.d.). Distribution of data helps in reducing the load on a single server.

The stored data has firstly been visualized in order to find important factors that influence the fate of projects. Subsequently, machine learning algorithms use these crucial parameters/attributes of the data according to the type of crowdfunding. Machine learning algorithms prevent us from hard-coding old-school algorithms to reach our results because of their ability to automatically make the programs learn and adapt to the increasing dataset.(Thrun & Pratt, 2012). Machine learning techniques are highly useful in the current era of big data since they help us in quickly develop models that can aid us in analyzing huge or convoluted datasets and delivering quick and precise results even on a very large scale. Most of the industries, such as healthcare, marketing, sales, transportation, financial services, etc, working upon large amounts of data have adopted machine learning technology. The current work predicts the outcome of a crowdfunding project using the machine learning algorithms, including the most commonly used ones like Naïve Bayes, Support Vector Machine (SVM) (Ng & Jordan, 2002) and Random Forest to train and test the data for prediction of success of any given project.

The remaining portion of the paper is organized in the following manner: Section II presents the background study of the domain of crowdfunding and work done in this field. We discuss the data analysis methodology applied in Section III. Results of visualization and outcome prediction are put forth in Section IV. Lastly, we conclude the study.

2. BACKGROUND AND RELATED WORK

2.1 Crowdfunding

Crowdfunding can be simply defined as a method of raising funds for a start-up or a small venture by obtaining small sums of money from numerous groups or individuals, generally using the Internet. Social networks have been instrumental in forming the basis of Crowdfunding. Here, consumers reach out to each other through networks formed on the internet to illuminate them about their start-ups to ask for a small funding, give valuable insights regarding improvements required or to just learn about the advancements in the industries. Crowdfunding sites on the web bring together investors, who are ready to put their investments, and aspiring entrepreneurs, who wish to put forth their ideas. Many projects can be separated based on the target amount and the means of return assured to backers or investors. Today, there exists a large number of crowdfunding platforms like Kickstarter, Indiegogo (Agrawal et al., 2015), etc. The biggest benefits of such platforms include an increase in the number of professional connections, accumulate the bare minimum amount of money to start the deployment on an industrial scale, gain in confidence and increased self-worth (if the project receives acceptance and faith from investors), complete autonomy over venture without selling equity to someone else.

They also help in emulating other's success which is much easier than starting everything from scratch and to promote it on social media websites without incurring any extra costs.

Crowdfunding can be characterized into four different types (Kuppuswamy & Bayus, 2018).

2.1.1 Reward-Based Crowdfunding

This is the most common version of crowdfunding. In this, the backers or investors invest generally a small amount in order to get a reward. The reward can be in the form of a gift or some other offers related to the project. These kinds of finance techniques are helpful to young start-ups, particularly in creative fields, that are not eligible for conventional small loans from banks but are potential game-changers in the industry. Anyone, known or unknown, can contribute to a reward crowdfunding campaign. The procedure is pretty straight-forward and can easily do without any kind of help, legal or professional. Reward-based crowdfunding enables small start-ups to get their ideas up and running unaccompanied by the pressure of investor deadlines or bank loans. One of the advantages of reward crowdfunding is that the promotion on these websites allows small companies to form a prior customer base and promote their ideas or products. Kickstarter is the harbinger of reward-based crowdfunding and has paved a way for other platforms as well.

2.1.2 Equity Based Crowdfunding

Equity based crowdfunding provides investors with a small chunk of equity in the company when they provide high amount of funding to it. Investors are not rewarded, rather given some shares in the company itself. Coverage of equity crowdfunding indicates that its potential is greatest with start-up businesses. StartEngine(https://www.startengine.com/), Indiegogo, EquityBender(https://equitybender.com)and FundPass(https://fund-pass.com) are some of the platforms for equity crowdfunding.

2.1.3 Donation Based Crowdfunding

The simplest type is 'donation-based' crowd-funding. It is no-return funding where people give money to an organization, cause, charity or small company just for the gratification of contributing to a project which they think would help people and improve the quality of life generally. Since crowdfunding is based on donations, donors do not get any equity or rights of the project, nor are they seen as contributors/creditors of the project. It is usually used for social causes, cultural and art projects by non-profit companies and organizations to gather money for help. Traditionally funders donate only small sums because they are not getting any reward from it. It is typically used to collect money for educating children whose parents cannot afford it, planting trees or for individual projects which intend to help any person, group or community in general. It is mainly done for social causes (Cordovaa et al., 2015). GoFundMe(https://www.gofundme.com), Crowdfunder(https://www.crowdfunder.com) and RocketHub(https://www.rockethub.com/) are a few well known donation-based crowdfunding websites.

2.1.4 Debt Based Crowdfunding

Debt-based crowdfunding is also known as peer-to-peer lending or P2P. With this type of crowdfunding, individuals lend money to businesses or other individuals with the expectation that it will be repaid together with added interest. Debt-based crowdfunding is now recognized as a great substitute for traditional forms of investments for start-ups. It provides a huge chance for young businesses and start-ups to gain pecuniary aid and resources apart from conventional loan forms like banks and investment firms. Unlike the other forms of crowdfunding, it does not provide any reward or shares in the project. P2P lending platforms such as Funding Circle (https://www.fundingcircle.com), lending to businesses and Zopa (https://www.zopa.com/) allow lenders to earn better returns.

Crowdfunding data has been utilized in (da Cruz, 2018) for testing the hypothesis that this data can act as an information provider about new products. It is observed that the chances of a product

getting released in the market are positively affected by contributions for those campaigns that fail to garner their goal amounts. The experiments are based upon a dataset built from multiple data from the Internet. It is seen that a project campaigner may introduce the product in the market if it gets a positive valuation from the crowd sources.

A dataset of around 674 technical projects of the Kickstarter website has been analyzed by authors in (Zhang & Chen, 2018). The objective of the analysis was to determine a correlation between the success of a crowdfunding project and its supporters or backers. The results confirm that there is indeed a relation between the two. Further, the number of backers is also dependent on the reward levels of the project. Another study on Kickstarter data (Kuppuswamy & Bayus, 2017) has examined the dynamics of support rendered to a crowdfunding project over time. It is observed that more people contribute to a project if they are convinced that their contribution will lead to a project's success. Therefore, a crowdfunding project attracts increased support once it approaches its target amount. Subsequently, on meeting the goal, the rendered support decreases over time.

The elaboration likelihood model was first introduced to crowdfunding literature in this study. The influence of online information on investors is examined in (Bi et al., 2017). Authors have utilized the elaboration likelihood model to perform an empirical investigation of a Chinese website implementing crowdfunding models. It was observed that a project that displays quality and is spread through electronic publicity finds more interest in investors. Moreover, this behaviour also varied according to the project's domain or category.

An interesting application of crowdfunding in the medical domain is illustrated in (Bassani et al., 2018). The correlation between medical crowdfunding and a country's healthcare systems is examined. It has also been pointed out that health-related projects are more successful on platforms not linked with investments and more importantly if these are exclusively for projects in healthcare.

2.2 Machine Learning Algorithms

The rapid expansion of big data in fields has resulted in a need to process it with efficient techniques. Machine learning (ML) has provided novel and effective methods to handle various issues related to big data handling. Machine learning is a smaller subset of artificial intelligence (AI) that has made it possible to automatically develop models that can aid us in analyzing huge or convoluted datasets and delivering quick and precise results even on a very large scale. Machine learning algorithms are being deployed extensively for data analysis in an increasing number of applications such as health care, marketing, sales, transportation, etc. A comprehensive review of research advances in machine learning focussing on the processing of big data is presented in (Qiu et al., 2016). Authors have discussed the different machine learning approaches, their applications to big data problems along with future research directions.

Machine learning-based techniques have been utilized for multiple big data-related applications in a variety of domains. One such application is presented in (Manogaran et al., 2018) to address the problems in DNA copy number change detection. The work has employed a Bayesian hidden Markov model (HMM) along with Gaussian Mixture (GM) Clustering mechanism for representing the DNA copy number mutations across the ordering of genes. The results of experiments have established the benefit of ML techniques. Another application of ML has been put forth in (Chang et al., 2018) for analyzing huge volumes of data traffic in wireless networks. Big data analytics using machine learning allows intelligence at the network components for an improved architecture of wireless networks. The work in (Chang et al., 2018) has employed ML for caching at network edges as well as estimation of content popularity estimation. A useful application of ML to big data analytics is presented in (Nair et al., 2018) for the healthcare domain. Users are required to tweet about their health. A Spark-based application for real-time or streaming data processes these tweets to extract their health attributes. Subsequently, ML is employed to predict the status of the health of the user and this information is passed to the user. An interesting application of machine learning algorithms to prediction of factors that affect electricity market is discussed in (Romero et al., 2019).

The current paper has utilized machine learning algorithms, namely Logistic Regression, Naive Bayes, Support Vector Machine (SVM) and Random Forest algorithms to predict the result of a crowdfunding venture (Caruana & Niculescu-Mizil, 2006; Pedregosa et al., 2011). Logistic Regression is a classification algorithm that helps in analyzing the dataset which may have more than one independent variable that decides the result. This algorithm uses a binary logistic model in order to roughly calculate the probability of a binary response depending on a few independent features or characteristics. (Cordovaa et al., 2015). The predictor variables are independent variables that decide the outcome of the variable referred to as the dependent variable by determining the relationship between the dependent and the group of predictor variables. The dependent variable or the outcome variable is binary, i.e., it assumes the value of true or false. Naive Bayes is a probability-based classifier that is essentially built on the Bayesian Theorem. It allocates class labels to problem instances, which are illustrated as vectors of the characteristic's values, where the class labels are drawn from some finite set. It predicts the probability of a given record belonging to each of these classes. The class for which the probability is maximum for a record is regarded as the most probable class. Naïve Bayes Algorithm has been used widely in the real-world and has many practical applications such as categorization of news, email spam detection, face recognition, medical diagnosis and weather prediction (Williams et al., 2006). SVM is a supervised learning algorithm used for classification. SVM works by segregating the two possible classes to which a record may belong using a hyperplane. In other words, it finds the best hyperplane that divides the data into 2 classes. SVMs are very helpful in text categorization, image classification, financial forecasting, etc. (Trafalis & Ince, 2000). The algorithm has also been applied in biological and other sciences (Yu et al., 2010) as it has produced extremely optimal and useful results in solving classification questions in bioinformatics and also in other biomedical disciplines. Random Forest is an ensemble classifier that creates various decision trees, by operating on a subset of sample data and features which are picked arbitrarily. It has produced amazing results and accuracy in the field of remote sensing, analyses of diabetic retinopathy and many more fields (Casanova et al., 2014).

3. DATA ANALYSIS METHODOLOGY

The current work analyzes the crowdfunding data from two perspectives. Firstly, the parameters or factors upon which the success of a crowdfunding project depends are looked into and results are visualized with suitable plots. Secondly, machine learning algorithms are used to train and subsequently test whether a crowdfunding project is likely to be successful or not. Thus, the objective of the work is to benefit both entrepreneurs who start a project to gauge the success determining factors. Additionally, this work will aid potential investors to contemplate where to invest for better results.

3.1 Data Collection and Storage

For the current analysis, data was collected from two crowdfunding platforms, namely Kickstarter, a reward-based crowdfunding website and Indiegogo. Data on projects was collected from these websites for a period over one year using the beautifulsoup library of python language (Crummy, n.d.). Thus, two datasets have been used, the first containing 45816 projects extracted from Kickstarter and the other containing 15248 projects extracted from the Indiegogo crowdfunding platform. The projects belong to various categories ranging from art, movies to comics. The dataset consists of various attributes for these projects such as category, duration of the project, the goal of the project (in terms of amount), number of investors, reward levels, number of comments, location, etc.

Considering the size and semi-structured nature of the data, it has been stored in a NoSQL database system, MongoDB (Győrödi et al., 2015; Truică et al., 2013). MongoDB is used since it is a schema free database suited to the considered datasets. Moreover, it is easy to scale since it provides horizontal fragmentation or sharding of documents. A shard is a horizontal fragment or a subset of records of the data. MongoDB shards or distributes the documents in a collection using an

attribute(s) of data known as the shard key. Only one shard key is permitted per sharded collection. In this work, the shard key is the "category" of the projects.

3.2 Data Analysis

From the background study and collected data, it was observed that several attributes have a huge part in deciding the fate of a project (failure/success)(Hekman & Brussee, 2013; Mollick & Kuppuswamy, 2014). The features which initiate a user to invest in projects can be categorized into 3 categories- projects based traits, location-based traits, and network-based traits. Project-based traits include the quality and novelty of the project, its category, goal amount requested, time period of the project, etc. Of these, the category of a project plays a significant role. Though the investors tend to invest in the projects belonging to the category of their interests it has been seen that the Film and music category attracts them the most as compared to other ones.

Crowdfunding makes an amalgamation of social media/network engagement with entrepreneurial activities, where the consumers act in an unconventional job, i.e. of investors. Varied from traditional investors in venture capital companies, consumers are likely to popularize their invested projects on social media among their family, friends and professional connections. Therefore, social networks like Twitter, LinkedIn, Instagram, Facebook, and Google Plus acts as a crucial factor in popularizing these crowdfunding projects. Moreover, the progress of a campaign can be tracked by not only the company but as well as the people who have invested in it. Interaction on the various crowdfunding platforms not only helps in establishing multiple connections with people but also replicates the successful experiences of others.

Although a crowdfunding project can be located in any part of the world and there are no restrictions to the location, location-based attribute plays some role in deciding the fate of the project. Crowdfunding platforms brings with itself obstacles along with a huge chance for entrepreneurs and investors who may not be located at the most optimum places for investment and growth.(Cheng et al., 2016; Greenberg et al., 2013). For instance, locally proximate investors have more mutual friends and are connected with not only other investors but with friends of the fund seekers too. This makes it possible to further expand the reachability of the project However, for distant investors, this may not be possible. Their investments and choices rely on the investments made by others. Another important factor is the sum of money accumulated till that date by a project. This value portrays how much investors really believe in the success or failure of the project as well as its worth. It shows the people's inclination towards a project.It also depicts the extent of social connect a fund seeker has.

With the above considerations, visual or graphical analysis of the data was performed with the help of Tableau software (n.d.) and python library plot.ly. Data visualization (Chabot, 2008; Romero et al., 2019) aided in determining the parameters that play the most significant role in the crowdfunding projects' outcome. For the Kickstarter data, the identified parameters were project category, number of backers, reward levels and number of comments. Three attributes were found to be of the maximum importance for the Indiegogo crowdfunding projects namely, the amount of backers, category of the project and the social media impact. The detailed results of data visualization are presented in Section IVA.

3.3 Prediction

In order to predict the success/failure of crowdfunded projects for both Kickstarter and Indiegogo projects (live or already finished), algorithms namely Naïve Bayes, Logistic Regression, SVM and Random Forest have been used considering the most important factors as mentioned above. These algorithms are then compared on the metric of accuracy. Results are discussed in the next section IVB.

4. RESULTS AND DISCUSSION

4.1 Data Visualization

Since Kickstarter and Indiegogo are websites built on different reward strategies, it was expected that the parameters affecting the outcome of a project will vary in these. Therefore, data from these platforms was visualized separately.

4.1.1 Kickstarter Dataset

Firstly, we analyzed the effect of number of comments garnered by a Kickstarter project on its final status.

Figure 1 depicts that successful projects get the maximum number of comments. This indicates that there is a direct proportion between social media popularity and the success of a project since a greater number of investors would be likely to invest in it. The more the number of comments, the more the popularity of the project and hence more chances of it being successful.

Another important factor was found to be the amount of backers of the project. If an investor is considering two projects, say A and B, for investment and finds that the project B has currently a significantly higher number of backers than project A, it is highly probable that he will invest in project B. This is termed as herd behaviour in social networking (Cheng et al., 2016; Hekman & Brussee, 2013) terms and is an important aspect in crowdfunding too. Therefore, we analyzed which categories of projects get maximum backers. Fig 2 depicts the number of backers in each category along with the status of these projects. It can be observed that most number of backers was found in successful projects in category of games.

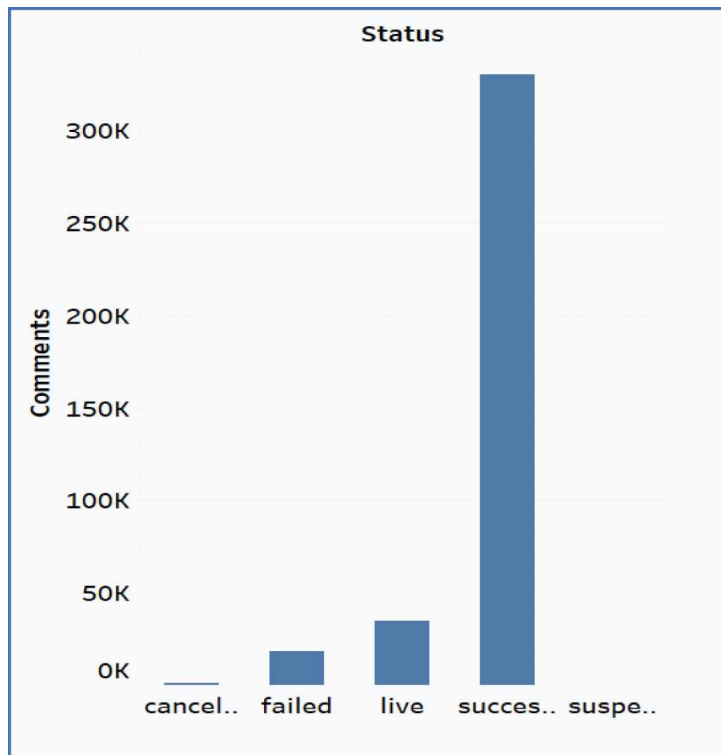Figure 1. Number of comments vs status of project

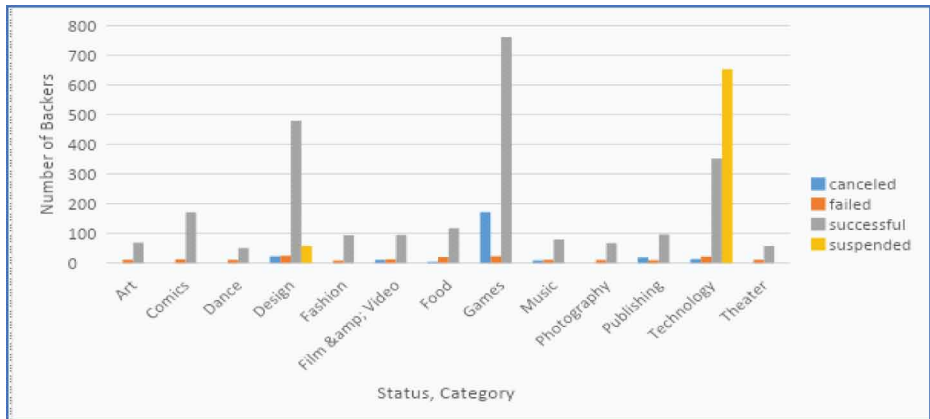**Figure 2. Backers vs category with status**

Figure 3 depicts the summation of the goal amount of all the projects for each category and it is observed that the maximum goal amount is observed for the "Film and Video" category.

Figures 4 a and b illustrate the reward levels for successful and unsuccessful projects respectively category-wise. It may be observed that failed projects offered higher levels of reward than the successful projects.

4.1.2 Indiegogo Dataset

Figure 5 illustrates the number of successful and unsuccessful ventures on the Indiegogo crowdfunding website. In all categories of projects, the number of successful projects is significantly higher than the number of failed projects.

Similar to Kickstarter, successful Indiegogo projects also observe a higher number of comments as compared to the ones that fail subsequently. Fig. 6 is a category wise depiction of the influence of the amount of comments on the fate of a project.

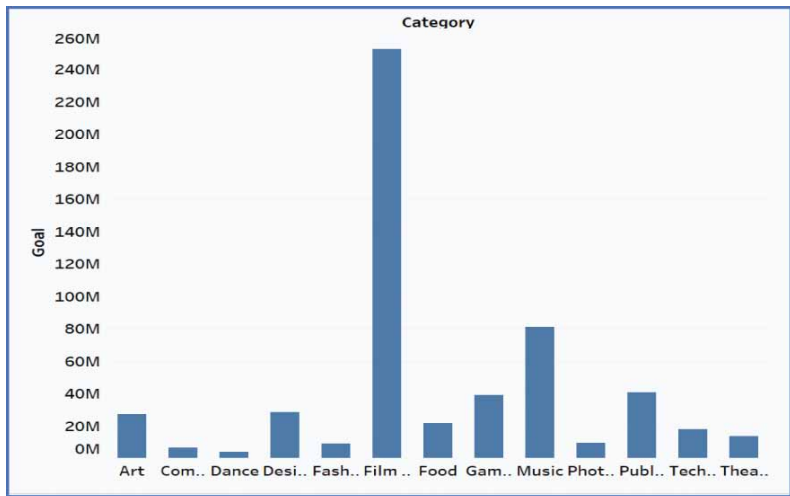**Figure 3. Category vs the sum of the goal (amount)**

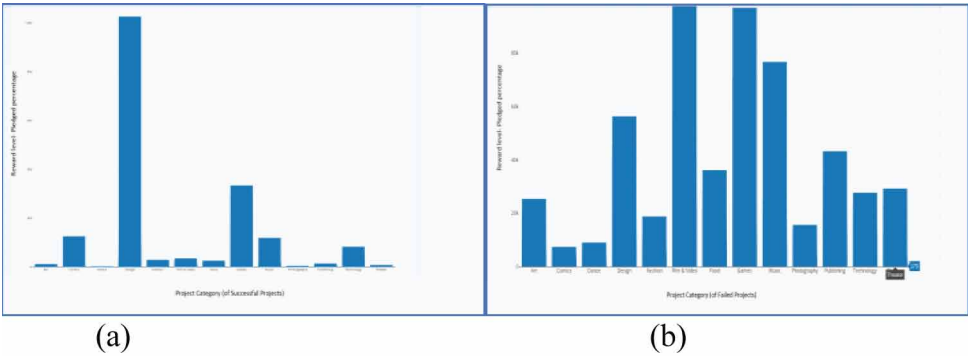**Figure 4. Reward vs category for successful and failed projects**

(a) (b)

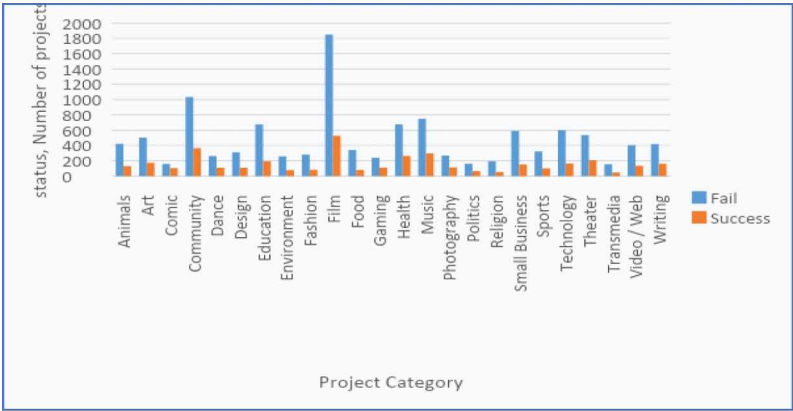**Figure 5. Category wise successful and failed projects**

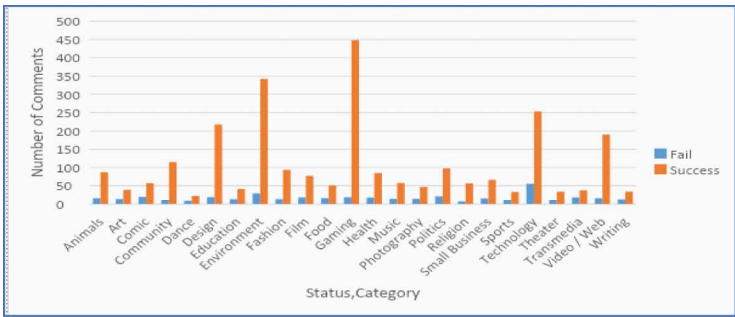**Figure 6. Number of comments vs. status of projects**
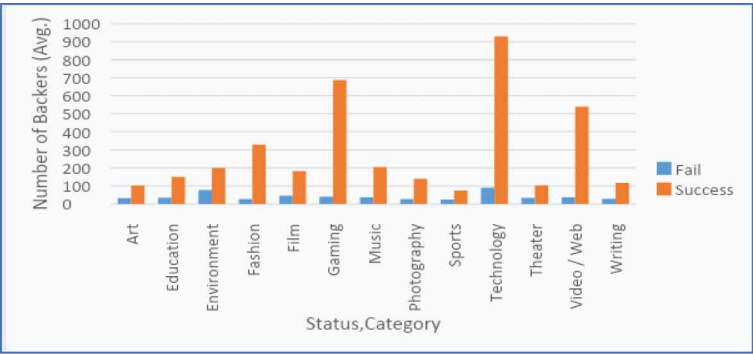

Figure 7 depicts that the projects belonging to the Technology category altogether received the maximum number of backers. Further, where the amount of backers is high (as in Technology and gaming categories), the count of successful projects exceeds the count of failed projects by huge margins.

Figure 7. Number of backers vs Category

Next, we examined each category of projects for goal amounts. It is seen that the projects of the Film category on Indiegogo set the maximum goal amount (Figure 8).

Category-wise distribution of funders is depicted in Figure 9.

4.2 Prediction

The algorithms used for prediction of success or failure of the crowdfunding projects are Support Vector Machines(SVM), Random Forests(RF), logistic regression and Naïve Bayes. The reason for selecting these algorithms is because these algorithms have varying complexity and hence, provide a clear and diverse comparison. The two datasets from both Kickstarter(a reward-based crowdfunding website) and Indiegogo(an equity-based crowdfunding website) are used in these experiments. For the prediction, we have utilized crucial parameters out of all the parameters that influence the success/ failure of projects the most.

For the Kickstarter data, after studying the literature and visualization, it was concluded that four parameters or factors are the most important ones which are – the category of the project, number of backers, reward level, number of comments on social media (social media activeness). For the Indiegogo data, we used its important parameters, namely the number of backers, category of project, and comments on social media. The accuracy for the Logistic regression algorithm, when running on the dataset, was 90% whereas that of Naïve Bayes was found to be 66%. Also, the SVM algorithm performs equally as the Logistic Regression algorithm as shown in Figure 10.
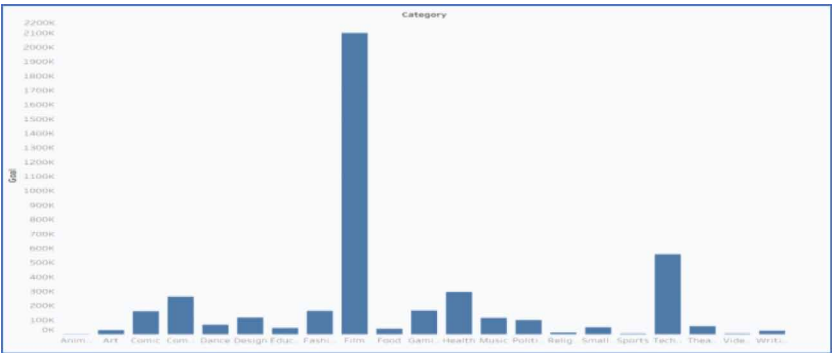
Figure 8. Category vs the sum of the goal (amount)

Figure 9. Percentage of funders for various categories of projects for Indiegogo data
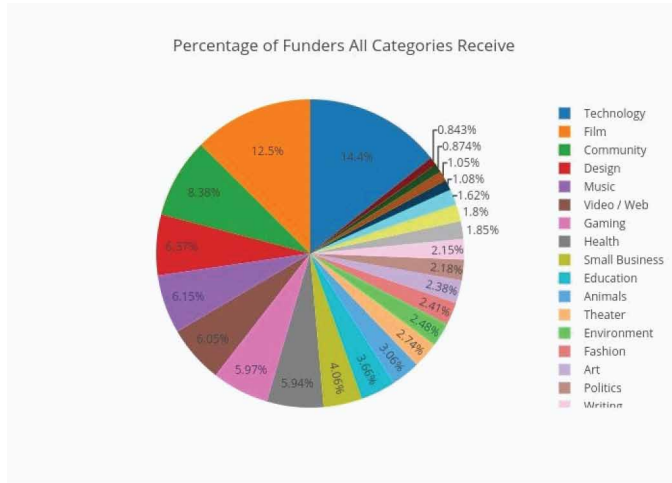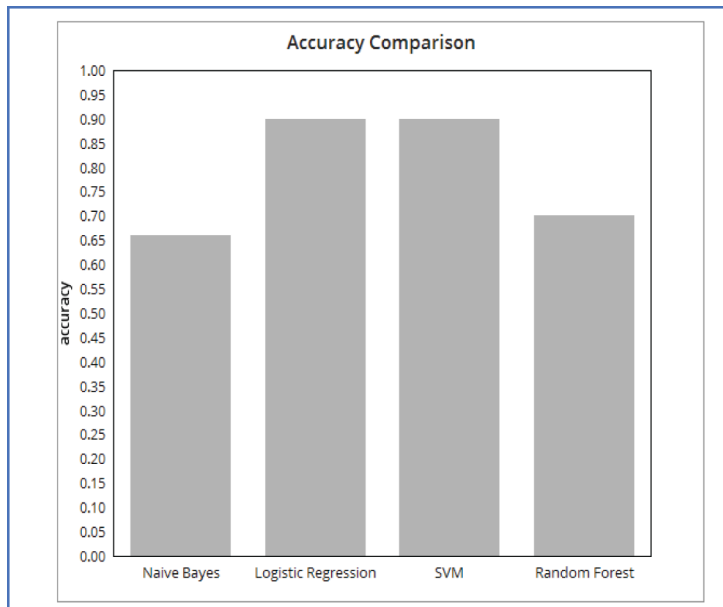

Figure 10. Comparison of accuracy of different algorithms for Kickstarter data


For the Indiegogo data, the Random Forest algorithm performed best with an accuracy of 70% followed by Logistic Regression having an accuracy of 63% (Figure 11).

The following observations have been made:

- Both Naïve Bayes and Logistic Regression algorithms can predict the future of the project in acquiring the funding proposed initially with high accuracy.
- For the Kickstarter data, the performance of Logistic regression and SVM was better than Naïve Bayes and Random Forest algorithm. The former yielded an accuracy of around 90% as compared

Figure 11. Accuracy comparison of algorithms for Indiegogo data
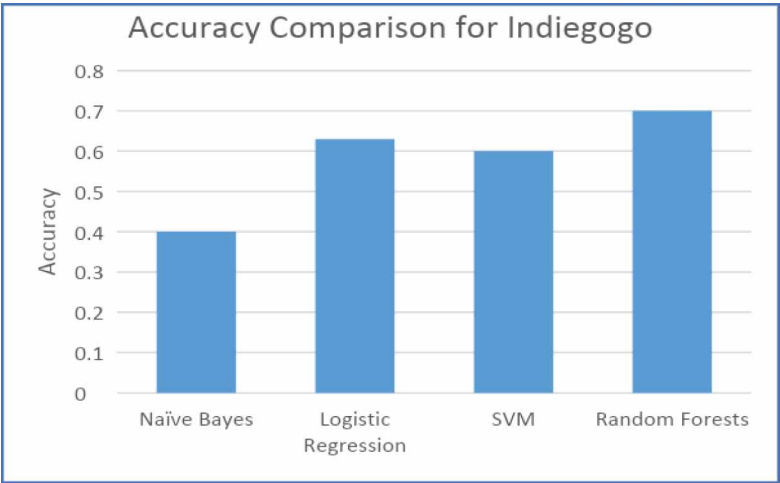
to 65% accuracy of Naïve Bayes. Random Forest performed better as compared to Naïve Bayes with an accuracy of around 70%.

- Generally projects with higher reward levels are funded the most by small investors because of the hope of getting substantial returns. Hence, it turns out to be the most important training feature among all.
- Higher the number of backers, the more is chance of getting a funding.

5. CONCLUSION

The paper has analyzed data on two types of crowdfunding projects. It was observed that social media has a very crucial impact and role in the success of a project. The more popular a project is on the website or the social media platforms, the more the probability of it being successful. Category of the projects observed to be another important attribute. The Film and Entertainment category tends to attract the maximum number of investors. Further, the reward level of the project also influences the outcome of a project. The investors find projects with higher reward levels more appealing. Investors feel that their contribution is substantial in a small project in comparison to huge ones as it would provide them more appreciation and credit. Hence It was discovered that investors were inclined to support smaller companies or even a single individual. This work can be utilized to predict how the project's outcome would be in terms of target amount raised by the project prior to the investment. This would prove to be helpful as they can invest their time and money wisely. Crowdfunding will gain more attention soon as it is estimated to prove to be a promising part of the fundraising society.

REFERENCES

Agrawal, A., Catalini, C., & Goldfarb, A. (2015). Crowdfunding: Geography, Social Networks, and the Timing of Investment Decisions. *Journal of Economics & Management Strategy*, *24*(2), 253–274. doi:10.1111/jems.12093

Bassani, G., Marinelli, N., & Vismara, S. (2018). Crowdfunding in healthcare. *The Journal of Technology Transfer*, 1–21.

Bi, S., Liu, Z., & Usman, K. (2017). The influence of online information on investing decisions of reward-based crowdfunding. *Journal of Business Research*, *71*, 10–18. doi:10.1016/j.jbusres.2016.10.001

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161-168). ACM. doi:10.1145/1143844.1143865

Casanova, R., Saldana, S., Chew, E. Y., Danis, R. P., Greven, C. M., & Ambrosius, W. T. (2014). Application of random forests methods to diabetic retinopathy classification analyses. *PLoS One*, *9*(6), e98587. doi:10.1371/journal.pone.0098587 PMID:24940623

Chabot, C. (2008). Demystifying visual analytics. IEEE Computer Graphics and Applications, 29(2), 84-87. doi:10.1109/MCG.2009.23

Chang, Z., Lei, L., Zhou, Z., Mao, S., & Ristaniemi, T. (2018). Learn to cache: Machine learning for network edge caching in the big data era. *IEEE Wireless Communications*, *25*(3), 28–35. doi:10.1109/MWC.2018.1700317

Cheng, M., Sriramulu, A., Muralidhar, S., Loo, B. T., Huang, L., & Loh, P.-L. (2016). Collection, exploration and analysis of crowdfunding social networks. In *Proceedings of the Third International Workshop on Exploratory Search in Databases and the Web*, (pp. 25-30). ACM. doi:10.1145/2948674.2948679

Cordovaa, Dolcib, & Gianfratec. (2015). Determinants of Crowdfunding Success: Evidence from technology projects. *3rd International Conference on Innovation, Technology & Management*.

Crummy. (n.d.). https://www.crummy.com/software/BeautifulSoup/doc

da Cruz, J. V. (2018). Beyond financing: Crowdfunding as an informational mechanism. *Journal of Business Venturing*, *33*(3), 371–393. doi:10.1016/j.jbusvent.2018.02.001

Gerber, E. M., Hui, J. S., & Kuo, P.-Y. (2012). Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms. *Proceedings of the International Workshop on Design, Influence, and Social Technologies: Techniques, Impacts and Ethics*, 2(11).

Greenberg, M. D., Pardo, B., Hariharan, K., & Gerber, E. (2013). Crowdfunding support tools: predicting success & failure. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 1815–1820). ACM. doi:10.1145/2468356.2468682

Győrödi, C., Győrödi, R., Pecherle, G., & Olah, A. (2015). A comparative study: MongoDB vs. MySQL. In *2015 13th International Conference on In Engineering of Modern Electric Systems (EMES)*, (pp. 1-6). IEEE. doi:10.1109/EMES.2015.7158433

Hekman & Brussee. (2013). *Crowdfunding and online social networks*. Academic Press.

Kuppuswamy, V., & Bayus, B. L. (2017). Does my contribution to your crowdfunding project matter? *Journal of Business Venturing*, *32*(1), 72–89. doi:10.1016/j.jbusvent.2016.10.004

Kuppuswamy, V., & Bayus, B. L. (2018). Crowdfunding creative ideas: The dynamics of project backers. In *The Economics of Crowdfunding* (pp. 151–182). Palgrave Macmillan. doi:10.1007/978-3-319-66119-3_8

Manogaran, G., Vijayakumar, V., Varatharajan, R., Malarvizhi Kumar, P., Sundarasekar, R., & Hsu, C.-H. (2018). Machine learning based big data processing framework for cancer diagnosis using hidden Markov model and GM clustering. *Wireless Personal Communications*, *102*(3), 2099–2116. doi:10.1007/s11277-017-5044-z

Mollick & Kuppuswamy. (2014). *After the campaign: Outcomes of crowdfunding*. Academic Press.

MongoDB. (n.d.). https://www.mongodb.com/

Nair, L. R., Shetty, S. D., & Shetty, S. D. (2018). Applying spark based machine learning model on streaming big data for health status prediction. *Computers & Electrical Engineering*, *65*, 393–399. doi:10.1016/j.compeleceng.2017.03.009

Ng, A. Y., & Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in Neural Information Processing Systems*, 841–848.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Blondel, M. et al.. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Qiu, J., Wu, Q., Ding, G., Xu, Y., & Feng, S. (2016). A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, *2016*(1), 67. doi:10.1186/s13634-016-0355-x

Romero, Á., Dorronsoro, J.R., & Díaz, J. (2019). Day-Ahead Price Forecasting for the Spanish Electricity Market. *International Journal of Interactive Multimedia and Artificial Intelligence, 5*, 42-50.

Tableau. (n.d.). https://www.tableau.com

Thrun, S., & Pratt, L. (Eds.). (2012). *Learning to learn*. Springer Science & Business Media.

Trafalis, T. B., & Ince, H. (2000). Support vector machine for regression and applications to financial forecasting. *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, 6, 348-353. doi:10.1109/IJCNN.2000.859420

Truică, C. O., Boicea, A., & Trifan, I. (2013). CRUD operations in MongoDB. In *Proceedings of the 2013 international Conference on Advanced Computer Science and Electronics Information*. Atlantis Press. doi:10.2991/icacsei.2013.88

Williams, N., Zander, S., & Armitage, G. (2006). A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. *Computer Communication Review*, *36*(5), 5–16. doi:10.1145/1163593.1163596

Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: The case of diabetes and pre-diabetes. *BMC Medical Informatics and Decision Making*, *10*(1), 16. doi:10.1186/1472-6947-10-16 PMID:20307319

Zhang, H., & Chen, W. (2018). Crowdfunding technological innovations: Interaction between consumer benefits and rewards. *Technovation*.

Parmeet Kaur is currently working in Jaypee Institute of Information Technology, Noida as an Assistant Professor (Senior grade) and has an academic experience of over 15 years. She received PhD(Comp Engg) from NIT Kurukshetra in 2016, M.Tech.(CSc) from Kurukshetra University in 2008 and B.E.(Hons)(CSc & and Engg.) from P.E.C., Chandigarh in 1998. She is presently supervising 3 doctoral students. Her research interests include distributed systems, cloud computing, big data, and security.