Analyze the role of Image Networks and Image Characteristics on the success of Kickstarter projects

Author

Imran Abdullah

Registration No. 00000327476

Supervisor

Prof Dr. Hammad Afzal

Department of Software Engineering

A thesis submitted to the faculty of Computer Science Engineering Department Military College of Signals, National University of Sciences and Technology, Islamabad, Pakistan in partial fulfilment of the requirements for the degree of MS in Software Engineering

(June 2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS Thesis written by **Mr Imran Abdullah**, Registration No. **00000327476**, of **Military College of Signals** has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulations/MS Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members and local evaluators of the scholar have also been incorporated in the said thesis.

Signature: HadAfzal

Name of Supervisor Prof Dr Hammad Afzal

Date: 23-6-2024

Signature (HOD): Adnan

Brig
Head of Dept of CSE
Mil College of Sigs (NUST)

Date: 24/6/24

Signature (Dean/Principal)

Date: 5/7/24

Brig
Dean, MCS (NUST)
(Asif Masood, Phd)

Form TH-4

NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by **Imran Abdullah,** Regn No **00000327476** Titled: "**Analyze the Role of Image Networks and Image Characteristics on the Success of Kickstarter Projects**" be accepted in partial fulfillment of the requirements for the award of **MS Software Engineering** degree.

Examination Committee Members

1. Name: **Prof Dr. Naima Iltaf** Signature: Naima

2. Name: **Lt Col Khawir Mahmood** Signature:

Supervisor's Name: **Prof Dr. Hammad Afzal** Signature:

Date: 12/6/2024

Brig
Head of Dept of CSE
Mil College of Sigs (NUST)

Head of Department

24/6/24
Date

COUNTERSIGNED

Brig
Dean, MCS (NUST)
(Asif Masood, PhD)

Dean

Date: 3/7/24

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled "**Analyze the role of image networks and image characteristics on the success of Kickstarter projects**" was conducted by **Mr Imran Abdullah** under the supervision of **Prof Dr. Hammad Afzal.** No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Computer Software Engineering** in partial fulfillment of the requirements for the degree of **Master of Science in Software Engineering** MCS, National University of Sciences and Technology, Islamabad.

Student Name: Imran Abdullah Signature: ______

Examination Committee:

a) External Examiner 1: Prof Dr. Naima Iltaf Signature: ______

(Designation and Office Address)

b) External Examiner 2: Lt Col Khawir Mahmood Signature: ______

(Designation and Office Address)

Name of Supervisor: Prof Dr. Hammad Afzal Signature: ______

Name of Dean/HOD: Brig Adnan Ahmed Khan, PhD Signature: ______

Brig
Head of Dept of CSE
College of Sigs (NUST)

Plagiarism Undertaking

I solemnly declare that research work presented in the thesis titled “Analyze the role of Image Networks and Image Characteristics on the success of Kickstarter projects” is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature:

Name: **Imran Abdullah**

Date: **12/06/24**

Author’s Declaration

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at MS Software Engineering at Military College Of Signals (MCS) or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at MCS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project’s design and conception or in style, presentation and linguistics which has been acknowledged.

Author Name: **Imran Abdullah**

Signature:

Dedication

This thesis is dedicated to all the deserving children who do not have access to quality education especially young girls.

Acknowledgments

Glory be to Allah (S.W.A), the Creator, the Sustainer of the Universe. Who only has the power to honour whom He please, and to abase whom He please. Verily no one can do anything without His will. From the day, I came to NUST till the day of my departure, He was the only one Who blessed me and opened ways for me, and showed me the path of success. Their is nothing which can payback for His bounties throughout my research period to complete it successfully.

Abstract

This thesis investigates the dynamics of crowdfunding success in Kickstarter's technology category using a blend of network analysis and machine learning. By sourcing a comprehensive dataset from Webrobots.io[20], which includes detailed monthly data dumps of Kickstarter projects, this study focuses on **U.S.!** (U.S.!)-based technology projects to uncover patterns and factors contributing to successful crowdfunding outcomes.

The methodology encompasses rigorous data preprocessing, advanced feature engineering, and graph-based modeling to represent relationships between projects, creators, and subcategories. Network analysis techniques, including centrality measures and community detection, identify influential projects and creators and uncover clusters with similar characteristics. A RandomForest machine learning model, integrating project-specific metrics and network-derived features, predicts project success with high accuracy.

Findings reveal significant patterns in project features, creator influence, and community structures that impact crowdfunding success. The predictive model serves as a practical tool for guiding project development and marketing strategies. This thesis enhances the understanding of crowdfunding platforms through the integration of network science and predictive analytics, offering insights for creators, backers, and platform moderators. The enriched dataset and methodologies are made publicly accessible to encourage further research and practical applications in crowdfunding.

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction and Motivation

1.1 Introduction and Motivation

In the contemporary era of digital transformation, crowdfunding platforms have emerged as pivotal mechanisms for project financing, particularly within the realm of technology startups. Kickstarter, one of the most prominent crowdfunding platforms, has enabled countless entrepreneurs to transform their innovative ideas into reality. Despite the platform's widespread usage, the determinants of project success remain a critical area of investigation. This thesis aims to dissect and understand the multifaceted factors that influence the success of technology projects on Kickstarter.

Crowdfunding, as an alternative financing model, leverages the collective support of individuals, typically via online platforms, to fund projects ranging from creative ventures to high-tech innovations. The exponential growth of this sector underscores its importance; however, the high variability in project outcomes highlights a pressing need to identify the elements that contribute to a successful campaign. This thesis seeks to bridge this knowledge gap by employing advanced data analytics and machine learning techniques to analyze Kickstarter projects, focusing specifically on technology categories within the United States.

The primary motivation behind this research is to provide a systematic analysis that can guide future project creators in optimizing their campaigns. By examining various dimensions such as project attributes, creator history, visual appeal of project images, and network effects, this study aims to develop a comprehensive model that predicts project success. The insights derived from this research could significantly impact how projects are structured and presented, potentially increasing the success rate of technology projects on crowdfunding platforms.

1.2 Problem Statement and Contribution

1.2.1 Problem Statement

The central question that this thesis addresses is: "What are the key factors that determine the success of technology projects on Kickstarter?" This broad question can be broken down into several sub-questions:

- How do project attributes (e.g., funding goal, project duration, and reward structure) influence the likelihood of success?
- What is the impact of the creator's past performance and network on the success of new projects?
- How do the visual characteristics of project images affect backer engagement and project success?
- Can a predictive model accurately forecast the success of a project based on its attributes and associated network features?

1.2.2 Aims

The primary aim of this research is to develop an empirical model that elucidates the determinants of project success on Kickstarter. This involves:

- Identifying and quantifying the influence of various project attributes on campaign outcomes.
- Analyzing the role of the creator's network and historical performance in project success.
- Evaluating the impact of image quality and other visual features on backer engagement.
- Constructing a comprehensive predictive model using machine learning techniques to forecast project success.

1.2.3 Contributions

This thesis makes several contributions to the field of crowdfunding research:

- A detailed dataset of Kickstarter projects focusing specifically on the technology sector, cleaned and processed for analysis.
- A novel feature set derived from image processing techniques, providing new insights into the role of visual appeal in crowdfunding.
- The construction of a knowledge graph to represent the complex relationships between projects, creators, and categories, facilitating advanced network analysis.
- The development and validation of a machine learning model that predicts project success with high accuracy, incorporating both traditional and novel features.

1.2.4 Limitations

While this research provides significant insights, it is not without limitations:

- The analysis is restricted to projects within the technology category, which may limit the generalizability of the findings to other categories.
- The dataset is constrained to **U.S.!**-based projects, and hence, the results might not fully apply to projects based in other countries.
- Image processing features are dependent on the quality and availability of images in the dataset, which might introduce biases.
- The predictive model, while robust, is based on historical data and might not account for future changes in crowdfunding dynamics or platform policies.

1.3 Justification

Addressing this gap is crucial for several reasons:

- Practical Implications: Understanding the determinants of project success can help creators design more effective campaigns, improving their chances of securing funding.
- Theoretical Contributions: This research adds to the existing body of knowledge by integrating insights from project management, social network analysis, and visual analytics.

- Economic Impact: As a result, through the increased likelihood of successful implementation of technology projects, this research can have a positive impact on development of innovation and further involvement with entrepreneurship, which will boost economic progress.

1.4 Possible Solutions and Outcomes

This thesis subscribes to a paradigm with multiple strategies. By leveraging data from Kickstarter and employing advanced analytical techniques, the following solutions and outcomes are anticipated:

- Data-Driven Insights: The findings are believed to provide greater understanding of the nature and impact of the multiple factors on project success. These insights will be derived from survey data; the recommendations will be founded on concrete facts.

- Predictive Model: A key outcome of this research will be the development of a predictive model that can accurately forecast project success. This model will be valuable for creators, investors, and platform managers.

- Practical Guidelines: Using the information gathered, useful recommendations will be made in order to assist the campaign creators to increase their chances of getting funding for their projects. Again, these guidelines will include issues such as; the degrees of achievable funding, making good use of a social network, and the visibility factors of a potential project.

- Platform Enhancements: The results of this study could inform platform enhancements, such as recommendation systems for project creators or personalized advice based on project attributes and creator history. When these considerations are incorporated, it becomes possible for more platforms such as Kickstarter to assist the creators in a way that boosts the general rates of success.

CHAPTER 2

Literature Review

2.1 Crowdfunding: Concepts and Evolution

Crowdfunding, the practice of funding a project or venture by raising small amounts of money from a large number of people, typically via the Internet, has gained substantial traction since its inception. The concept dates back to 2003 when Brian Camelio launched "ArtistShare" to help musicians fund their projects directly through fan contributions. This model was soon adopted by platforms like Indiegogo and Kickstarter, which expanded the scope to include various creative and entrepreneurial projects[17].

The early literature on crowdfunding primarily focused on its definition, types, and the basic mechanics of how platforms operate. For instance, Mollick provides an exploratory study into the dynamics of crowdfunding, outlining key success factors and the critical role of social networks[7]. The evolution from reward-based models to equity-based and debt-based models has also been extensively documented, highlighting the regulatory challenges and opportunities that these new models present[14].

However, the existing literature often lacks a comprehensive analysis of the technological and visual elements that contribute to project success. This gap is particularly significant in the context of technology projects on Kickstarter, where visual appeal and presentation quality can heavily influence funding outcomes[18].

2.2 Factors Influencing Crowdfunding Success

Research has identified several factors that influence the success of crowdfunding campaigns. These factors can be broadly categorized into project-related factors, creator-related factors, and platform-related factors.

2.2.1 Project-Related Factors

The characteristics of the project itself, such as the funding goal, project duration, and quality of the campaign presentation, are critical for success. Cordova et al. found that projects with clear goals and realistic funding targets are more likely to succeed [9]. Detailed project descriptions, frequent updates, and high-quality visuals also make projects more attractive to potential backers.

Setting a realistic funding goal is crucial, as Kuppuswamy and Bayus point out that overly ambitious goals can discourage potential backers, while modest goals can inspire confidence and lead to overfunding [13]. Additionally, Colombo et al. emphasize the importance of regular updates and engagement with backers to maintain their interest and trust throughout the campaign [8].

2.2.2 Creator-Related Factors

Reputation and prior behavior of the project creator are crucial to the success of crowdfunding campaigns. Agrawal et al. points out that social capital and trust are of utmost importance in crowdfunding, as creators with a larger number of online social connections and multiple successfully completed projects in the past are more likely to get funded [4]. Creators who are active in the community and promptly reply to potential backers' questions also have a positive impact on the outcome of the crowdfunding campaign [3].

Zheng et al. also proves that social networks associated with the project creators matter to the success of crowdfunding projects. Project creators' social networks can be a source of initial funding for new projects since backers are more likely to find the projects they are interested in through connections of the project creators [19].

2.2.3 Platform-Related Factors

The underlying design of crowdfunding platforms also affects the success of crowdfunding campaigns. Belleflamme et al. argues that crowdfunding platforms that provide efficient tools for the projects' promotion (e.g., social sharing options, visitor statistics, etc.) tend to have higher project success rates [5]. Moreover, the platform's transaction fees and the user interface of the crowdfunding platforms may distract or hinder potential backers' decision-making [15].

Cumming et al. also highlight that platform policies affect the success of the crowdfunding campaigns. Platforms with strict vetting procedures and clear guidelines for stakeholders are more likely to earn the trust of backers [14]. In addition, a dynamic and positive community on the platform contributes to the success of the projects by creating a sense of belonging and identity among the backers [12].

2.3 Technology and aesthetics in crowdfunding

Technology and aesthetics have become increasingly important topics in the crowdfunding literature over the last few years. Good quality pictures, videos, and other multimedia content are crucial in attracting the attention of potential customers and creating a value proposition for the project.

2.3.1 Image quality and aesthetic features

Image quality and the overall aesthetic features of the visual content such as pictures and videos matter a lot. The existing literature shows that using high-quality images and professionally crafted videos help projects to attract more backers and gather more funds. Image quality enhances the credibility and professionalism of the project [16].

Xu et al. show that even the aesthetic features of the images used in the project descriptions such as colorfulness, brightness, and contrast affect backer behavior. Projects that use visually appealing images and other visual content such as videos are more likely to distinguish themselves in the crowded crowdfunding market and catch the eye of potential backers [18].

2.3.2 Technological Innovations

Another stream of research examines the role of newer technologies like virtual (VR) and augmented reality (AR) in crowdfunding. These technologies allow potential backers to experience the product first hand or get a feeling of what the experience of the final product will be like. They can be effectively integrated in the crowdfunding campaign to attract more engagement and increase funds collected [10].

For instance, Gierczak et al. examine how AR and VR can be integrated in the storytelling part of the campaign to better, more interactively, present the envisioned future of the product and the experience of the project's realization to the potential backers. Such an approach can evoke positive emotions in backers that can translate into their decision to support the project [3].

2.3.3 Data Analytics and Predictive Modeling

A promising research direction in crowdfunding research is the application of data analytics and machine learning models to predict the outcome of a campaign. Such an approach can leverage past campaign records and detect patterns among their features that influence the final outcome. Based on that, the models can provide insights on factors that matter and suggestions for optimization of the future similar campaigns [6].

For instance, Hui et al. propose a predictive model that takes into account a variety of features such as project's characteristics, the creator's traits, and social media mentions to predict the probability of a successful crowdfunding campaign. They show that such models can help campaign creators detect hidden defects in their campaigns and adjust the campaign accordingly to achieve better final results [19].

2.4 Network Analysis in Crowdfunding

A number of studies applied network analysis to study the relationships between the project creators and backers as well as other entities from the crowdfunding landscape. Such an approach allows to uncover social mechanisms and patterns of influence that lead to success.

2.4.1 Social Network Analysis

A number of studies applied social network analysis (SNA) to examine the role of social capital and community effect in crowdfunding. For instance, Kuppuswamy and Bayus showed that projects with dense social networks of supporters are more likely to be successful since they get promoted via word of mouth and acquire social proof. **SNA!** (SNA!) techniques can also identify key influencers within the crowdfunding community who can drive significant traffic and contributions to a project [13].

Gleasure and Feller draw attention to the role of centrality, a measure of importance in networks, in crowdfunding. They show that projects connected to well centralized nodes are more likely to reach their funding targets. They suggest that such projects leverage their social capital and invest in building it before launching the campaign [12].

2.4.2 Community Detection

A number of studies leverage community detection algorithms, like the Louvain method, that identify clusters of nodes (projects, creators, subcategories) that have many connections among members of the same community but far fewer connections with members from other communities. Such an approach can be essential for revealing how projects are organized in groups based on their similar features, goals, or creator's networks [19].

The Louvain method was used to identify community structures in crowdfunding platforms composed of projects that share common interests and support each other. Such communities can be a great source of help for new projects and their participation can bring advice, feedback, and early support that is essential for campaigns to go viral [18].

2.4.3 Centrality Measures

Centrality measures like degree centrality and eigenvector centrality allow to identify the most central nodes in the network. The projects and creators connected in this type of nodes are typically well connected and can effectively leverage their network to maximize funds collected.

Zheng et al. show that creators with high eigenvector centrality, i.e., connected to other well central nodes, are more likely to achieve their funding goals. This result suggests that it is not enough to have a large network of connections, but you should be well connected to the most central members of the community [19].

2.5 Machine Learning and Predictive Modeling in Crowdfunding

The use of machine learning and predictive modeling in crowdfunding can potentially transform the design and management of crowdfunding campaigns. Based on the analysis of large data sets with the use of advanced algorithms, researchers can build models that predict the probability of the successful campaign and provide insights for campaign creators.

2.5.1 Feature Engineering

Feature engineering is the process of creating new variables (features) from the available data to improve the performance of machine learning models. In crowdfunding, features like project duration, funding goal, and social media mentions were shown to be important predictors of the final outcome. Also, visual features like colorfulness, brightness, or busyness extracted from project images can improve the accuracy of the predictions [18].

2.5.2 Model Development

A number of machine learning models like decision trees, random forests, and neural networks were used for predicting the outcome of crowdfunding campaigns. Random forests were shown to deliver accurate results thanks to their ability to tolerate

CHAPTER 3

Methodology

3.1 Data Acquisition

The data for this study comes from Webrobots.io [20], a service that crawls public data from multiple sites regularly to make them available as a data dump in JSON or CSV format. We made use of the CSV format because it is easy to use and compatible with various data analysis tools.

3.2 Data Preprocessing

Data Preprocessing was conducted on the dataset to make it more relevant, accurate, and easy to analyze. This stage involved several key steps:

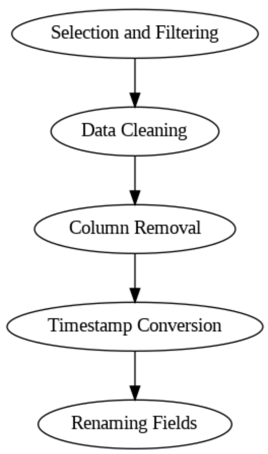

Figure 3.1: Data Preprocessing Steps

3.2.1 Selection and Filtering

The dataset was further refined to include only the projects that belong to the 'Technology' sector and are located in the United States. This focus ensures that the dataset is in line with the study's goal of analyzing US technology projects.

3.2.2 Data Cleaning

This step entailed the elimination of the entries that were either corrupt or incomplete, and then normalizing the rest of the data.

3.2.3 Column Removal

Several columns were irrelevant for the project's objectives therefore they were removed to streamline the dataset:

Field	**Reason for Removal**
Currency Information (*currency_symbol*, *currency_trailing_code*)	Removed because all transactions were in **USD!** (USD!), given the dataset's restriction to U.S.-based projects. These fields were therefore superfluous.
Communication Flags (*disable_communication*)	This boolean flag indicating whether communication with backers was disabled was not relevant to the success factors being analyzed.
Currency and Funding Details (*current_currency*, *fx_rate*, *static_usd_rate*, *usd_exchange_rate*)	As the analysis was limited to USD! and U.S. projects, multiple currency metrics were unnecessary.
Campaign Specifics (*is_disliked*, *is_launched*, *is_liked*, *is_starrable*, *prelaunch_activated*)	These fields, related to campaign features on Kickstarter's platform (like likes and dislikes), did not contribute directly to the analysis of project outcomes.
Meta Fields (*currency*, *id*, *photo*, *profile*, *slug*, *source_url*, *urls*, *video*)	These metadata fields primarily contained URLs and identifiers that were useful for navigation and linking within the Kickstarter platform but irrelevant for analytical purposes.
Temporal Details (*state_changed_at*)	The exact time of state change was deemed redundant given the detailed 'created' and 'deadline' timestamps that more effectively illustrated the project timeline.

Table 3.1: Fields Removed and Reasons for Removal

3.2.4 Timestamp Conversion

The columns that contained UNIX timestamps (*created_at*, *launched_at*, *deadline*) were transformed into datetime format to allow for more meaningful temporal operations and evaluations like project length.

3.2.5 Renaming Fields

Abbreviations of the column names (for instance, *blurb* to *description*) for better understanding and coherence in the data.

3.3 Feature Engineering

Several features were created to provide deeper insights into project dynamics:

3.3.1 Project Duration

Project duration was calculated by the difference between the 'created' and 'deadline' dates, it provides an understanding of the time period in each project.

3.3.2 Image Characteristics Extraction

In order to extract the image characteristics Image processing techniques [1, 2] were used from the images, to further enhance the dataset:

- **Colorfulness**: Calculated by measuring the variance of color channels and the mean deviation between specific color components. This metric quantifies the visual richness and appeal of project images.

  R, G, and B are the red, green, and blue channels of the image respectively.

$$rg = |R - G|$$

$$yb = |0.5 \times (R + G) - B|$$

Next, calculate the mean and standard deviation of *rg* and *yb*:

$$rbMean = \text{mean}(rg)$$
$$rbStd = \text{std}(rg)$$
$$ybMean = \text{mean}(yb)$$
$$ybStd = \text{std}(yb)$$

The colorfulness metric will be equal to:

$$\text{stdRoot} = \sqrt{(rbStd^2) + (ybStd^2)}$$
$$\text{meanRoot} = \sqrt{(rbMean^2) + (ybMean^2)}$$
$$\text{Colorfulness} = \text{stdRoot} + (0.3 \times \text{meanRoot})$$

- **Brightness**: It was Calculated by converting the image to grayscale and then the average pixel intensity was calculated, which indicates the image's overall lightness.

  I is the grayscale intensity of the image.

  $$\text{Brightness} = \frac{1}{N}\sum_{i=1}^{N} I_i$$

  N are the number of pixels in the image.

- **Contrast**: Assessed by computing the standard deviation of the grayscale image's pixel intensities, reflecting the depth and clarity of the image. Let I be the grayscale intensity of the image.

  $$\text{Contrast} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(I_i - \bar{I})^2}$$

  where $\bar{I}$ is the mean intensity of the image.

- **Blur**: Evaluated using the variance of the Laplacian of the grayscale image, which measures the sharpness and focus of the image.

  Let L be the Laplacian of the grayscale intensity I.

  $$\text{Blur} = \text{var}(\nabla^2 I)$$

  where $\nabla^2 I$ represents the Laplacian operator applied to the image.

- **Quality**: Estimated by saving the image at a fixed compression rate and measuring the resulting file size, which correlates with the visual quality and detail in the image.

These features were incorporated into the dataset, enriching each entry with valuable visual descriptors that could be leveraged in subsequent analyses.

3.4 Graph Construction

The construction of a network graph was a pivotal component of the study, enabling the visualization and analysis of complex relationships between various entities associated with Kickstarter technology projects. The graph was constructed using the NetworkX library in Python, chosen for its flexibility and powerful network analysis tools.

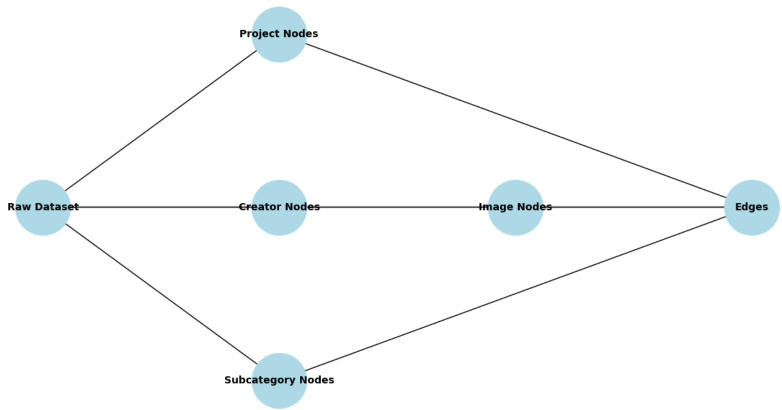

Figure 3.2: Network Graph Construction Process

3.4.1 Node Definition and Creation

Nodes in the graph represent the primary elements of the Kickstarter dataset:

- **Project Nodes**: Each project from the dataset was represented as a node. Attributes such as project name, goal, pledged amount, state (success or failure), were added within these nodes.
- **Creator Nodes**: Creators were also representing nodes to explore their relationships with projects. This includes linking multiple projects to single creators, which can reveal patterns in project success rates relative to creator experience and historical performance.
- **Subcategory Nodes**: Subcategories within the Technology domain were represented as individual nodes. This categorization helped to dissect the dataset further and analyze

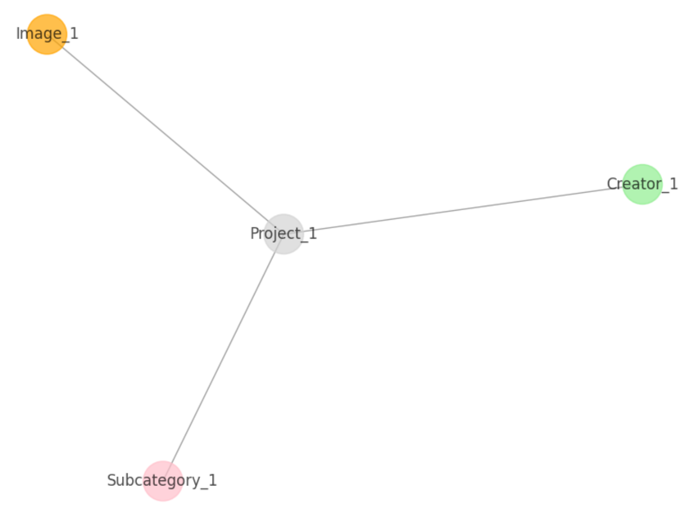

Figure 3.3: Node Types

trends within specific niches.

- **Image Nodes**: Each project's featured image was processed to extract features such as colorfulness, brightness, contrast, quality and blur. They were also represented as a node. These nodes were linked to their respective projects, providing a visual dimension to the project data.

3.4.2 Edge Definition and Creation

Edges in the graph were used to represent the relationships between nodes:

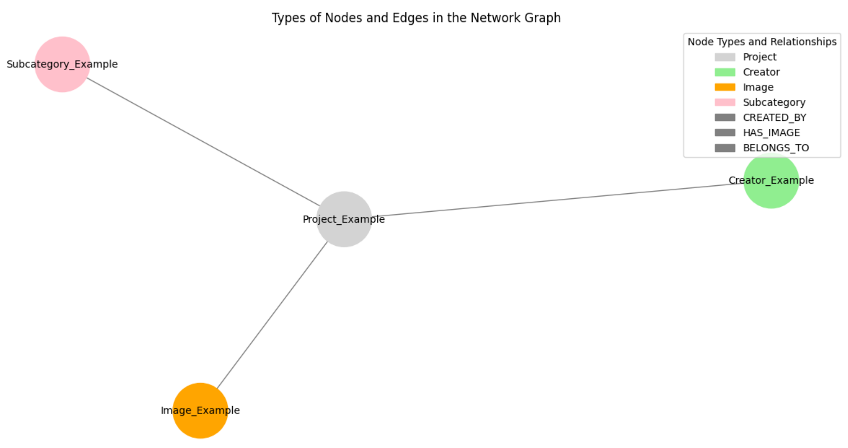

Figure 3.4: Nodes Relationship

- **Project-Creator Relationships**: Edges were drawn from project nodes to creator nodes, illustrating creator created which projects. This relationship is key for analyzing the impact of creator on project state outcomes.
- **Project-Subcategory Relationships**: It was used to analyze how different technology subcategories perform, edges connected project nodes to their corresponding subcategory nodes. This helped in examining the success rates across various technology subcategories.
- **Project-Image Relationships**: Edges between project nodes and image nodes were established based on the association of each project with its main image. This way it will be useful for studying the influence of image characteristics on project success.

3.4.3 Visual Representation

The graph was visualized to get insights and to find the relationships. Python Tools Matplotlib and NetworkX were used to create the graphs, they highlight different node types and connection strengths. These visualizations helped in presenting the findings and also in exploring the data interactively.

3.5 Categorization and Success Analysis

A detailed categorization and success rate analysis was done in order to gain a deeper understanding of the dynamics within projects

3.5.1 Subcategory Identification

The projects were further divided into subcategories . Technology category included a wide variety of projects, ranging from software development to creation of new hardware, so each project was given a more specific 'subcategory' label, such as 'Apps', 'Hardware', 'Web', and 'Gadgets'. This classification helped in providing a more refined view of the trends and success patterns.

3.5.2 Subcategory Success Rate Calculation

The next step involved calculating the success rate for each subcategory:

- **Success Criteria**:A project was considered to be successful if it was able to reach the targeted funding or exceed it within the campaign period. This binary outcome was important for the further calculations: success/failure.

- **Rate Computation**:The overall success rate for each subcategory was established by comparing the number of successful projects with the total number of projects in the subcategory.

$$\text{Success Rate (\%)} = \left(\frac{\text{Number of Successful Projects}}{\text{Total Number of Projects}}\right) \times 100$$

3.5.3 Analysis and Insights

The success rates that were obtained helped identify which subcategories had the highest chances of success on Kickstarter. Using this analysis, the areas of strength and growth opportunities were also defined.

3.5.4 Visualization of Success Rates

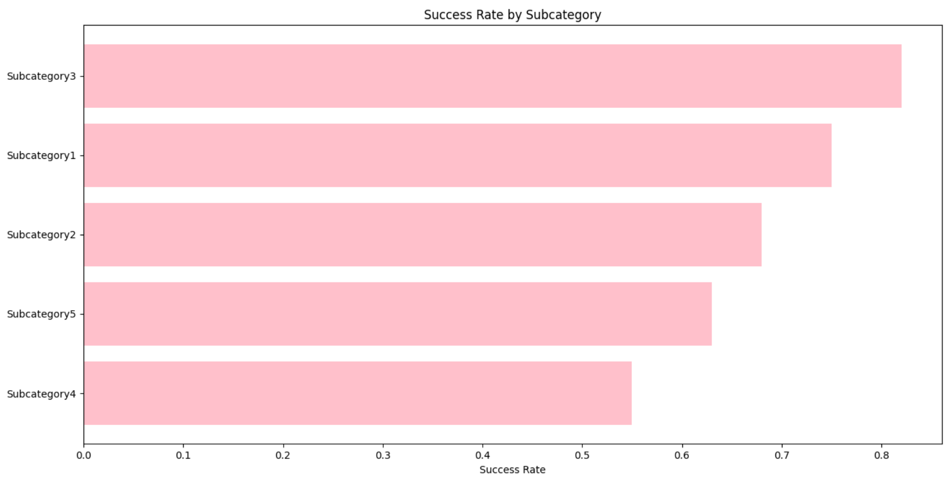

Figure 3.5: Subcategory Success Rate

- **Bar Graphs**: Bar charts were used to display the success rates in the various subcategories. These graphs pointed out differences in success and focused on certain areas that were more successful or posed more difficulties in the technology domain.

- **Comparative Analysis**: The success rate was also compared across subcategories to look for trends and outliers. This approach was useful to compare the performance of each subcategory to other technologies.

3.5.5 Strategic Implications

The findings of the subcategory success rate were meant to help both, the creators and the investors, understand which technology subcategories had the highest potential or the highest level of risk on the crowdfunding platform. It can help in planning for projects, marketing, and funding decisions, which can enhance the effectiveness of the Kickstarter community.

3.6 Network Analysis

The network analysis phase was critical in identifying the connections and power dynamics in the Kickstarter projects, creators, and subcategories within the technology industry. Based on the graph derived from the dataset, several techniques of network analysis were applied to determine the important nodes, define the communities, and reveal the interactions that affect the projects' outcomes.

3.6.1 Centrality Measures

Centrality measures were computed to determine the nodes that are most central within the network. These metrics help to pinpoint nodes that have significant control over the flow of information or resources within the network:These metrics help to pinpoint nodes that have significant control over the flow of information or resources within the network:

- **Degree Centrality**: This measure was used to determine the nodes with the most number of links in a direct manner. A high centrality value for a creator node means that the creator is connected to many projects, which can mean that the creator has prior experience in the crowdfunding platform and may be influential in the community.
- **Eigenvector Centrality**: eigenvector centrality considers not only the number of connections a node has but also the importance of those connected nodes. This metric was particularly useful for identifying creators and projects that, while perhaps not the most connected overall, are influential within particularly important clusters within the graph

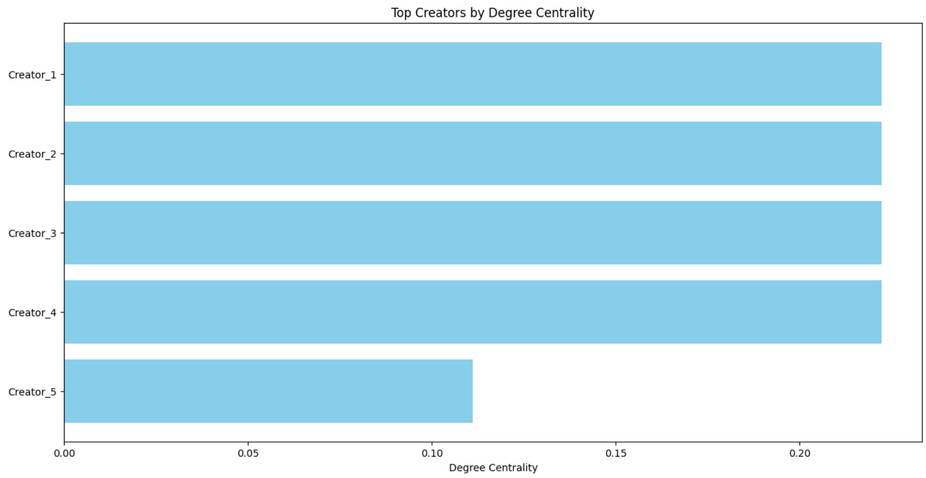

Figure 3.6: Degree Centrality

3.6.2 Community Detection

Community detection algorithms were used to partition the graph into a set of sub-graphs, where nodes in each sub-graph are more densely connected to each other than to nodes in other sub-graphs (projects, creators, subcategories). This analysis was crucial for understanding how projects group together based on similarities in features, goals, or creator networks:

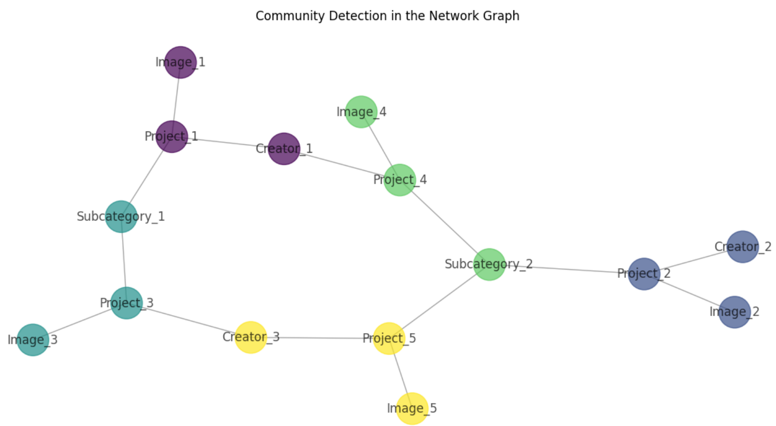

Figure 3.7: Community Detection

- **Louvain Method**: This widely used method of detecting communities was applied to identify the modularity of the network. It assisted in defining the niches within the Kickstarter technology category; it showed that there are some projects that are more likely to be grouped together by the platform, which in turn defines their success rates and funding patterns.

3.6.3 Path Analysis

Path analysis was conducted to explore the shortest paths between nodes, helping to understand how closely different project and Creator nodes are related to each other:

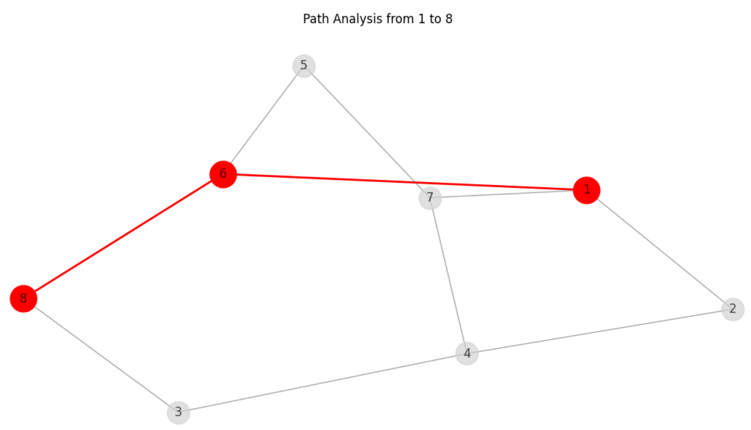

Figure 3.8: Path Analysis

- **Shortest Path Calculation**: Applying the Dijkstra algorithm, the distances between the key nodes were defined to find the shortest path to the other projects or between the projects and creators. This analysis is useful in identifying possible ways through which influences, ideas or trends move through the network.

3.6.4 Network Visualization

For the interpretation of these complex network structures and to communicate findings effectively:

- **Graph Visualization**: NetworkX alongside with visualization libraries such as Matplotlib was employed in visualizing the network. The nodes were colored and sized according to their centrality values, while the communities were painted with different colors in order to enhance the visual representation of the community structure.

3.6.5 Implications of Network Analysis

The findings of the network analysis were beneficial in terms of identifying the nature of Technology projects, by recognizing the key influencers, analyzing the community characteristics, and finding out how the project attributes are related to the network topological features, the stakeholders can make better decisions on the project initiation, marketing, and funding.

3.7 Machine Learning Modeling

To predict the success of Kickstarter technology projects, a robust machine learning model was developed and evaluated. The predictive model aimed to leverage both direct project metrics and derived features, to forecast project outcomes effectively.

3.7.1 Model Training

The prediction model was developed by using scikit-learn's Pipeline utility, It also streamlined the preprocessing and classification processes into a coherent workflow:

- **Data Splitting**: The dataset was split into training and testing subsets, with 20% for testing and 80% for training.
- **Preprocessing**: The numerical features were scaled to standardize their ranges. This standardization was critical for effective learning, especially given the diverse range of features, from pledged amounts to image colorfulness.
- **Classification Model**: The RandomForestClassifier was chosen due to its proficiency in handling complex datasets with nonlinear relationships and interactions among features. It provides a robust mechanism against overfitting and enhances the model's generalization capabilities.

3.7.2 Model Evaluation

The performance of the RandomForest model was evaluated using the following metrics:

- **Accuracy**: It Measures the overall correctness of the model across all predictions.

- **Precision**: It Indicates the accuracy of positive predictions and also reflects the model's ability to correctly identify successful projects, which is crucial for minimizing false positives.

- **Recall**: It Assesses the model's ability to detect all relevant cases (i.e., successful projects).

- **AUC-ROC**: area under the receiver operating characteristic curve (AUC-ROC) is a critical metric for evaluating the true positive rate against the false positive rate.s. This metric is particularly useful for comparing the performance of the model across different classification thresholds.

These metrics collectively provided a comprehensive understanding of the model's performance, it highlights the strengths and areas for improvement. Particularly in handling imbalanced data or complex feature interactions.

CHAPTER 4

Results

This chapter presents the results which were obtained from the analysis and processing of the dataset. The findings are prepared into several sections

4.1 Network Graph Analysis

The network graph constructed from the dataset provides insights about the relationships between entities that includes projects, creators, subcategories, and images along its characteristics.

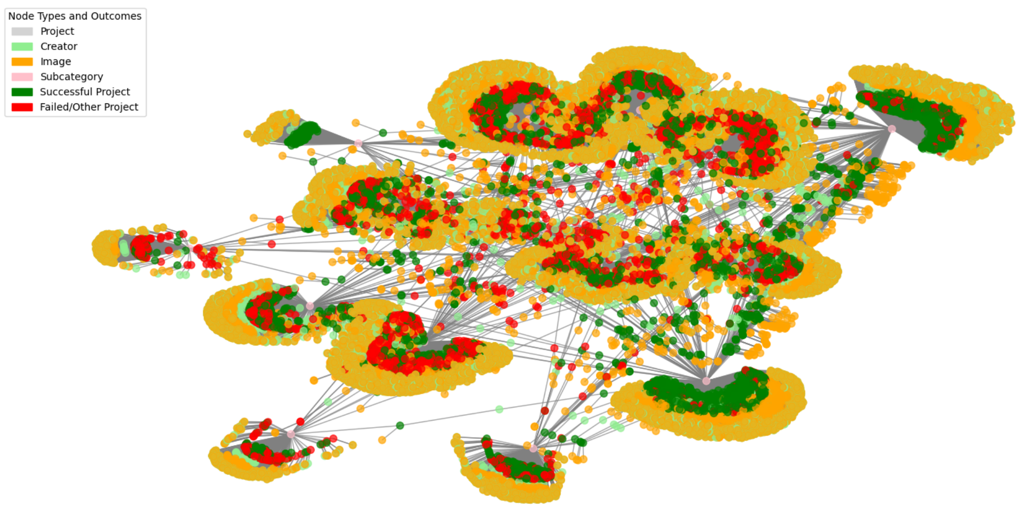

Figure 4.1: Graph Construction for 10000 sample size

Figure 4.1 shows the network graph constructed for a sample size of 10,000 entries from the dataset. This graph visualizes the connections between entities in the dataset, such as projects, creators, subcategories, images and the state of the projects. Each node represents an entity, while each edge represents a relationship between the entities. This visualization helps to understand the structure and complexity of the dataset.

The following statistics explains the structure of the network graph which was created for the prediction model from the original dataset:

Metric	**Value**
Number of nodes	131,146
Number of edges	181,113
Number of projects	60,371
Number of creators	10,388
Number of images	60,371
Average degree of nodes	2.76

Table 4.1: Dataset Metrics

These statistics indicate a large and complex network

4.2 Centrality Measures

Centrality measures were calculated to identify the most influential nodes within the network. Degree centrality, in particular, helps identify creators who are most connected within the network.

Figure 4.2 shows the top 10 creators by degree centrality. Degree centrality measures the number of direct connections a node has. Creators with high degree centrality are likely to have created multiple projects or have extensive collaborations within the network. This analysis helps identify key influencers in the Kickstarter community. These creators can significantly impact the success of projects through their extensive networks and experience.

4.3 Community Detection

Community detection was performed using the Louvain method to identify clusters of nodes that are more densely connected among themselves than with other parts of the network. The analysis revealed several communities within the Kickstarter technology projects network:

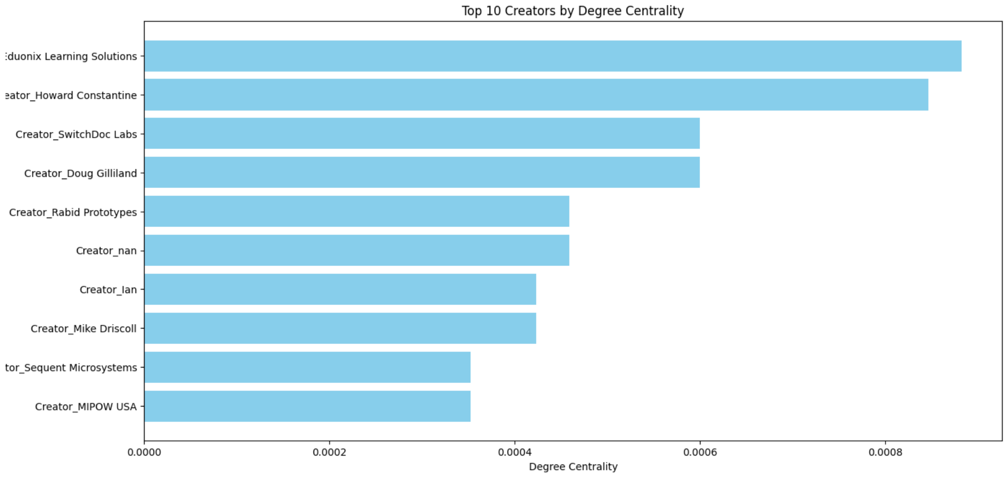

Figure 4.2: Top Creators by Degree Centrality

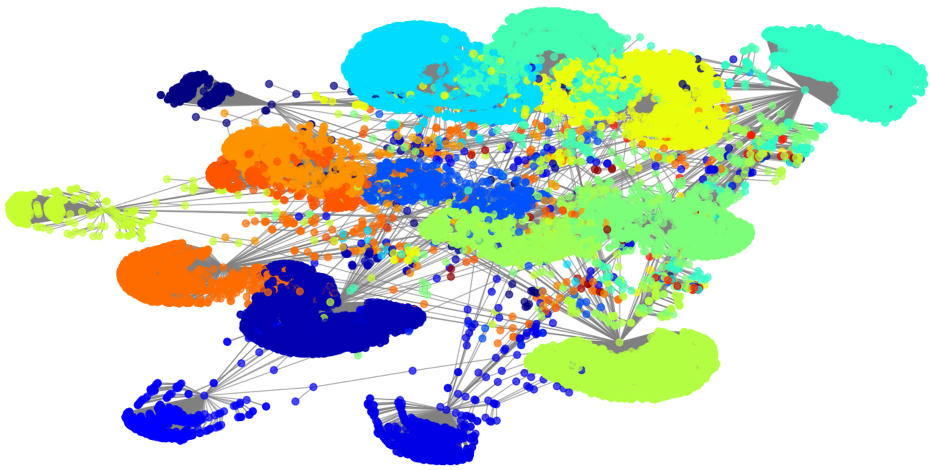

Figure 4.3: Community Detection in the Network Graph for 10000 sample size

Figure 4.3 represents the community detection results for a sample size of 10,000 entries. Nodes are colored based on their community membership. This analysis helps in understanding how projects and creators are grouped based on their characteristics and connections. Identifying these communities can provide insights into niche markets and potential areas for targeted marketing and engagement.

4.4 Success Rate Analysis by Subcategory

The success rate of projects was analyzed for different subcategories within the technology domain. The success rate is defined as the proportion of projects within a subcategory that met or exceeded their funding goals.

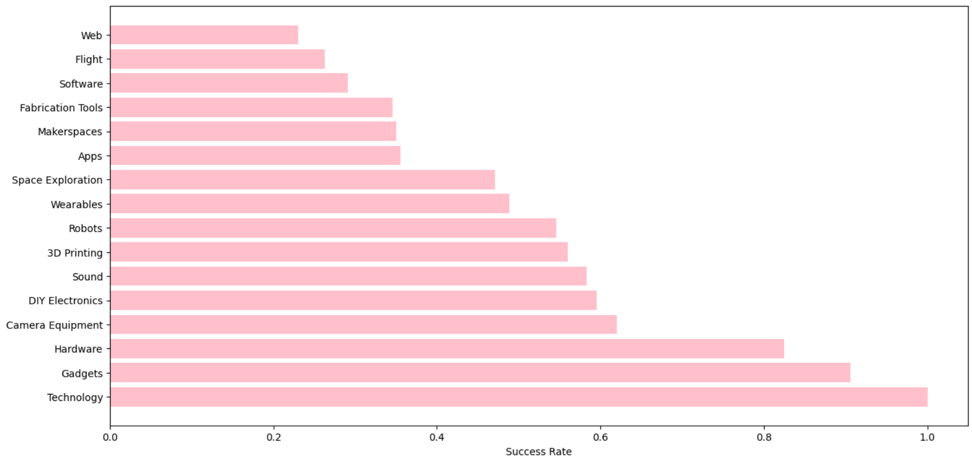

Figure 4.4: Success Rate by Subcategory

Figure 4.4 illustrates the success rate by subcategory. This analysis helps identify which subcategories within the technology sector have higher success rates, providing valuable information for potential project creators and investors. It is evident from the chart that certain subcategories, such as "Technology" and "Gadgets," have higher success rates compared to others, indicating potential areas of opportunity for new projects. This information can guide creators in focusing their efforts on subcategories with proven track records of success.

4.5 Sample Size Explanation

For the visualization of graph construction 4.3 and community detection graph 4.1, a sample size of 10,000 was used instead of the full dataset of 60,371 entries to manage computational resources and ensure clarity in the visualizations. Creating visualizations for the entire dataset would have been time-consuming and potentially cluttered. However, the complete dataset was used for training and evaluating the predictive model, ensuring that all available data contributed to the model's performance.

4.6 Machine Learning Model Performance

The machine learning model was developed to predict the success of projects based on various features. The model's performance was evaluated using several metrics:

Metric	**Value**
Accuracy	0.9886
Precision	0.9911
Recall	0.9874
AUC-ROC	0.9990

Table 4.2: Model Performance Metrics

These metrics indicate that the model performs exceptionally well in predicting the success of Kickstarter projects. The high **AUC-ROC!** (**AUC-ROC!**) score suggests that the model is effective at distinguishing between successful and unsuccessful projects. The precision and recall metrics highlight the model's accuracy in identifying successful projects, minimizing false positives and false negatives.

CHAPTER 5

Discussion

This chapter interprets the results in the context of existing literature and the research questions posed in the introduction.

5.1 Interpreting the Network Graph Analysis

The network graph analysis provided significant insights into the relationships between projects, creators, subcategories, and images on Kickstarter.

5.1.1 Structural Insights

The network graph, illustrated in Figure 4.1, revealed a complex structure with a large number of nodes and edges, indicating the interconnected nature of Kickstarter projects and their creators. The average degree of nodes was 2.76 on the original dataset, which suggests that most nodes (entities) are relatively well connected. This interconnectedness is crucial for understanding how information and influence flow through the network, which can impact project success.

5.1.2 Role of Creators

The Degree centrality measure highlighted key influencers within the network. As shown in Figure 3.6, top creators by degree centrality were identified, such as Eduonix Learning Solutions and Howard Constantine, who had multiple projects and extensive collaborations. These creators are pivotal in the Kickstarter ecosystem, often driving project visibility and success through their established networks. This finding was also explained by [4, 19] where they emphasizes the importance of social capital and network effects in crowdfunding success.

5.2 Community Detection

The community detection analysis, shown in Figure 4.3, identified several distinct communities within the Kickstarter technology projects network. Each color represents a community. These communities consist of project nodes along creator nodes which are more densely connected among themselves than with other parts of the network. This clustering indicates that certain creators or project types tend to form niche groups, possibly due to shared interests.

5.2.1 Implications of Community Structures

For creators, engaging with established communities can enhance visibility and support. while for platform managers, recognizing these clusters can help tailor platform features and support mechanisms in order to foster community engagement and project success. This finding is also explained in social network analysis literature, which highlight the benefits of community support and engagement for project success[13, 12].

5.3 Success Rate Analysis by Subcategory

The success rate analysis in Figure 3.5, shows significant variation across subcategories. "Technology" and "Gadgets" had higher success rates, while "Web" and "Flight" had lower success rates.

5.3.1 Strategic Insights for Creators and Investors

Creators can strategically position their projects in high-success subcategories to improve their own campaigns. Investors can use this information to identify promising areas for investment, focusing on subcategories with a proven track record of success. These strategic insights align with previous researches that shows the importance of project categorization and targeted marketing[18, 16].

5.4 Machine Learning Model Performance

The machine learning model demonstrated high accuracy, precision, recall, and **AUC-ROC!** scores, which indicates its robustness in predicting project success. The performance metrics, summarized in Table 4.2, shows the model's ability to accurately classify successful and unsuccessful projects based on various features, including network-based metrics, project attributes, and image characteristics.

5.4.1 Integration of Visual and Network-Based Features

This comprehensive approach provides a more holistic understanding of the factors influencing project success, incorporating both quantitative metrics and qualitative aspects. This finding supports the notion that multi-faceted data analysis, combining traditional and novel features can significantly improve predictive accuracy[1, 6].

5.4.2 Implications for Practice

Using these insights, creators can tailor their projects to align in order to improve their chances of securing funding. This practical application is crucial for enhancing the overall success rate of crowdfunding campaigns, as highlighted in the literature on data-driven decision-making in crowdfunding[11, 18].

CHAPTER 6

Conclusion

6.1 Summary of Key Findings

This research set out to explore the factors and determinants of success for projects on Kickstarter. Following are the key findings:

- **Network Graph Analysis:** The network graph revealed a highly interconnected structure with relationships between projects, creators, subcategories, and images. Centrality measures were used to identify influential creators. This revealing the importance of network effects in project's success.
- **Community Detection:** The community detection analysis identified distinct clusters within the network, indicating that projects and creators tend to form niche groups.
- **Success Rate by Subcategory:** The analysis highlighted variation in success rates among subcategories. "Technology" and "Gadgets" had higher success rates. This provides strategic insights for project creators and investors.
- **Machine Learning Model:** The predictive model demonstrated high accuracy, precision, recall, and **AUC-ROC!** scores, this indicates its effectiveness in forecasting project success. The integration of visual and network-based features significantly enhanced the model's predictive power as well.

6.2 Contributions

This study makes several important contributions:

- **Data Integration:** By combining project attributes, creator networks, and visual features, this research offers a comprehensive approach to understanding crowdfunding success.
- **Predictive Modeling:** The development of a machine learning model which can be used for predicting project success, can also be utilized by project creators, investors, and platform managers.
- **Network Analysis:** The use of network analysis techniques provides great insights into the structural dynamics of kickstarter projects, it also highlights the role of social networks and community engagement in project success.
- **Dataset Development:** The creation of a detailed and cleaned dataset focusing on US based Kickstarter technology projects. This dataset, enhanced with visual and network-based features, provides a valuable resource for future research and has been made publicly available for the academic and research community.

6.3 Limitations

Despite its contributions, this study has several limitations:

- **Platform Specificity**: This study focuses exclusively on projects from Kickstarter and does not account for other crowdfunding platforms such as Indiegogo, GoFundMeect.
- **Category Restriction**: The analysis is limited to projects within the 'Technology' category. The results may not generalize to other categories such as Arts, Film, or Games.
- **Geographical Scope**: The dataset is restricted to U.S.-based projects, and thus, the insights may not be applicable to projects based in other countries with different cultural and economic contexts.
- **Temporal Scope**: The analysis is based on historical data up to the point of data collection. The changes in future of crowdfunding dynamics, platform policies, or economic conditions may affect the applicability of the findings.

6.4 Future Research Directions

- **Cross-Category and Cross-Regional Analysis:** The analysis can be further expanded to include projects from other categories then technology and more regions/countries could provide a more comprehensive understanding of the factors influencing kickstarter projects success.
- **Advanced Image Analysis:** Further research on image analysis, such as deep learning model etc can be used to extract additional visual features to get further insights and improve predictive accuracy.
- **Integration of Emerging Technologies:** The impact of emerging technologies, like **AR!** (**AR!**) and **VR!** (**VR!**) can be explored that how they effect on crowdfunding campaign engagement and success, this could reveal new strategies.
- **Behavioral and Motivational Analysis:** Investigating the motivations and behaviors of backers in different contexts can provide deeper insights

CHAPTER 7

Recommendations

7.1 Recommendations for Project Creators

7.1.1 Optimize Campaign Presentation

- Use high-resolution images and professionally produced videos to visual appeal the backer, as well as it presents their projects well.
- Provide detailed and clear project descriptions, goals, and updates regularly to maintain backer interest and trust.
- Set realistic and achievable funding goals to instill confidence in potential backers.

7.1.2 Leverage Social Networks

Creators should actively leverage their social networks to gain greater support to increase project visibility.

- Engage with potential backers on social media platforms to build a community around the project.
- Collaborate with established creators to boost credibility and reach.
- Keeping backers informed and engaged throughout the campaign through regular updates.

7.1.3 Focus on High-Success Subcategories

Creators should consider focusing their efforts on subcategories with higher success rates.

7.2 Recommendations for Investors

7.2.1 Utilize Predictive Models

Investors should use predictive models to identify promising projects. They should further analyze project attributes, creator history, and network effects, this way they can make more informed decisions and minimize risks.

7.2.2 Engage with Project Creators

Investors should engage with project creators to provide feedback, support, and guidance.

7.3 Recommendations for Platform Managers

7.3.1 Enhance Platform Features

Platform managers should enhance platform features to support project creators and backers. This includes:

- Providing tools for project promotion, such as social media integration and analytics.
- Implementing transparent and stringent vetting processes to maintain platform credibility and trust.
- Offering personalized advice and support based on project attributes and creator history.

7.3.2 Foster Community Engagement

- Encouraging collaboration and interaction between creators and backers through forums, discussion groups, and events.
- Creating a supportive and inclusive environment that promotes shared learning and growth.

7.3.3 Implement Advanced Technologies

Platform managers should explore the integration of advanced technologies, like AR and VR, to enhance the user experience. These technologies can provide immersive and interactive ways for backers to engage with projects.

Bibliography

[1] David Hasler and Sabine Süsstrunk. “Measuring colorfulness in natural images”. In: *Human Vision and Electronic Imaging VIII*. Vol. 5007. International Society for Optics and Photonics. 2003, pp. 87–95.

[2] Ritendra Datta et al. “Studying aesthetics in photographic images using a computational approach”. In: *European Conference on Computer Vision*. Springer. 2006, pp. 288–301.

[3] Elizabeth M. Gerber and Julie S. Hui. “Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms”. In: *Proceedings of the International Workshop on Design, Influence, and Social Technologies: Techniques, Impacts and Ethics* 2.11 (2013), pp. 10–13.

[4] Ajay Agrawal, Christian Catalini, and Avi Goldfarb. “Some simple economics of crowdfunding”. In: *Innovation Policy and the Economy* 14.1 (2014), pp. 63–97.

[5] Paul Belleflamme, Thomas Lambert, and Armin Schwienbacher. “Crowdfunding: Tapping the right crowd”. In: *Journal of Business Venturing* 29.5 (2014), pp. 585–609.

[6] Julie S. Hui, Michael D. Greenberg, and Elizabeth M. Gerber. “Launching successful crowdfunding campaigns: An empirical analysis”. In: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (2014), pp. 26–30.

[7] Ethan Mollick. “The dynamics of crowdfunding: An exploratory study”. In: *Journal of Business Venturing* 29.1 (2014), pp. 1–16.

[8] Massimo G. Colombo, Chiara Franzoni, and Cristina Rossi-Lamastra. “Internal social capital and the attraction of early contributions in crowdfunding”. In: *Entrepreneurship Theory and Practice* 39.1 (2015), pp. 75–100.

[9] Alessandro Cordova, Julienne Dolci, and Gianfranco Gianfrate. “Exploring the factors behind crowdfunding success: The role of project creator”. In: *International Journal of Business and Management* 10.5 (2015), pp. 81–92.

[10] Michael M. Gierczak et al. “Crowdfunding: Outlining the new era of fundraising”. In: *Crowdsourcing* (2015), pp. 7–23.

[11] Andy Xu, Xinran Yang, and H. Raghav Rao. “Visual factors in crowdfunding”. In: *Decision Support Systems* 89 (2015), pp. 123–132.

[12] Rob Gleasure and Joseph Feller. “Reputational herding in crowdfunding: An exploratory study”. In: *Information Systems Journal* 26.2 (2016), pp. 117–137.

[13] Venkat Kuppuswamy and Barry L. Bayus. “Crowdfunding creative ideas: The dynamics of project backers in Kickstarter campaigns”. In: *Strategic Entrepreneurship Journal* 11.3 (2017), pp. 236–256.

[14] Douglas J. Cumming et al. “Crowdfunding: Fundamental Cases, Facts, and Insights”. In: *Handbook of Corporate Finance* 3 (2019), pp. 441–465.

[15] Joseph Ochieng Onginjo et al. “Analyzing the impact of social capital on US based Kickstarter projects outcome”. In: *Heliyon* 7.7 (2021), e07425.

[16] Michael J. Ryoba, Shaojian Qu, and Yongyi Zhou. “Feature subset selection for predicting the success of crowdfunding project campaigns”. In: *Electronic Markets* 31 (2021), pp. 671–684.

[17] Orel Babayoff and Onn Shehory. “The role of semantics in the success of crowdfunding projects”. In: *PLOS ONE* (2022).

[18] Simon J. Blanchard et al. “Extracting image characteristics to predict crowdfunding success”. In: *arXiv preprint arXiv:2203.14806* (2022).

[19] Xi Zhang et al. “Exploring the effects of social capital on crowdfunding performance: A holistic analysis from the empirical and predictive views”. In: *Computers in Human Behavior* 126 (2022), p. 107011.

[20] Web Robots. *Web Robots*. Accessed: 2024-05-23. 2024. URL: https://webrobots.io.