INSY 662: Individual Project

2 Dec 2024

Submitted by: Hazel Foo

Student ID: 261194705

Introduction

Kickstarter is a crowdfunding platform where creators seek financial support for their innovative projects, offering backers incentives like exclusive rewards or early access. Projects are funded only if funding goals are achieved. This project aims to develop a classification model to predict the success of projects and perform clustering to identify patterns in project characteristics.

Data Pre-processing

First, missing values (1.8%) were removed. Data available at the end or mid-way through the campaign period were excluded. Projects that neither succeeded nor failed (5%), along with identifiers were filtered out as they do not provide meaningful data for prediction. Features like 'staff pick' and 'spotlight' were disregarded as these labels are applicable from pre-launch to post-campaign. Univariate predictors (disable_communication and proj_we_love), and show_feature_image, which showed perfect collinearity with failure, were dropped. To avoid overfitting and reduce dimensionality, the top 10 countries (90% of data) remained intact, while the remaining were recategorised as 'Others'. Finally, all categories were dummified and outliers were removed (5%), and all numerical values were standardised.

Feature Engineering

Granular data on creation, launch, and deadline were excluded due to their specificity. Instead, campaign duration and time between creation and launch were calculated for more generalisable insights. Data integrity was also maintained by ensuring campaign durations were within the 60-day limit. However, the year, month, and weekdays of project launch and deadline were kept to capture seasonality. Project goals were recalculated in USD and scaled over the campaign duration for consistency. Logarithmic transformation was applied to address the skewness in duration till launch and average goal per day before standardisation.

Classification Model & Insights

The best performing model was the Gradient Boosting Classifier, optimised through hyperparameter tuning with k-fold cross-validation and evaluated using validation set approach.

	Precision	Recall	F1-score	Support
Failed	0.76	0.71	0.74	1703
Successful	0.80	0.83	0.82	2339
Accuracy	0.78		0.78	4042

Table 1: Performance Metric of Gradient Boosting Model

Based on Table 1 metrics, it is evident that the model performs marginally better at predicting successful projects, reflecting the imbalance in project states. While effective, its bias requires careful application. Investors with larger budgets can allocate resources to high-success projects, while regular backers can focus on initiatives with higher chances of success, enhancing their experience and the impact of their contributions.

Average daily goal (log)	Campaign duration	Time from creation to launch (log)	Inclusion of video in project launch	Year of launch
0.158287	0.112667	0.104089	0.049162	0.047908

Table 2: Top 5 features and their respective importance in descending order from left to right

Based on Table 2, the average daily goal is the most influential predictor, highlighting the importance of setting realistic funding targets. Unrealistically high goals may deter backers, while attainable ones signal careful planning. The time between project creation and launch reflects readiness, making it a critical factor for campaign optimisation. Additionally, the inclusion of a video is significant, as it captures attention and generates excitement. Creators launching new projects should consider these factors to increase their chances of success.

Clustering Model

To cluster projects, an autoencoder was used for dimensionality reduction followed by KMeans clustering, achieving optimal within-cluster cohesion and between-cluster separation. The original dataset, with over 190 features, saw limited success with KMeans or DBSCAN. After hyperparameter tuning, the autoencoder reduced the feature set to 5 while preserving key characteristics and capturing non-linear relationships. The optimal number of clusters (k=11) was determined using the elbow method and silhouette score (0.245).

Clustering Results & Insights

The clustering analysis revealed key insights into success and failure patterns. Distinctive clusters emerged, with some containing only failed projects, while others showed varying levels of success. Projects with excessively high funding goals, minimal preparation time before launch, less descriptive names, and no image in project description were significantly more likely to fail. These characteristics were often associated with food-related categories which came as no surprise as the food industry is highly saturated with trendy concepts and fast-paced, visually driven content. Younger consumers are increasingly drawn to innovative, unique experiences and products influenced by social media trends and influencers. Conversely, successful projects like fiction, playing cards, and children's books thrive by leveraging creativity, individuality, and emotional connection—qualities often absent in mass production. By offering unique, personal experiences, these projects engage backers on a deeper level, fostering emotional ties and enhancing creative expression. Additionally, strategic campaign timing such as ending on Sundays further enhances chances of success, emphasising the importance of aligning with key factors that drive positive outcomes.

Conclusion

Overall, this clustering analysis provides valuable insights for businesses to optimise their investment strategies. By understanding the traits of successful campaigns, businesses can identify high-potential projects and allocate resources more efficiently, particularly in competitive markets like wearables. This analysis helps businesses target categories with proven success, allowing for more strategic investments and higher return on investments. For creators, the analysis offers valuable guidance to refine their campaigns, meet backer expectations, and improve success rates. By aligning projects with key success drivers, creators can increase their chances of securing investment and market traction, ultimately enhancing their business prospects.