

УДК 330.322:004.9

DOI: <https://doi.org/10.17323/ej.2026.38567>

Мультимодальное прогнозирование успешности краудфандинговых проектов: сигнальная роль текстового и визуального контента¹

Федорова В.С., Завертяева М.А.

Краудфандинг – это относительно новый метод финансирования проектов, который заключается в привлечении средств от большого количества непрофессиональных инвесторов. Несмотря на рост этого направления, только около 41,98% краудфандинговых кампаний являются успешными, что требует выявления факторов, способствующих повышению вероятности успешного сбора средств. Исследование направлено на решение вопроса о том, какие текстовые и визуальные характеристики проектов в наибольшей степени определяют их успех и насколько их учет повышает точность прогнозирования.

В работе разрабатывается мультимодальная модель прогнозирования вероятности сбора средств на платформе Kickstarter. Выдвинуты гипотезы о том, что включение текстовых (*H1*) и визуальных (*H2*) характеристик значительно повышает прогностическую способность моделей, а их совместная интеграция (*H3*) обеспечивает наилучшее качество прогноза за счет учета дополнительной информации.

Эмпирической базой исследования послужили данные о 89557 проектах, размещенных на платформе Kickstarter в период с 2010 г. по 2024 г. Помимо базовых характеристик проектов (целевая сумма, длительность кампании, категория, страна, год публикации), типичных для анализа детерми-

¹ Исследование осуществлено в рамках программы фундаментальных исследований НИУ ВШЭ (HSE-BR-2025-82).

Федорова Валерия Сергеевна – стажер-исследователь Международной лаборатории экономики нематериальных активов. Национальный исследовательский университет «Высшая школа экономики» (Пермь). E-mail: VS.Fedorova@hse.ru

Завертяева Марина Александровна – к.э.н., PhD, профессор департамента экономики и финансов, старший научный сотрудник Международной лаборатории экономики нематериальных активов. Национальный исследовательский университет «Высшая школа экономики» (Пермь). E-mail: moskolkova@hse.ru

Статья поступила: 05.02.2026/Статья принята: 14.05.2026.

нант успешности краудфандинговых кампаний, в анализ включен расширенный набор текстовых и визуальных факторов. Текстовые характеристики извлекались с применением методов обработки естественного языка, включая оценку тональности, сложности и логичности описаний проектов и рисков. Визуальные признаки получены с использованием предобученной нейронной сети ResNet50 с последующим снижением размерности методом главных компонент. Для целей прогнозирования рассмотрены логистическая регрессия, деревья решений, ансамблевые методы (Random Forest, XGBoost, LightGBM) и многослойный перцептрон.

Результаты исследования показывают, что включение текстовых и визуальных факторов существенно повышает качество прогноза. Лучшая модель (LightGBM) достигает значения ROC AUC около 0,82. Установлено, что наибольшее влияние на вероятность успеха оказывают целевая сумма, длительность кампании, характеристики текстового описания и визуального контента. На основе полученных результатов сделан вывод о значимой роли информационных сигналов в краудфандинге и предложены практические рекомендации по улучшению стратегии привлечения финансирования через краудфандинговые платформы.

Ключевые слова: краудфандинг; прогнозирование вероятности успеха проекта; Kickstarter; финансирование проектов; текстовые характеристики; визуальные характеристики.

Для цитирования: Федорова В.С., Завертяева М.А. Мультимодальное прогнозирование успешности краудфандинговых проектов: сигнальная роль текстового и визуального контента. *Экономический журнал ВШЭ*. 2026; 30(2): 276–312.

For citation: Fedorova V.S., Zavertyaeva M.A. Multimodal Prediction of Crowdfunding Project Success: The Signaling Role of Text and Visual Content. *HSE Economic Journal*. 2026; 30(2): 276–312. (In Russ.)

Введение

Краудфандинг является относительно новым способом финансирования проектов, который заключается в сборе средств у большого количества непрофессиональных инвесторов через онлайн-платформу. На данный момент эта сфера растет быстрыми темпами, особенно в Европе и США. Благодаря краудфандингу многие бизнесмены и творческие личности собрали крупные средства на осуществление своих идей.

Первая краудфандинговая платформа ArtistShare, основанная в 2000/2001 г. в США, была предназначена для сбора средств на создание музыкальных композиций [Can You Spare a Quarter, 2010]. На данный момент большое количество сайтов также предоставляет возможность народного финансирования для осуществления творческих идей, однако краудфандинговые площадки сейчас развиваются также в сфере стартапов, технологий, инноваций и других бизнес-идей.

Краудфандинг можно считать успешной альтернативой традиционным методам финансирования. Рынок краудфандинга во многих развитых и развивающихся странах демонстрирует устойчивый рост: согласно Statista Market Insights, стабильная положительная динамика рынка краудфандинга вознаграждений фиксируется с 2017 г. По прогнозным оценкам, к 2028 г. рынок краудфандинга вырастет на 7,63% по сравнению с 2025 г. [Reward-Based Crowdfunding, 2026]. Однако в настоящее время предприниматели сталкиваются с большим количеством трудностей в сборе средств через краудфандинговые платформы. Согласно данным Statista, на 2025 г. только 41,98% проектов с площадки Kickstarter являются успешными [Kickstarter, 2025]. В рамках краудфандинга проект считается успешным, если до официального окончания кампании он собрал целевую сумму или более, в противном случае он классифицируется как неуспешный. Стоит отметить, что в выборке, сформированной для данного исследования, наблюдается сопоставимая доля успешных проектов – 43,37%, а в некоторых категориях процент успеха даже ниже 30. Неуспешные проекты более чем в 80% случаев не собирают и 20% от целевой суммы. Это говорит о неправильном построении стратегии привлечения средств, которое объясняется в том числе и недостаточной изученностью данного феномена. Поэтому прогнозирование вероятности успеха краудфандинга при планировании проекта может показать его неэффективные стороны, что положительно скажется на качестве стратегии и популярности краудфандинга как альтернативного способа финансирования.

Первые научные работы на тему прогнозирования вероятности успеха краудфандинга были написаны только в середине – конце 2010-х годов, т.е. изучение этой сферы началось сравнительно недавно [Fan-Osuala et al., 2018; Song et al., 2019]. Кроме того, использование более современных методов прогнозирования, к которым относятся модели машинного обучения и нейронные сети, можно встретить в работах, написанных после 2020 г. [Kaminski, Hopp, 2020; Tang et al., 2022; Liu, Hu, 2024; Elitzur et al., 2024]. На данный момент в прогнозных моделях в научной литературе используются в основном базовые характеристики проектов, к которым можно отнести целевую сумму, длительность кампании, категорию и другие. Значительно меньше работ концентрируются на исследовании текстовых характеристик, таких как настроение текста, сложность и логичность описания проекта [Kaminski, Hopp, 2020; Tang et al., 2022; Liu, Hu, 2024]. Что касается визуальной составляющей проекта, то она практически не исследуется. Фотографии и видео в основном учитываются как количество прикрепленного визуального материала, а его контент и наполнение не используются в моделях из-за сложностей анализа.

Данная работа призвана заполнить описанный выше научный пробел, поэтому ее целью является разработка прогнозной модели вероятности успешного сбора средств проектами на краудфандинговой платформе Kickstarter на основе мультимодальных данных: базовых характеристик проекта, его текстовых и визуальных характеристик. Исследовательский вопрос состоит в определении того, какие текстовые и визуальные характеристики краудфандинговых проектов в наибольшей степени определяют вероятность их успеха и насколько их включение улучшает прогностическую точность по сравнению с моделями, использующими только базовые характеристики.

Таким образом, в отличие от большинства предыдущих исследований, работа переходит от анализа базовых характеристик проектов (целевая сумма, длительность и т.д.) к исследованию содержательных параметров, таких как описание проекта и рисков и визуальное наполнение. Для достижения цели используются различные методологии, соче-

тающие эконометрический анализ с методами машинного и глубинного обучения. Для анализа текстовых характеристик используется методология анализа естественного языка с применением модели ModernBERT, а для анализа изображений выбраны методы компьютерного зрения, позволяющие получить полное цифровое описание фотографии (ResNet50). При интерпретации полученной прогнозной модели использованы значения Шепли, которые позволяют выявить силу и направление влияния признаков на итоговую вероятность успеха.

Основные выводы исследования указывают на наличие существенного влияния текстовых и визуальных характеристик проекта на его вероятность успеха. Можно сказать, что более подробное и длинное описание проекта и его рисков связано с более высокой вероятностью успеха. При этом описание проекта не должно содержать излишне сложные смысловые конструкции и должно быть понятно широкой аудитории. Что касается визуального наполнения, то минимализм и наличие элементов брендинга могут повысить вероятность успеха проекта, при этом наличие людей на фотографиях, наоборот, отрицательно связано с успехом.

Помимо анализа мультимодальных данных для прогнозирования успешности проекта, работа дополняет существующую литературу анализом описания рисков, так как в предыдущих исследованиях риски учитывались преимущественно с помощью бинарной переменной, характеризующей их наличие или отсутствие. Кроме того, в отличие от предыдущих работ, использующих более простые инструменты для анализа настроения текста, была применена языковая модель, которая позволила улучшить методологию большинства существующих работ на тему краудфандинга. Наконец, существенным дополнением предшествующих исследований стал анализ визуального наполнения, который заключался в формировании цифрового описания изображений, прикрепленных к краудфандинговым проектам. До этого научные работы учитывали визуальные характеристики как количество фотографий и видео без детального анализа представленного контента.

Работа содержит введение, три раздела и заключение. В первом разделе представлена теория сигналов как базовая теория исследования, приведен анализ научной литературы, посвященный факторам успеха краудфандинговых проектов и методам прогнозирования, а также подчеркивается научный пробел существующих исследований. Во втором разделе подробно описаны используемые данные и методология исследования. Третий раздел посвящен результатам проведенного анализа. В нем обсуждаются результаты анализа текстового описания проектов и его рисков, визуального наполнения проекта, а также представлены результаты сравнения разных методов прогнозирования и выбрана лучшая модель.

1. Анализ литературы

Теоретическим фундаментом исследования детерминант успеха краудфандинговых проектов является теория сигналов [Li et al., 2022; Steigenberger, Wilhelm, 2018; Mollick, 2014]. Взаимодействие между инициатором проекта, обладающим полнотой информации относительно качества проекта, и потенциальными инвесторами через краудфандинговые платформы характеризуется сильной информационной асимметрией.

В работе Спенса [Spence, 1978] вложение денег выступает инвестиционным решением в условиях неопределенности. Основная сложность заключается в наличии ненаб-

людаемых характеристик, которые Спенс определяет как внутренние возможности или «качество», недоступные для оценки в момент совершения сделки. В контексте краудфандинга к таким ненаблюдаемым факторам относятся истинный уровень компетенций предпринимателя (команды), реальная стадия готовности проекта и уровень добросовестности инициатора проекта. Поскольку инвестор не может оценить эти параметры напрямую, он вынужден принимать решение, опираясь на доступные данные. Поэтому страница проекта на краудфандинговой платформе становится единственным источником информации для принятия решения о вложении средств.

Для преодоления барьеров, вызванных асимметрией, предприниматели используют информационные сигналы. Спенс разделяет наблюдаемые характеристики на две категории: индексы – неизменные атрибуты, которые предприниматель не может менять в рамках кампании (к ним можно отнести страну регистрации проекта, год, категорию продукта), и сигналы – характеристики, поддающиеся воздействию и требующие инвестиций, которые служат индикаторами скрытых качеств проекта. В контексте краудфандинга под такими сигналами подразумеваются любые доступные инвестору характеристики проекта – от базовых параметров (целевая сумма, длительность кампании) до качества контента (структура текста, наличие видео, качество изображений).

Для того чтобы взаимодействие было эффективным, рынок должен прийти к состоянию разделяющего равновесия, при котором предприниматели разного качества выбирают разные стратегии поведения, позволяя инвесторам однозначно определить тип проекта (качественный или низкокачественный). Это равновесие обеспечивается соблюдением условий самоотбора: проекты высокого качества стремятся подать сильный сигнал, чтобы избежать недофинансирования, в то время как проекты низкого качества осознанно избегают «дорогостоящих» сигналов. Основным механизмом равновесия выступает структура издержек: затраты на подачу сигнала для низкокачественных проектов должны быть выше получаемой выгоды. Подготовка детализированного мультимедального контента (качественные фотографии и видео, проработанное описание проекта) требует значительных ресурсов, которые для компетентного предпринимателя относительно ниже, так как он опирается на реальные наработки. Для «слабого» проекта попытка имитации такого уровня качества оказывается экономически нецелесообразной, так как издержки на фальсификацию сложного сигнала превышают потенциальную выгоду от сборов [Spence, 1978; Connelly et al., 2011].

Однако на практике рынок краудфандинга редко достигает идеального равновесия из-за возможности существования неэффективных состояний: рынок может застрять в «ловушках», где сигналы интерпретируются искаженно или имитируются [Spence, 1978]. Именно отсутствие устойчивого равновесия делает необходимым изучение сигналов качества и их скрытых закономерностей, что позволяет выявлять механизмы, помогающие предпринимателям и инвесторам преодолевать информационные барьеры в неравновесной среде.

Итак, исходя из теории сигналов, далее важно определить, какие характеристики проекта считаются наиболее ценными сигналами для инвесторов, поэтому первым шагом необходимо выделить факторы, являющиеся предикторами успеха (табл. 1). Краудфандинговый проект считается успешным, если он привлек 100% или более от целевой суммы, в противном случае проект считается неуспешным.

Таблица 1.

Факторы успеха краудфандингового проекта,
используемые в научной литературе

Фактор	Направление влияния	Методология	Теоретическое основание
Общие характеристики проекта			
Целевая сумма	Отрицательное влияние	–	Mollick, 2014; Ahlers et al., 2015; Cordova, Dolci, Gianfrate, 2015; Aleksina et al., 2019; Kim et al., 2017; Ferreira, Pereira, 2018; Elitzur et al., 2024; Fan-Osuala et al., 2018; Liu, Hu, 2024; Ralcheva, Roosenboom, 2020
Длительность кампании	Смешанное влияние	–	Bo Xu et al., 2016; Cordova et al., 2015; Mollick, 2014; Boeuf et al., 2014; Kim et al., 2017; Ferreira, Pereira, 2018; Fan-Osuala et al., 2018; Liu, Hu, 2024
Страна	–	–	Di Pietro, Buttice, 2020; Elitzur et al., 2024; Сухановский, Миронова, 2019
Категория проекта	–	–	Mollick, 2014; Elitzur et al., 2024
Год публикации проекта	–	–	Elitzur et al., 2024
Текстовые переменные			
Длина описания	Положительное влияние	Количество слов	Aleksina et al., 2015; Koch, Siering, 2015; Liu, Hu, 2024
Сложность восприятия текста описания	Перевернутый U-паттерн	Индекс Ганнинга	Liu, Hu, 2024
Логичность текста описания	Смешанное влияние	Отношение слов связок ко всем словам в тексте	Liu, Hu, 2024
Настроение текста описания	Смешанное влияние	Мешок слов	Jiang et al., 2020; Liu, Hu, 2024; Zhai, Shen, 2024
Описание рисков	Положительное влияние	Фиктивная переменная: 1 – описание рисков присутствует; 0 – иначе	Ahlers et al., 2015; Koch, Siering, 2015
Визуальные переменные			
Изображения	Положительное влияние	Количество прикрепленных изображений	Liu, Hu, 2024
Видео	Положительное влияние	Наличие видео, длительность видео, анализ речи из видео	Kaminski, Hopp, 2020; Tang et al., 2022

В эмпирической литературе, посвященной моделированию вероятности успеха краудфандинговых проектов, факторы успеха классифицируют на три основные группы: базовые характеристики проекта, текстовые и визуальные переменные. Первая группа факторов включает базовые характеристики проекта: целевую сумму, период кампании, страну, категорию и год публикации. Их значимость подтверждена в ряде исследований [Mollick, 2014; Ahlers et al., 2015]. Однако несмотря на высокую прогностическую способность, данные факторы не полностью объясняют вариативность успеха, поскольку не учитывают содержательные и качественные аспекты проекта. Вторая группа факторов охватывает лингвистические и семантические характеристики описания проекта. В частности, авторы работы [Liu, Hu, 2024] используют Индекс Ганнинга для оценки сложности восприятия текста и долю слов-связок для учета логичности. В данной работе эти показатели также включены в модели как базовая методология учета лингвистических особенностей. Оценка тональности чаще всего проводится методом «мешка слов» [Jiang et al., 2020; Liu, Hu, 2024; Zhai, Shen, 2024]. Несмотря на то, что этот подход считается общепринятым, он не позволяет учитывать контекст и нюансы языка при определении сентимента. Кроме того, научная литература часто учитывает описание рисков в качестве фиктивной переменной, игнорируя содержательную сторону этого фрагмента текста. Хотя объясняющая и предсказательная способность текстовых факторов доказана, применение этих факторов затруднено из-за неструктурированности текстовых данных, субъективности интерпретации и необходимости использования сложных методов анализа. Третья группа факторов связана с визуальным контентом, который чаще всего учитывается количественно или с помощью фиктивных переменных и является наименее изученной [Liu, Hu, 2024; Kaminski, Nopp, 2020]. Такой подход не отражает качественный аспект визуальных материалов, несмотря на их значимое влияние на решения инвесторов и восприятие профессионализма команды.

На данном этапе также важно рассмотреть методы прогнозирования вероятности успеха краудфандинга, которые используются в научной литературе (рис. 1). Их можно разделить на три основные группы: линейные модели, модели машинного обучения и методы глубинного обучения [Liu, Hu, 2024].

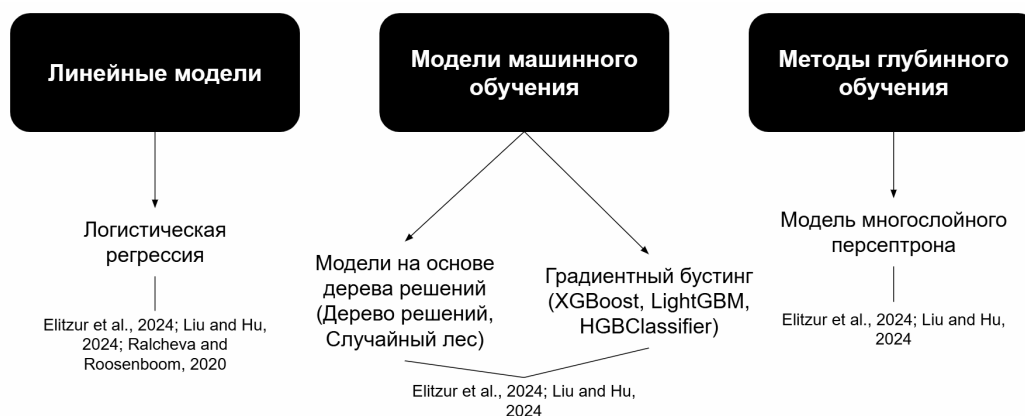


Рис. 1. Методы прогнозирования вероятности успеха краудфандинга

Первый вид моделей – линейные, прежде всего логистическая регрессия. Она проста в реализации, не требует значительных вычислительных ресурсов и позволяет интерпретировать влияние факторов, однако уступает другим подходам по качеству прогноза. Несмотря на это, логистическая регрессия остается наиболее распространенной при оценке вероятности успеха проектов [Song et al., 2019; Ralcheva, Roosenboom, 2020; Kaminski, Hopp, 2020; Liu, Hu, 2024; Elitzur et al., 2024]. Второй вид – это модели машинного обучения, включая деревья решений и ансамбли градиентного бустинга. Они обеспечивают более высокую точность и учитывают нелинейные зависимости, но требуют значительных вычислительных ресурсов и тщательной настройки гиперпараметров. Их применение в задачах прогнозирования краудфандинга активно расширяется в последнее время [Elitzur et al., 2024; Liu, Hu, 2024; Kaminski, Hopp, 2020]. Третий вид – глубокие нейронные сети, преимущественно модель многослойного персептрона (Multilayer Perceptrons – MLP), способные выявлять сложные закономерности на больших выборках. Однако чувствительность к объему данных, склонность к переобучению и сложность настройки ограничивают практическое использование этой модели, несмотря на то, что отдельные работы все же ее рассматривают [Elitzur et al., 2024; Liu, Hu, 2024; Kaminski, Hopp, 2020; Tang et al., 2022].

Все три вида моделей обладают своими преимуществами и недостатками, что не позволяет выделить из них лучшую. Исследователи при анализе приходят к разным результатам при выборе наиболее подходящей модели для прогнозирования вероятности успеха краудфандингового проекта. Например, Р. Элитзур и его коллеги при построении прогноза для вероятности успеха краудфандингового проекта сравнивают модели логистической регрессии (Logistic Regression), градиентного бустинга (Boosted trees), случайного леса (Random Forest), однослойной нейронной сети (Shallow NN) и глубокой нейронной сети (Deep NN). Исследователи приходят к тому, что наилучшей моделью на их данных оказалась модель градиентного бустинга (Boosted trees) с ROC AUC на тестовой выборке около 0,93 [Elitzur et al., 2024]. Почти не уступили в качестве глубокая нейронная сеть (Deep NN) и модель случайного леса (Random Forest) с ROC AUC, равными 0,91 и 0,89 соответственно. К подобным результатам также пришли и другие исследователи [Liu, Hu, 2024]. В работе Камински и Хоппа в качестве метрики сравнения используется Ассигасу [Kaminski, Hopp, 2020]. Исследователи сравнивают по качеству модели логистической регрессии (Logistic Regression), наивного байесовского классификатора (Gaussian NB), метод опорных векторов (SVM), градиентного бустинга (XGBoost) и нейронную сеть (Neural Network). Самыми точными оказались модели логистической регрессии и метод опорных векторов с показателем Ассигасу, равным 0,71.

Видно, что выбор лучшей модели при прогнозировании вероятности успеха краудфандингового проекта зависит от большого количества факторов, таких как выборка данных, подбор параметров модели, выбор метрики эффективности. Поэтому построение прогнозных моделей вероятности успеха кампании актуально и требует более тщательного исследования.

Проведенный обзор литературы показывает, что исследования в области прогнозирования успеха краудфандинговых проектов за последние годы существенно продвинулись, и качество прогнозов в ряде работ достигает высоких значений. Вместе с тем эмпирические исследования в большей степени опираются на относительно простые модели, тогда как потенциал современных методов машинного и глубинного обучения ис-

пользуется ограниченно. Помимо этого, в качестве факторов успешности зачастую используются базовые характеристики проектов, а текстовые и особенно визуальные характеристики краудфандинговых проектов остаются недостаточно изученными из-за сложностей их измерения. Лишь некоторые работы используют такие характеристики текстов как сложность, логичность и эмоциональная окраска описания проекта, а описание рисков чаще всего сводится к формальному признаку наличия соответствующего блока. Визуальная составляющая кампаний, как правило, учитывается лишь количественно, без анализа содержания изображений и видео, что оставляет без внимания важный источник сигналов для потенциальных инвесторов.

Таким образом, в существующей литературе сохраняется исследовательский пробел, связанный с ограниченным использованием мультимодальных данных и недостаточным вниманием к содержательному анализу текстовых и визуальных сигналов. В настоящем исследовании предлагается восполнить данный пробел путем интеграции классических текстовых метрик (Индекс Ганнинга и логичность) с глубоким семантическим анализом тональности на базе нейронных сетей (ModernBERT), что позволяет учесть контекстуальные связи и нюансы языка, которые теряются при использовании словарных методов. Также предлагается перейти от учета наличия описания рисков к содержательному анализу текста блока «Риски и проблемы». Кроме того, работа дополняет предшествующие исследования анализом содержательных аспектов изображений, переходя от количественного учета к извлечению признаков визуального контента методами машинного зрения (ResNet50).

На основе выявленного пробела и поставленного исследовательского вопроса можно сформулировать следующие гипотезы:

H1: Включение текстовых характеристик значимо повышает прогностическую способность моделей по сравнению с моделями, включающими только базовые характеристики.

H2: Включение содержательных визуальных характеристик обеспечивает прирост качества прогноза относительно моделей, ограничивающихся только базовыми характеристиками проектов.

H3: Мультимодальные характеристики, объединяющие текстовые и визуальные характеристики, повышают прогностическую способность моделей по сравнению с моделями, основанными на включении только текстовых или только визуальных характеристик.

2. Данные и методология

Дальнейший анализ будет проведен в несколько основных этапов: предварительная обработка данных, анализ текста описания с помощью методов анализа естественного языка (NLP), анализ фотографий с помощью методов машинного зрения, построение прогнозных моделей.

Первый шаг – предварительная обработка и подготовка данных. Предварительная подготовка текстовых данных будет включать в себя фильтрацию текстов по языку: выбраны только тексты на английском языке для упрощения анализа и обеспечения однородности выборки; удаление стоп-слов и приведение слов к одному регистру. Подготовка текстовых данных позволяет унифицировать тексты для дальнейшего извлечения харак-

теристик. Предварительная подготовка фотографий включает в себя извлечение фотографий по URL и приведение изображений к одному формату и разрешению.

Второй шаг – анализ текста описания с помощью методов анализа естественного языка. На этом этапе проведен анализ тональности (сентимента) текста описания. В отличие от предшествующих работ, основанных на словарных методах [Zhou et al., 2018; Jiang et al., 2020], в данном исследовании использована нейросетевая архитектура ModernBERT (версия – base-multilingual-sentiment, доступная в репозитории Hugging Face: `clapAI/modernBERT-base-multilingual-sentiment`). Выбор данной модели обусловлен ее архитектурным превосходством над классическими трансформерами (BERT, FinBERT, DeBERTa), подтвержденным в исследованиях [Antoun et al., 2025]. Во-первых, ModernBERT поддерживает окно контекста до 8192 токенов (около 6000 слов на английском языке). Это позволяет проводить полный анализ длинных описаний проектов без потери информации, что невозможно при использовании стандартных моделей с лимитом в 512 токенов. Во-вторых, модель демонстрирует высокую эффективность классификации, согласно оценке, на тестовом датасете `clapAI/MultiLingualSentiment`, ModernBERT достигает метрики F1-score около 80,16, что сопоставимо с результатами более тяжелых моделей при значительно большей скорости обработки [clapAI, 2024]. В работе использована предобученная модель с длиной токенизации до 4096 токенов. Процедура дообучения не проводилась, так как архитектура ModernBERT обладает высокой обобщающей способностью. Использование предобученных весов позволяет минимизировать риск переобучения на специфической лексике выборки, были использованы стандартные веса [clapAI, 2024].

Также на основе текста определены логичность и сложность восприятия текста описания и описания рисков проекта. Сложность восприятия текстов рассчитана с помощью индекса Ганнинга (Fog Index), который определяет долю сложных слов в тексте и присваивает уровень сложности прочтения этого текста. Индекс измеряет читаемость текста и отражает уровень усилий, необходимый для его понимания. Интерпретация значений индекса Ганнинга приведена в Приложении 1 (табл. П1.) Логичность текста определяется отношением количества слов-связок ко всем словам в тексте [Liu, Hu, 2024]. В Приложении 2 представлен список причинных и соединительных слов, которые были использованы для анализа.

Третий шаг – анализ фотографии проекта с помощью методов машинного зрения. В настоящее время в научной литературе фотографии, прикрепленные к краудфандинговому проекту, чаще всего учитываются в моделях как переменная, отражающая количество прикрепленных изображений [Zhou et al., 2018; Liu, Hu, 2024], однако данная работа предполагает, что контент изображения может иметь значительное влияние на вероятность успеха краудфандинга. Поэтому в прогнозную модель были добавлены дополнительные числовые параметры, которые являются цифровым описанием фотографии.

Для извлечения визуальных сигналов использована глубокая сверточная нейронная сеть ResNet50, реализованная на базе библиотеки torchvision и предобученная на наборе данных ImageNet-1K со стандартным набором весов (ResNet50_Weights.DEFAULT) [He et al., 2016]. Процесс извлечения признаков включал процедуру нормализации входных данных, необходимую для корректной работы предобученной модели. Все изображения были приведены к разрешению 224 на 224 пикселя, после чего значения интенсивности пикселей были масштабированы к диапазону [0; 1] и нормализованы с исполь-

зованием среднего значения ($\text{mean} = [0.485, 0.456, 0.406]$) и стандартного отклонения ($\text{std} = [0.229, 0.224, 0.225]$) набора данных ImageNet. В качестве цифрового описания фотографии был использован выход слоя глобального усредняющего пулинга, представляющий собой вектор из 2048 признаков.

Поскольку этот вектор обладает высокой размерностью и избыточностью, для выделения наиболее значимых характеристик был применен метод главных компонент (PCA). Выбор данного метода обусловлен необходимостью получения стабильных характеристик для последующего обучения прогностических моделей. В отличие от нелинейных методов снижения размерности, таких как t-SNE или UMAP, которые ориентированы в первую очередь на визуализацию данных, PCA обеспечивает детерминированное преобразование, которое сохраняет структуру дисперсии и позволяет обрабатывать новые наблюдения без необходимости полного пересчета модели. Применение PCA позволяет сохранить основную информацию, необходимую для обучения моделей, и снизить влияние «шумовых» признаков. Основным ограничением данного подхода является невозможность интерпретации извлеченных признаков, что делает прогностические модели менее прозрачными. Однако данная методология релевантна при ответе на исследовательский вопрос, так как позволяет сравнивать фотографии с разным стилем и контентом и учитывать этот контент в прогностических моделях, что не позволяют сделать другие методы анализа изображений.

Четвертый шаг – построение прогностических моделей вероятности успеха краудфандинга. Проект считался успешным и оценивался как 1, если он собрал целевую сумму до окончания срока, и неуспешным в противном случае (оценивался как 0) [Mollick, 2014; Liu, Hu, 2024]. В работе использованы три основных типа моделей – линейные модели, модели машинного обучения и глубокие нейронные сети. Сравнение большого количества моделей разных типов позволяет определить модель, которая наилучшим образом будет описывать данные, что позволяет добиться максимальной точности прогноза.

Перед построением моделей проведено масштабирование данных с помощью логарифмирования количественных переменных и стандартизации всей выборки с применением метода StandardScaler в Python, что позволяет избежать влияния масштаба на результаты моделей. Данные были случайно поделены на обучающую и тестовую выборку в пропорции 80/20 для оценки качества моделей. Случайное разбиение выборки позволяет минимизировать риск исторического переобучения. При хронологическом разбиении существует высокая вероятность того, что модель «выучит» специфические для ранних этапов развития платформы закономерности, которые потеряли актуальность. Использование случайного деления позволяет модели сфокусироваться на устойчивых во времени информационных сигналах, что подтверждается практикой аналогичных исследований [Koch, Siering, 2015; Elitzur et al., 2024], где временные эффекты контролируются через включение переменной года в модель, а не через структуру разбиения.

Кроме того, важным шагом является использование кросс-валидации при выборе оптимальной модели каждого типа, так как кросс-валидация позволит избежать переобучения и добиться максимальной точности.

После построения моделей, оценивающих вероятность успешного финансирования посредством краудфандинга, рассчитываются общепризнанные метрики их качества Accuracy (точность классификации) и ROC AUC (Area Under the ROC Curve, площадь под кривой ошибок). Accuracy отражает точность предсказания, т.е. долю правильно классифицированных наблюдений от общего количества наблюдений:

$$(1) \quad Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)},$$

где TP – True Positive, количество верно предсказанных успешных проектов; TN – True Negative, количество верно предсказанных неуспешных проектов; FP – False Positive, количество неверно предсказанных успешных проектов; FN – False Negative, количество неверно предсказанных неуспешных проектов.

Метрика ROC AUC отражает производительность модели и измеряет способность модели различать два класса (успешные и неуспешные проекты). Значения ROC AUC варьируются в пределах от 0 до 1, причем значения ниже 0,5 свидетельствуют о том, что модель хуже случайного классификатора, а выше 0,5 – лучше. Поскольку выборка исследования не сбалансирована, в частности по количеству успешных и неуспешных проектов, соотношению проектов из разных стран и категорий, наиболее предпочтительной метрикой качества моделей является ROC AUC, дающая точные результаты на несбалансированных данных.

На основе дополнительного анализа лучшей модели будут выдвинуты рекомендации по повышению вероятности успеха краудфандингового проекта. В качестве методов дополнительного анализа будут использованы график пчелиного роя, значение Шепли и статистические тесты.

Для проведения анализа используются данные о краудфандинговых проектах с платформы Kickstarter с 2010 по 2024 гг. Сбор данных осуществлялся с помощью веб-скрейпинга непосредственно с официального сайта платформы. В данных представлена информация об ID, создателе и названии проекта, описании проекта и рисков, статусе проекта, стране, в которой реализуется проект, о целевой и собранной сумме в долларах США, о дате и времени запуска и окончании проекта, ссылка на фотографии и видео, прикрепленные к проекту.

Исходный массив содержал 103703 наблюдения. Были исключены полные дублирующие записи, возникшие в процессе обработки текстов и агрегации датасета. С помощью библиотеки `langdetect` были отобраны проекты, описания которых написаны на английском языке. Это необходимо для корректного извлечения текстовых переменных. Также были удалены наблюдения с пропущенными значениями в используемых переменных. Для большинства количественных переменных вместо удаления экстремальных значений было применено логарифмирование, что позволило сохранить объем выборки и масштабировать распределение. Удаление выбросов было проведено точечно только для переменной индекса Ганнинга. Наблюдения со значением выше 50 были исключены, так как это свидетельствует об ошибках измерения (наличие HTML-кода в тексте, ошибки кодировки, отсутствие пунктуации). Также некоторые проекты были исключены из анализа из-за невозможности корректной обработки визуального контента (битые ссылки, ошибки при загрузке изображений, несоответствия форматов).

В результате многоступенчатой очистки итоговый набор данных составил 89557 наблюдений. Данные релевантны для построения прогнозных моделей, так как представляют собой большую выборку за длительный промежуток времени с одного из самых популярных краудфандинговых сайтов, что снижает влияние случайных факторов на модель.

В табл. 2 представлено описание факторов, используемых для построения прогнозной модели вероятности успеха краудфандингового проекта. Эти переменные были выделены на основе анализа научной литературы. Также факторы были дополнены более глубоким анализом текста описания и изображений, что позволяет дополнить существующую литературу.

Таблица 2.

Независимые переменные	
Фактор	Описание
Целевая сумма	Сумма, которую предприниматель хочет собрать на осуществление своего проекта (долл. США)
Период кампании	Длительность краудфандинговой кампании (дни)
Страна	Фиктивная переменная: 1, если проект реализуется в США, 0 – проект реализуется в другой стране
Категория проекта	Категория проекта в соответствии с категориями на сайте Kickstarter (пр.: Технологии, Искусство и т.д.)
Год публикации проекта	Год, в который проект был анонсирован
Длина описания	Количество слов в описании проекта
Сложность восприятия описания проекта	Индекс Ганнинга [Liu, Hu, 2024] в соответствии с табл. П1 Приложения 1
Логичность описания проекта	Отношение количества слов-связок ко всем словам в тексте. Чем выше, тем текст логичнее [Liu, Hu, 2024] (Приложение 2)
Настроение описания проекта	Категориальная переменная, принимающая значения: POSITIVE – позитивный сентимент; NEGATIVE – негативный сентимент; NEUTRAL – нейтральный сентимент. Оценка осуществляется с помощью нейросетевой модели ModernBERT (Bidirectional Encoder Representations from Transformers)
Длина описания рисков	Количество слов в описании рисков
Сложность восприятия описания рисков	Индекс Ганнинга [Liu, Hu, 2024] текста описания рисков в соответствии с табл. П1 Приложения 1
Логичность описания рисков	Отношение количества слов-связок ко всем словам в тексте описания рисков. Чем выше, тем текст логичнее [Liu, Hu, 2024] (Приложение 2)
Цифровое описание изображения	Цифровое описание изображения, полученное с помощью предобученной нейронной сети ResNet50
Наличие видео	Фиктивная переменная: 1, если к проекту прикреплено видео, 0 – иначе

В табл. 3 представлены описательные статистики независимых количественных переменных в исходном виде. Видно, что все количественные переменные имеют сильный разброс относительно среднего значения, особенно целевая сумма. Поэтому они были логарифмированы перед дальнейшим анализом. Кроме того, был проведен анализ корреляций между независимыми переменными: значения не превышают 0,15 для большинства переменных, что минимизирует риск возникновения проблемы мультиколлинеарности (табл. ПЗ Приложения 1). Корреляции около 0,3 наблюдаются для индекса Ганнинга описания проекта и рисков и логичности описания проекта и рисков, что закономерно.

Таблица 3.

Описательные статистики количественных переменных

	Целевая сумма, долл. США	Период кампании, дни	Длина описания проекта, слова	Длина описания рисков, слова
Наблюдения	89557,00	89557,00	89557,00	89557,00
Среднее	57109,37	33,73	508,77	100,11
Стандартное отклонение	1203466,00	12,76	487,85	94,63
Минимум	1,00	1,00	1,00	0,00
Квантиль 25%	1500,00	29,00	193,00	44,00
Квантиль 50%	5000,00	30,00	361,00	76,00
Квантиль 75%	15000,00	37,00	657,00	125,00
Максимум	100000000,00	90,00	7132,00	5073,00

Для выявления статистических различий между успешными и неуспешными проектами был проведен t-тест Стьюдента на равенство средних количественных переменных, результаты представлены в табл. 4. По всем анализируемым характеристикам наблюдается статистически значимое различие между успешными и неуспешными проектами на уровне значимости 1%. Успешные проекты, как правило, обладают более низкой целевой суммой и длительностью кампании, а также более длинным описанием проекта и его рисков.

Таблица 4.

Результаты теста Стьюдента для количественных переменных

Признак	t-statistic	Среднее значение для успешных проектов	Среднее значение для неуспешных проектов
Целевая сумма	-11,132***	12298,195	91486,586
Период кампании	-60,620***	30,905	35,885
Длина описания	56,755***	614,641	427,682
Длина описания рисков	18,981***	106,843	94,959

Примечание: ***p-value < 0,01.

Также стоит рассмотреть качественные переменные. Зависимой переменной при построении моделей выступает бинарный статус проекта. Неуспешных проектов в данных немного больше (56,7%), однако, в целом, данные сбалансированы относительно этого показателя. Кроме того, соотношение успешных и неуспешных проектов в выборке соответствует статистике по платформе Kickstarter, что говорит о релевантности данных.

В разрезе географического распределения проектов наблюдается дисбаланс: доля проектов из США составляет 68,59% от общего объема выборки. В связи с этим для обеспечения устойчивости оценок и предотвращения разреженности данных переменная страны была преобразована в бинарный формат (США, Другие страны). Кроме того, неуспешных проектов в США 56,6%, а в других странах 56,8%. Статистическая значимость данных различий подтверждаются тестом на равенство пропорций на уровне значимости 1%.

Что касается наличия или отсутствия видео, то большинство проектов имели видео. Причем только 30% проектов без видео стали успешными, проекты с видео же становились успешными в 51% случаев. Эту тенденцию также подтверждает z-тест на равенство пропорций ($p\text{-value} < 0,01$).

Анализ годовой динамики (рис. 2) выявил неоднородность: период 2014–2015 гг. характеризуется резким ростом числа проектов со снижением доли успешных проектов, что, вероятно, связано с эффектом масштабирования платформы. Данная структурная особенность выборки делает случайное разбиение на тестовый и обучающий сетки наиболее обоснованным методом. Хронологическое деление в таких условиях может привести к проблеме временного дрейфа данных и сдвига концепта [Gama et al., 2014; Lu et al., 2018]. Как отмечают исследователи, при наличии нестабильности в целевой переменной модель, обученная на одном временном периоде, может потерять предсказательную способность на другом из-за изменения рыночного контекста. Случайное разбиение выборки позволяет нивелировать эту проблему, обеспечивая присутствие проектов из разных временных периодов в обеих выборках. Это гарантирует, что анализируемые предикторы являются инвариантными к внешним шокам и структурным сдвигам рынка краудфандинга, что соответствует практикам анализа данных в этой области [Koch, Siering, 2015; Elitzur et al., 2024].

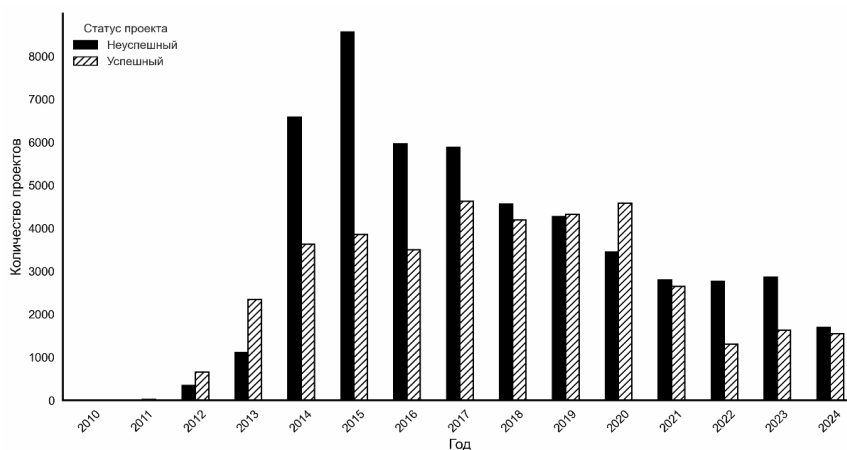


Рис. 2. Соотношение успешных и неуспешных проектов по годам

На рис. 3 представлено соотношение успешных и неуспешных проектов в разрезе категорий.

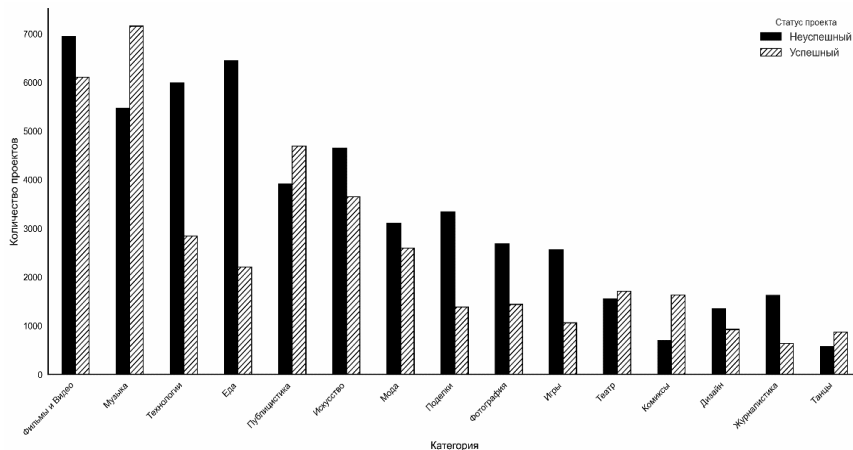


Рис. 3. Соотношение успешных и неуспешных проектов в разрезе категорий

Видно, что в некоторых категориях проекты более успешны, к таким категориям можно отнести «Комиксы» (70% успешных проектов), «Танцы» (60%), «Музыка» (57%). А наименьшая доля успешных проектов встречается в категориях «Еда» (26% успешных проектов), «Журналистика» (28%), «Игры» (29%) и «Подарки» (29%).

3. Результаты

После предварительной обработки данных был проведен анализ текстов описаний проектов и их рисков на основе методов обработки естественного языка (Natural Language Processing, NLP). Были рассчитаны такие показатели, как сложность восприятия описания проекта и его рисков, логичность текстов описания и рисков и настроение (сентимент) текста описания. Сентимент описания рисков не рассматривался, так как он всегда будет определяться моделью как негативный из-за слов, связанных с риском.

Распределение индекса Ганнинга, позволяющего определить сложность восприятия текстов, для описания проекта (а) и для описания рисков (б) относительно статуса проекта представлено на рис. 4. Медианные значения индекса для обоих типов текстов находятся в диапазоне от 11 до 12, что свидетельствует об уровне читаемости, доступном широкой аудитории (табл. П1 Приложения 1), в том числе потенциальным инвесторам. Также был проведен t-тест, который показал, что успешные проекты отличаются более простым для восприятия языком описания проекта и рисков на уровне значимости 1%. Кроме того, индекс Ганнинга имеет достаточно большой разброс как для описания проектов, так и для описания рисков. Для решения этой проблемы значения были логарифмированы с применением квартильного метода, так как в данных присутствовали нулевые значения [Smith, 2018].

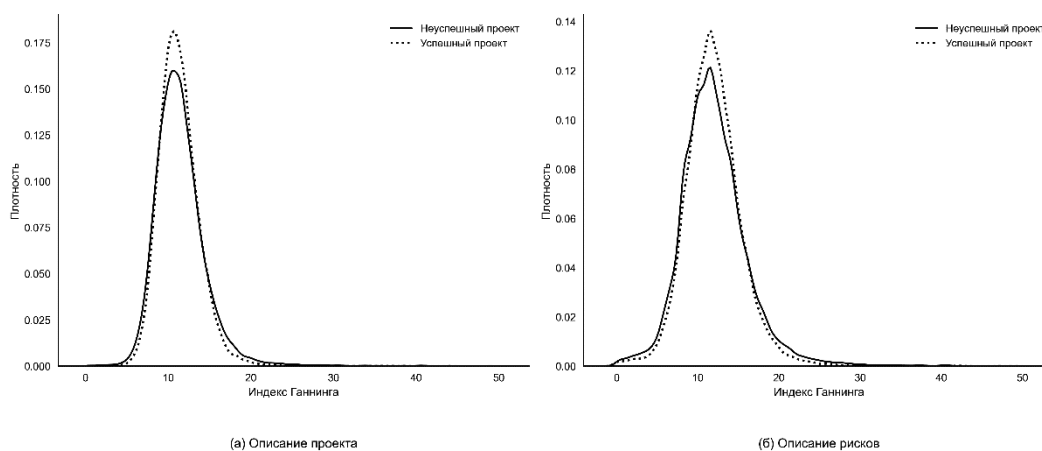


Рис. 4. Распределение Индекса Ганнинга:
(а) описание проекта, (б) описание рисков

Распределение показателя логичности текста для описания проектов (а) и для описания рисков (б) представлено на рис. 5. Медианные значения показателя равны 0,068 для описания проекта и 0,071 для описания рисков, соответственно, можно сказать, что около 7% слов в текстах являются словами-связками. Эти значения указывают на средний уровень связности текстов. Результаты теста Стьюдента показали, что есть статистические различия между успешными и неуспешными проектами по показателю логичности текста описания и описания рисков на уровне значимости 1%: логичность описания проекта ниже, а логичность описания рисков выше для успешных проектов. Важно отметить, что высокая доля слов-связок не всегда коррелирует с информативностью текста, поэтому иногда логичность текста может быть высокой, однако сам текст будет бесполезен для потенциальных инвесторов.

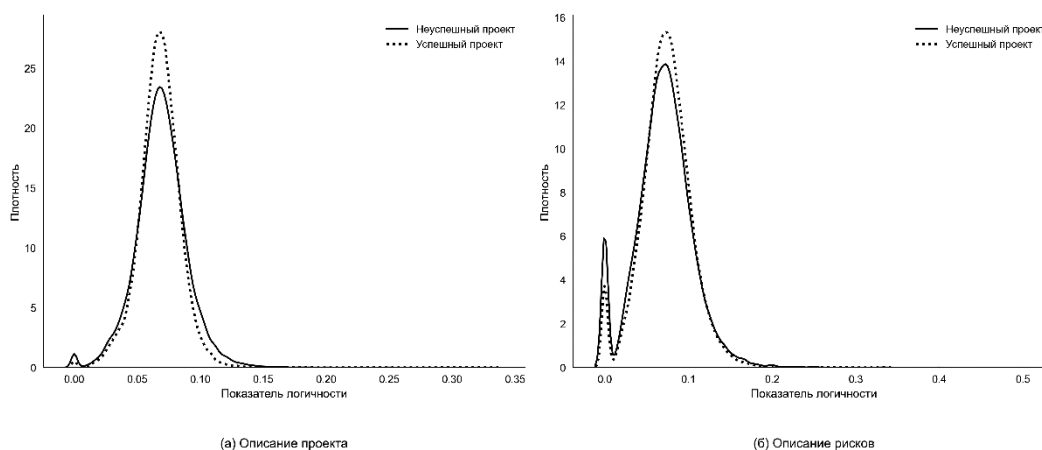


Рис. 5. Распределение показателя логичности текста:
(а) описание проекта, (б) описание рисков

Результаты применения модели ModernBERT для оценки настроения текста показали, что 92,6% текстов имеют положительный sentiment, 6,7% – нейтральный и 0,8% – негативный. Видно, что данные несбалансированы по этому признаку, что может вызвать смещение при обучении моделей и снизить точность прогноза, особенно для нейтрального и негативного sentiment. При этом среди проектов с негативным sentiment около 80% оказались неуспешными, среди проектов с нейтральным настроением – 73% проектов, а среди проектов с положительным sentiment – 55%.

На следующем шаге анализа были получены визуальные характеристики проекта, полученные с помощью глубокой сверточной нейронной сети ResNet50 (Residual Network, 50 слоев). Данная модель широко используется в задачах классификации и извлечения признаков из изображений, так как она предварительно обучена на большом наборе данных и способна эффективно распознавать сложные визуальные паттерны. Каждое изображение было преобразовано в вектор фиксированной длины, содержащий информацию о 2048 признаках. Вектор признаков был сформирован с выходного слоя перед финальной классификацией, т.е. получен «цифровой отпечаток» изображения без определения класса, к которому принадлежит объект. Это позволяет учитывать в прогнозных моделях такие особенности изображений, как форма объектов, текстура, цветовые сочетания, контрасты и композиционные элементы.

Однако использование высокой размерности признаков при дальнейшем анализе может привести к переобучению моделей и значительным вычислительным затратам. Также большое количество признаков при ограниченном объеме данных снижает способность модели к обобщению и качество прогноза. Для борьбы с этими проблемами было проведено снижение размерности с помощью метода главных компонент (PCA).

На рис. 6 представлены (а) график каменистой осыпи и (б) график накопленной объясненной дисперсии для первых 500 компонент. Анализ графика каменистой осыпи демонстрирует точку изгиба («локоть») в области первых 50–100 компонент, однако использование данного подхода приводит к низкому уровню накопленной объясненной дисперсии (около 55%), что признается недостаточным для сохранения репрезентативности многомерных визуальных характеристик. Согласно установленным правилам [Hair et al., 2019], число выбранных факторов должно быть достаточным для достижения порога объясненной дисперсии, который составляет 60% и выше. В данном исследовании выбор 200 компонент позволяет достичь кумулятивного показателя в 67,97%, что соответствует критерию и обеспечивает сохранение значимой информации при существенном сокращении размерности. Таким образом, выбор 200 компонент является компромиссом между минимизацией избыточного шума и обеспечением информативности данных для дальнейшего прогнозирования.

При этом важно отметить, что компоненты не являются точно интерпретируемыми, т.е. невозможно определить, за какую именно характеристику отвечает каждая компонента (цвет, форма, конкретный объект и так далее). В дальнейшем анализе эти переменные используются при построении прогнозных моделей для кодирования информации об изображении.

После предварительного анализа текстовых и визуальных характеристик были оценены прогнозные модели вероятности успеха краудфандинга. На этапе оценки моделей первым шагом выборка была поделена на обучающую и тестовую: 80% наблюдений были отнесены к обучающей выборке, а 20% – к тестовой. Деление производилось слу-

чайным образом, что снижает вероятность смещения и повышает устойчивость моделей к переобучению. Далее количественные переменные были стандартизированы, т.е. приведены к распределению с нулевым средним и стандартным отклонением, равным единице. Это позволяет исключить влияние масштаба данных на результаты модели.

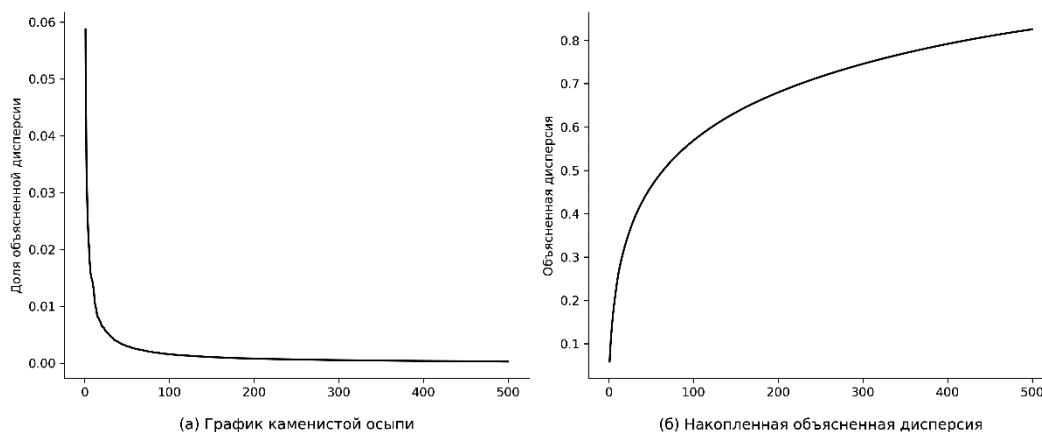


Рис. 6. Графики каменной осыпи и объясненной дисперсии

Подбор гиперпараметров моделей машинного и глубинного обучения осуществлялся методом случайного поиска (Random Search) с использованием перекрестной валидации по пяти фолдам. Такой подход позволяет эффективно подобрать оптимальные гиперпараметры и избежать переобучения моделей, когда модель демонстрирует высокую точность на обучающих данных, но неспособна эффективно предсказывать тестовые наблюдения. Также использование кросс-валидации повышает прогностическую способность моделей и надежность результатов. Предварительные шаги позволили подобрать оптимальные параметры для всех моделей.

Для прогнозирования вероятности успеха краудфандингового проекта были использованы линейные модели, методы машинного обучения (Decision Tree, Random Forest, XGBoost, LightGBM, HGBClassifier) и глубокие нейронные сети (MLP). Результаты ROC AUC представлены в табл. 5. Узкие границы 95-процентных доверительных интервалов для ROC AUC свидетельствуют о статистической устойчивости полученных оценок. В табл. П4 Приложения 1 также представлены показатели Accuracy, Precision, Recall и F1-score.

На первом этапе (столбец 1 табл. 5) модели оценивались только на базовых характеристиках проекта (целевая сумма, период кампании, категория, страна и год публикации). Все модели показали результаты лучше случайного классификатора ($\text{ROC AUC} > 0,5$), при этом наилучшие показатели ROC AUC ($\sim 0,77$) и точности ($\sim 70\%$) продемонстрировали модели градиентного бустинга: XGBoost, LightGBM и HGBClassifier.

Таблица 5.

Оценка качества прогноза (ROC AUC)

	Базовые	Базовые и текстовые	Базовые и визуальные	Все
Logistic Regression	0,6983 (0,6911; 0,7050)	0,7722 (0,7658; 0,7788)	0,7383 (0,7312; 0,7450)	0,7913 (0,7844; 0,7975)
Decision Tree	0,7355 (0,7282; 0,7424)	0,7657 (0,7588; 0,7721)	0,7094 (0,7019; 0,7162)	0,7553 (0,7481; 0,7618)
Random Forest	0,7592 (0,7520; 0,7658)	0,8067 (0,8003; 0,8127)	0,7635 (0,7565; 0,7699)	0,8032 (0,7964; 0,8094)
XGBoost	0,7656 (0,7588; 0,7725)	0,8147 (0,8084; 0,8211)	0,7892 (0,7825; 0,7956)	0,8248 (0,8188; 0,8310)
LightGBM	0,7656 (0,7583; 0,7721)	0,8148 (0,8086; 0,8212)	0,7851 (0,7781; 0,7918)	0,8249 (0,8188; 0,8311)
HGBClassifier	0,7646 (0,7576; 0,7711)	0,8132 (0,8070; 0,8194)	0,7869 (0,7798; 0,7935)	0,8222 (0,8157; 0,8281)
MLP	0,7559 (0,7488; 0,7629)	0,7763 (0,7697; 0,7829)	0,6819 (0,6740; 0,6897)	0,7165 (0,7090; 0,7233)

Примечание: в ячейках указан ROC AUC и его 95-процентный доверительный интервал (в скобках).

На втором этапе к базовым характеристикам были добавлены текстовые признаки из описаний проектов и рисков. Это привело к росту качества прогнозирования для всех архитектур. ROC AUC лучших моделей увеличился до 0,81, а точность – примерно до 74%, что подтверждает значимость текстовых характеристик. Согласно результатам теста Делонга, представленным в табл. 6, наблюдаемый при добавлении текстовых характеристик прирост является статистически значимым для всех моделей на уровне 1%. Причем этот прирост неоднороден: наибольший вклад текстовые характеристики имеют для модели логистической регрессии (+0,07 ROC AUC), а наименьший – для нейронной сети MLP (+0,02 ROC AUC).

Таблица 6.

Результаты теста Делонга на разницу показателей ROC AUC

	Текстовые и базовые/ Базовые	Визуальные и базовые/ Базовые	Все/ Текстовые и базовые
Logistic Regression	0,0738*** (14,173)	0,0399*** (7,448)	0,0192*** (4,005)
Decision Tree	0,0302*** (5,860)	-0,0261*** (-4,751)	-0,0104** (-2,049)
Random Forest	0,0474*** (9,933)	0,0042 (0,844)	-0,0035 (-0,769)

Окончание табл. 6.

	Текстовые и базовые/ Базовые	Визуальные и базовые/ Базовые	Все/ Текстовые и базовые
XGBoost	0,0491*** (10,471)	0,0236*** (4,872)	0,0101** (2,323)
LightGBM	0,0492*** (10,476)	0,0195*** (4,020)	0,0101** (2,329)
HGBClassifier	0,0486*** (10,328)	0,0223*** (4,582)	0,009** (2,052)
MLP	0,0204*** (4,120)	-0,0740*** (-13,849)	-0,0598*** (-11,706)

Примечание: Δ ROC AUC – разность показателей площади под ROC-кривой сравниваемых моделей; в скобках указана z-статистика теста Делонга; * p-value < 0,1, ** p-value < 0,05, *** p-value < 0,01.

Далее были оценены модели с включением визуальных характеристик, таких как наличие видео и числовые признаки изображений, полученные с использованием ResNet50 и метода главных компонент. Добавление визуальных данных повысило точность моделей градиентного бустинга (LightGBM, XGBoost, HGBClassifier) на 1–2 п.п. и ROC AUC на 0,02 по сравнению с базовыми моделями (статистическая значимость также подтверждается тестом Делонга на уровне 1%). Однако реакция алгоритмов на визуальные данные оказалась разной: в то время как ROC AUC моделей логистической регрессии и градиентного бустинга вырос, ROC AUC для Decision Tree и MLP значимо снизился (–0,03 и –0,07 соответственно). Это может указывать на чувствительность данных архитектур к «шуму», возникшему при добавлении дополнительного 201 признака.

Аналогичные результаты были получены при добавлении всех признаков одновременно. При сравнении с моделями, оцененными на базовых и текстовых признаках, видно, что для моделей градиентного бустинга ROC AUC вырос и достиг рекордных 0,82. Этот рост также статистически значим на уровне 5%. Это свидетельствует о том, что визуальные факторы добавляют ценность дополнительно к текстовой информации и использование мультимодального подхода при прогнозировании вероятности успеха краудфандинга оправдано. Кроме того, тест Делонга подтверждает превосходство моделей градиентного бустинга над другими моделями на уровне значимости 1%, однако статистических различий в результатах моделей градиентного бустинга не обнаружено.

Результаты данной работы развивают некоторые исследования. В работе [Kaminski, Норр, 2020] лучшей моделью для прогнозирования вероятности успеха краудфандинга стала модель линейной регрессии с Ассигасу, равным 0,71. Данная работа достигла показателя Ассигасу 0,74 для лучшей модели. Также результаты данной работы частично совпадают с предшествующими научными исследованиями. Лучшей оказалась модель градиентного бустинга, что подтверждает результаты [Elitzur et al. 2024; Liu, Hu, 2024]. Однако в данной работе качество прогноза, измеренное показателем ROC AUC, меньше. Авторы работы [Elitzur et al., 2024] достигли показателя, равного 0,93, для лучшей модели градиентного бустинга, а [Liu, Hu, 2024] – 0,89, в то время как данная работа достигла показателя

ROC AUC, равного 0,82. На качество прогноза в данном случае, во-первых, могла повлиять невозможность учета личных характеристик предпринимателя, таких как предыдущие успешные и неуспешные проекты, социальный капитал, измеряемый в количестве комментариев в период кампании, которые были учтены в исследовании [Elitzur et al., 2024]. Во-вторых, в работе [Liu, Hu, 2024] использовались данные о 3101 проекте с другой краудфандинговой площадки, что могло существенно повлиять на разницу в результатах моделей. Результаты работы [Elitzur et al., 2024] получены на выборке 108223 проектов, собранных за период с 2013 по 2016 гг., что отличается от выборки данных в этой работе. Также значительную роль играют подбор гиперпараметров и качество обработки данных, так как модели чувствительны к этим характеристикам.

Для интерпретации результатов выбрана модель LightGBM, оцененная на всех признаках (параметры: num_leaves = 70, max_depth = 14, learning_rate = 0.1, random_state = 42). Выбор обусловлен ее вычислительной эффективностью при работе с массивами данных среднего и большого объема (около 90000 наблюдений в данной работе). Благодаря использованию градиентного одностороннего сэмпирования (GOSS, Gradient-based One Side Sampling), LightGBM обеспечивает высокую скорость обучения с минимальными затратами памяти при идентичных показателях точности по сравнению с XGBoost и HGBClassifier [Ke et al., 2017]. Поскольку все модели градиентного бустинга показали схожие результаты и практически идентичны в ранжировании признаков, LightGBM была выбрана как наиболее оптимальная с точки зрения производительности.

Был построен график «пчелиного роя» (Bee Swarm Plot), основанный на значениях Шепли [Liu, Hu, 2024]. Данный график иллюстрирует вклад каждого признака в предсказанную вероятность успеха краудфандинговой кампании (рис. 7). На графике отражены первые 20 признаков, имеющие наиболее значимый вклад в прогноз и представляющие 81,33% объясняющей способности модели. Полученные значения Шепли устойчивы на уровне значимости 5% (результаты проведения проверки на устойчивость методом бутстрапирования представлены в табл. П5 Приложения 1).

Признаки на графике отсортированы в порядке важности. Видно, что наибольшее влияние на вероятность успеха краудфандинговой кампании оказывают целевая сумма, длина описания проекта, год публикации и длительность проекта. Шкала показывает значение признака: темные точки – высокие значения признака, светлые точки – низкие значения признака. Горизонтальная ось отражает значение Шепли: выше 0 – признак повышает вероятность успеха, ниже 0 – признак снижает вероятность успеха.

Для формулирования конкретных количественных рекомендаций были рассчитаны оптимальные пороговые значения факторов на основе значений Шепли. Методология заключалась в анализе векторов признаков успешных проектов, для которых вклад конкретного параметра в итоговый прогноз был положительным (SHAP > 0). В качестве оптимального диапазона использованы значения между 25-м и 75-м перцентилями распределения, что позволяет выделить интервалы факторов, которые инвесторы воспринимают как сигналы качества. Результаты подтверждают некоторые предшествующие исследования [Mollick, 2014; Liu, Hu, 2024]: высокие значения целевой суммы и длительности кампании снижают вероятность успеха краудфандингового проекта. Наиболее благоприятно рынок реагирует на запросы суммы в диапазоне 600–3000 долл. США при оптимальной длительности кампании 25–30 дней.

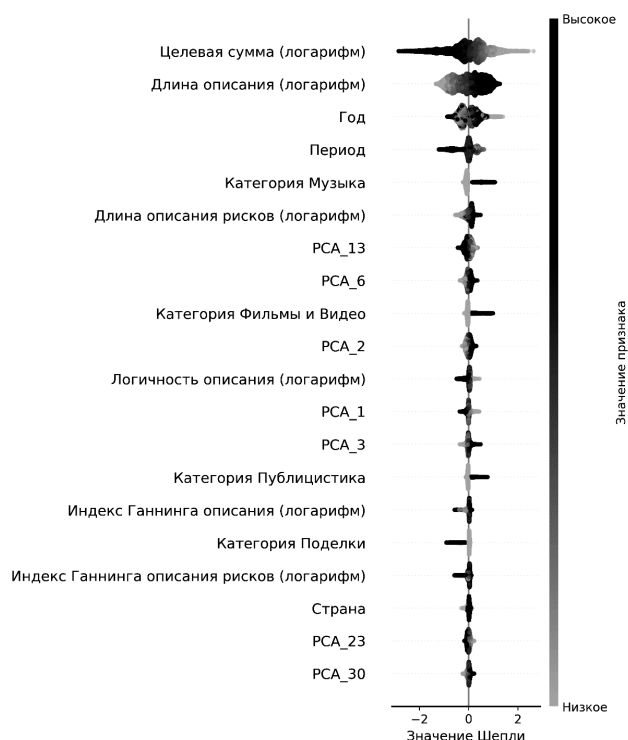


Рис. 7. График «пчелиного роя» для модели LightGBM

Что касается текстовых характеристик, длина описания проекта и рисков положительно связаны с вероятностью успеха: оптимальный объем описания проекта составляет 424–910 слов, а описания рисков – 99–175 слов, что согласуется с предыдущими исследованиями [Ahlers et al., 2015; Koch, Siering, 2015; Liu, Hu, 2024]. Однако слишком сложный текст или большое количество слов-связок могут снижать вероятность успеха. Оптимальный индекс Ганнинга для описания проекта находится в диапазоне от 11 до 13, для описания рисков допускается более сложная лексика – 18–21. Долю слов-связок в описании проекта лучше сохранять в диапазоне от 10 до 11%.

Важную роль также играют некоторые визуальные признаки. Так как они были извлечены с помощью модели ResNet50 и подвергнуты процедуре снижения размерности, их прямая интерпретация невозможна. Тем не менее в работе был проведен качественный анализ изображений с экстремальными значениями компонент. В табл. 7 представлено описание визуальных паттернов, соответствующих компонентам с наибольшей прогностической силой.

Наиболее сильное влияние на вероятность успеха краудфандинга имеют компоненты PCA_6, PCA_2, PCA_1 и PCA_3. Компонента PCA_13 была опущена из-за редкости использования данного паттерна в изображениях на краудфандинговых платформах. Обе компоненты PCA_6 и PCA_3 характеризуются минималистичным изображением, однотонным фоном и наличием большого количества свободного пространства. PCA_6 отражает использование логотипов на темном фоне, что может придавать проекту более завер-

шенный вид и давать сигнал о высокой проработанности потенциальным инвесторам. PCA_3 сконцентрирован на считывании текста на однотонном фоне. Текст может добавлять прозрачности и снижать когнитивную нагрузку на инвесторов, которые хотят быстро понять суть проекта и его качество [Pieters, Wedel, 2004].

Таблица 7.

Пример интерпретации изображений с экстремальными значениями компонент

Компонента	Влияние на успех	Визуальное содержание (интерпретация)	Пример
PCA_6	Положительное	Использование логотипов, графический минимализм, темный фон, большое количество свободного пространства	Проект 1 ² , Проект 2 ³ , Проект 3 ⁴
PCA_2	Положительное	Яркая цветовая гамма, большое количество деталей	Проект 4 ⁵ , Проект 5 ⁶ , Проект 6 ⁷
PCA_1	Отрицательное	Наличие людей на фотографии	Проект 7 ⁸ , Проект 8 ⁹ , Проект 9 ¹⁰
PCA_3	Положительное	Текст, графический минимализм, однотонный фон, большое количество свободного пространства	Проект 10 ¹¹ , Проект 11 ¹² , Проект 12 ¹³

² См.: https://www.kickstarter.com/projects/2016612727/7-corners-debut-album?ref=discovery&term=7%20cjrntsr&total_hits=1&category_id=43

³ См.: https://www.kickstarter.com/projects/bloodypineapple/bloody-pineapples-new-album?ref=discovery&term= Bloody%20Pineapples%20New%20Album&total_hits=1&category_id=241

⁴ См.: https://www.kickstarter.com/projects/1186040528/cub-magazine-wwi-centenary-issue?ref=discovery&term=CUB%20Magazine%20-%20WWI%20Centenary%20Issue&total_hits=1&category_id=359

⁵ См.: https://www.kickstarter.com/projects/carnetdevoyage/make-100-essential-oil-diffusers-100-diffuseurs?ref=discovery&term=Make%20100%20Essential%20Oil%20Diffusers%20-%20100%20Diffuseurs&total_hits=1&category_id=26

⁶ См.: https://www.kickstarter.com/projects/jabartmx/jabart-una-opcion-natural-para-tu-limpieza-persona?ref=nav_search&result=project&term=Jabart%3A%20una%20opcion%20natural%20para%20tu%20limpieza&total_hits=1

⁷ См.: https://www.kickstarter.com/projects/1030007474/building-a-home-for-brigadeiro-bakery?ref=nav_search&result=project&term=Building%20a%20Home%20for%20Brigadeiro%20Bakery&total_hits=1

⁸ См.: https://www.kickstarter.com/projects/1573132265/edge-of-eternitys-full-length-cd-dreams-of-complac?ref=nav_search&result=project&term=Edge%20of%20Eternity%27s%20Full%20Length%20CD%20%22Dreams%20of%20&total_hits=1

⁹ См.: https://www.kickstarter.com/projects/thaddwilliamsmusic/take-a-knee?ref=discovery&term=Take%20a%20Knee&total_hits=5&category_id=39

¹⁰ См.: https://www.kickstarter.com/projects/1764720044/deadbeat-dads-0?ref=discovery&term=DEADBEAT%20DADS&total_hits=12&category_id=330;%20https://www.kickstarter.com/projects/362268404/colorful-murals-tell-the-story-of-one-houston-neig?ref=nav_search&result=project&term=Colorful%20Murals%20Tell%20the%20Story%20of%20One%20Houston&total_hits=1

Компонента PCA_2, наоборот, имеет яркую цветовую гамму и большое количество деталей, однако также положительно коррелирует с вероятностью успеха краудфандинга. Это объясняется наличием корреляции между цветовыми решениями в маркетинге и восприятием продукта, так правильно подобранная цветовая гамма может стать решающим моментом в восприятии контента потенциальными инвесторами [Gorn et al., 1997; Labrecque, Milne, 2012].

PCA_1 отвечает за наличие людей на фотографиях и отрицательно связана с вероятностью успеха краудфандинга. Это противоречит теории о том, что наличие лица или «очеловеченность» бренда повышают доверие к продукту, а также противоречит некоторым исследованиям [Wang et al., 2024] и требует дальнейшего более глубокого анализа.

Заключение

В заключение можно сказать, что краудфандинг в настоящее время является одной из основных альтернатив традиционному финансированию, и его популярность продолжает расти, поэтому изучение этой сферы актуально. В статье проводится прогнозирование успешности проектов, размещенных на популярной краудфандинговой платформе Kickstarter, на основе мультимодальных данных: базовых характеристик, к которым относятся целевая сумма, категория, период, страна и год публикации проекта; текстовых характеристик проектов, а также характеристик изображений, приложенных к проектам.

Прогнозирование успешности проектов осуществлялось с использованием семи моделей, принадлежащих трем разным типам: линейные модели, модели машинного обучения и методы глубинного обучения. Результаты сравнительного анализа показали, что лучшее качество прогноза демонстрирует модель LightGBM, обученная на всех трех типах характеристик, которая представляет собой усовершенствованный ансамбль деревьев решений на основе градиентного бустинга. На тестовой выборке модель достигла показателя ROC AUC, равного 0,82, что говорит о ее высокой производительности и способности выявлять сложные закономерности в больших данных, особенно на бинарных факторах.

Теоретическая значимость исследования заключается в расширении существующих подходов к анализу факторов успеха краудфандинговых проектов за счет интеграции мультимодальных данных и современных методов машинного обучения. В отличие от большинства предшествующих работ, фокусирующихся преимущественно на базовых характеристиках кампаний, в статье теоретически и эмпирически обоснована роль содержательных текстовых и визуальных сигналов как значимых факторов успешности. Полученные результаты развивают положения теории сигналов в контексте цифровых финансовых платформ, демонстрируя, что информационное наполнение проектов играет важную

¹¹ См.: https://www.kickstarter.com/projects/howtoquityourdayjob/baby-its-cold-outside-silver-spring-stage-hvac-fun?ref=nav_search&result=project&term=Baby%20It%27s%20Cold%20Outside%20Silver%20Spring%20Stage%20&total_hits=1

¹² См.: https://www.kickstarter.com/projects/2082673896/bloghouse-the-book?ref=nav_search&result=project&term=Bloghouse%3A%20The%20Book&total_hits=1

¹³ См.: https://www.kickstarter.com/projects/heavygemdev/our-live-are-data-game-development?ref=nav_search&result=project&term=%20Our%20live%20are%20data_%22%20Game%20Development&total_hits=1

сигнальную функцию и выступает каналом снижения асимметрии информации между авторами проектов и потенциальными инвесторами.

Что касается практической значимости работы, то результаты данного исследования могут быть использованы предпринимателями на практике. Так, может быть разработано приложение, позволяющее прогнозировать вероятность успешного сбора средств через краудфандинговые платформы и давать персональные рекомендации. Кроме того, результаты работы могут быть полезны краудфандинговым платформам, которые на основе полученных данных смогут скорректировать работу сайтов. Результаты позволяют выявить закономерности между характеристиками проектов и вероятностью их успеха, что может помочь в улучшении рекомендательных систем краудфандинговой платформы.

По итогам исследования можно сформировать список рекомендаций, которого стоит придерживаться при публикации проекта на краудфандинговой площадке. Во-первых, предпринимателю не стоит завышать целевую сумму, так как проекты с умеренными финансовыми целями демонстрируют более высокую вероятность достижения успеха. Это может быть связано с тем, что непрофессиональные инвесторы менее склонны к риску и их отталкивают крупные суммы. Во-вторых, оптимальной стратегией является выбор более короткой длительности кампании: около 30 дней (среднее значение длительности кампании для успешных проектов), поскольку затяжные проекты менее успешны. В-третьих, особое внимание стоит уделить качеству описания проекта и его рисков. Инвесторы склонны вкладываться в проекты с более подробным описанием, однако оно не должно быть трудным для восприятия. Использование большого количества слов-связок, которые не несут дополнительной информации, также не приветствуется. Непрофессиональные инвесторы, не склонные к риску, чаще хотят подробно ознакомиться с проектом и его рисками перед вложением средств. В-четвертых, предпринимателям важно учитывать и визуальное сопровождение своего проекта. Стоит уделить отдельное внимание изображению, которое прикрепляется к проекту, так как его качество и привлекательность могут значительно повлиять на вероятность успеха.

Также стоит учесть некоторые ограничения исследования, которые могли повлиять на результат. Во-первых, использование модели ModernBERT для анализа настроения текстов имеет ряд ограничений. Она выделяет три категории: положительное настроение, нейтральное и негативное, не давая непрерывного показателя. Это приводит к несбалансированности данных: более 90% проектов имели положительную окраску, что не позволяет повысить прогностическую способность модели за счет этого параметра. Кроме того, модель ModernBERT обучена на корпусе текстов общего назначения и может не учитывать специфику описания краудфандингового проекта. Во-вторых, показатель логичности текста оценивался как отношение количества слов-связок ко всем словам в тексте, однако такой подход, хотя и является базовым в научной литературе, не учитывает семантическую логичность и информативность текста. В-третьих, из-за высокой вариативности визуального контента изображения проектов трудно поддаются объективному сравнению, что может искажать выводы моделей. В-четвертых, большинство моделей чувствительны к «шумовым» показателям и подбору гиперпараметров, что могло снизить качество моделей. В-пятых, для масштабирования данных было использовано логарифмирование с применением квартильного метода, который мог сместить данные. В-шестых, при интерпретации вклада визуальных признаков в прогнозную вероятность

использовался качественный анализ изображений с экстремальными значениями соответствующих компонент, что является субъективным методом анализа. В-седьмых, к ограничениям можно также отнести невозможность учета всех факторов, влияющих на вероятность успеха краудфандингового проекта. К таким факторам можно отнести: личные характеристики предпринимателя, предыдущие успешные и неуспешные проекты предпринимателя, географическую удаленность инвестора, внешние экономические факторы и другие [Colombo et al., 2015; Agrawal et al., 2015; Zheng et al., 2014; Snyder et al., 2017; Van Duynhoven et al., 2019].

В качестве дальнейших исследований можно рассмотреть учет дополнительных факторов при оценке прогнозной модели вероятности успеха краудфандинга, более подробный анализ визуального контента проекта, например, анализ видео, а также переход от качественного анализа изображений к количественным методам. Кроме того, дальнейший анализ может строиться на более сложных методах машинного и глубинного обучения, которые позволяют выявлять нестандартные закономерности в данных, что повышает прогностическую способность.

Приложение 1.

Таблица П1.

Интерпретация индекса Ганнинга

Значение	Интерпретация
До 6	Простой текст, подходящий для массового читателя
7–12	Текст не вызовет затруднений у большинства взрослых
13–17	Средняя сложность, нужно приложить незначительные усилия
18–24	Высокая сложность, для образованной аудитории или специалистов конкретного профиля
Более 24	Крайне трудный материал даже для экспертного читателя

Источник: [Вересова, 2024].

Таблица П2.

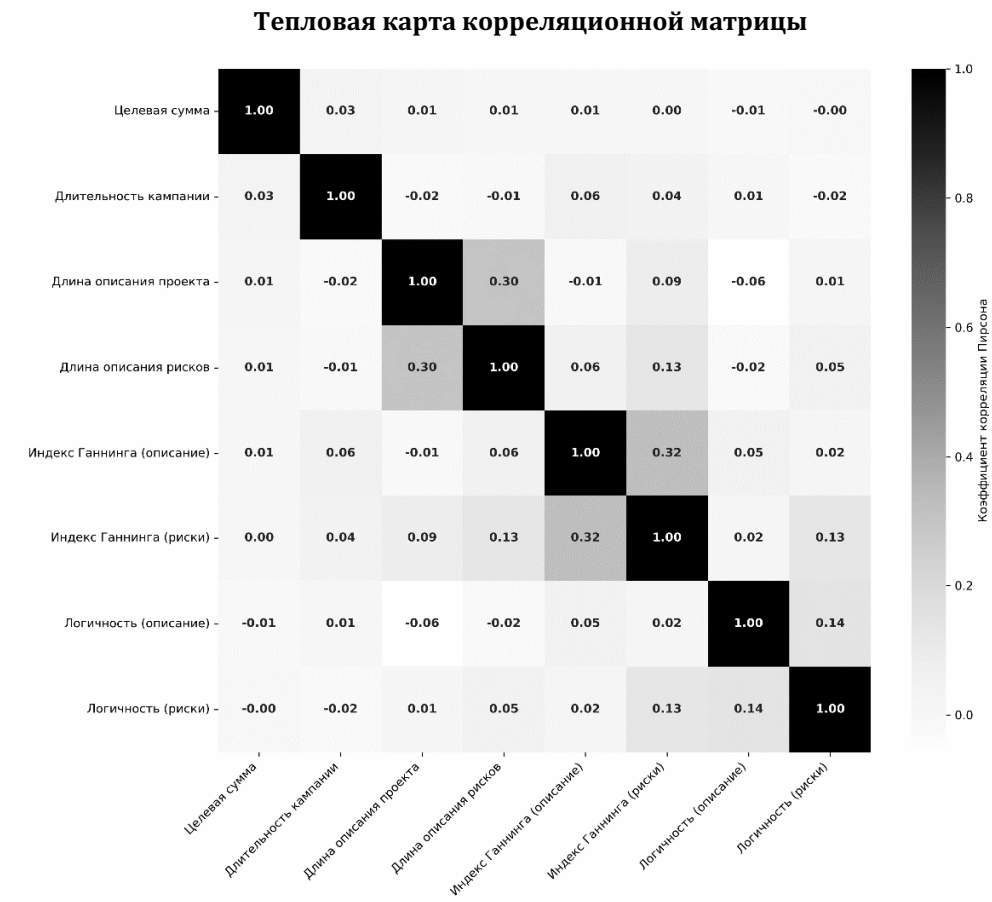


Таблица П3.

Метрики качества оцененных моделей								
Набор факторов	Модель	ROC AUC	95% ДИ	Std Error	Accuracy	Precision	Recall	F1-Score
Базовые	Decision Tree	0,7355	(0,7282; 0,7424)	0,0036	0,6762	0,6475	0,5556	0,5981
Базовые	Random Forest	0,7592	(0,7520; 0,7658)	0,0036	0,6967	0,6608	0,6172	0,6383
Базовые	XGBoost	0,7656	(0,7588; 0,7725)	0,0035	0,6979	0,6607	0,6231	0,6414

Продолжение табл. ПЗ.

Набор факторов	Модель	ROC AUC	95% ДИ	Std Error	Accuracy	Precision	Recall	F1-Score
Базовые	LightGBM	0,7656	(0,7583; 0,7721)	0,0035	0,6988	0,6603	0,6288	0,6442
Базовые	HGBClassifier	0,7646	(0,7576; 0,7711)	0,0035	0,697	0,6576	0,6283	0,6426
Базовые	MLP	0,7559	(0,7488; 0,7629)	0,0036	0,6925	0,6529	0,6207	0,6364
Базовые	Logistic Regression	0,6983	(0,6911; 0,7050)	0,0036	0,6539	0,6173	0,5308	0,5708
Базовые + текстовые	DecisionTree	0,7657	(0,7588; 0,7721)	0,0034	0,6979	0,6499	0,6574	0,6536
Базовые + текстовые	Random Forest	0,8067	(0,8003; 0,8127)	0,0031	0,7317	0,6959	0,6773	0,6865
Базовые + текстовые	XGBoost	0,8147	(0,8084; 0,8211)	0,0031	0,7373	0,6988	0,6926	0,6957
Базовые + текстовые	LightGBM	0,8148	(0,8086; 0,8212)	0,0032	0,7387	0,6995	0,6964	0,6979
Базовые + текстовые	HGBClassifier	0,8132	(0,8070; 0,8194)	0,0032	0,7355	0,6974	0,6888	0,6931
Базовые + текстовые	MLP	0,7763	(0,7697; 0,7829)	0,0034	0,7081	0,6708	0,6416	0,6559
Базовые + текстовые	Logistic Regression	0,7722	(0,7658; 0,7788)	0,0034	0,7051	0,6714	0,6264	0,6481
Базовые + визуальные	DecisionTree	0,7094	(0,7019; 0,7162)	0,0036	0,656	0,651	0,4451	0,5288
Базовые + визуальные	Random Forest	0,7635	(0,7565; 0,7699)	0,0036	0,6881	0,7047	0,4833	0,5733
Базовые + визуальные	XGBoost	0,7892	(0,7825; 0,7956)	0,0034	0,7181	0,6927	0,6286	0,6591
Базовые + визуальные	LightGBM	0,7851	(0,7781; 0,7918)	0,0034	0,7097	0,6816	0,62	0,6494
Базовые + визуальные	HGBClassifier	0,7869	(0,7798; 0,7935)	0,0034	0,7151	0,688	0,6273	0,6563
Базовые + визуальные	MLP	0,6819	(0,6740; 0,6897)	0,004	0,6381	0,5814	0,5898	0,5856
Базовые + визуальные	Logistic Regression	0,7383	(0,7312; 0,7450)	0,0037	0,6771	0,6487	0,557	0,5994
Все	DecisionTree	0,7553	(0,7481; 0,7618)	0,0036	0,6943	0,6465	0,651	0,6487

Окончание табл. ПЗ.

Набор факторов	Модель	ROC AUC	95% ДИ	Std Error	Accuracy	Precision	Recall	F1-Score
Все	Random Forest	0,8032	(0,7964; 0,8094)	0,0033	0,7235	0,7259	0,582	0,646
Все	XGBoost	0,8248	(0,8188; 0,8310)	0,0031	0,7432	0,7114	0,6861	0,6985
Все	LightGBM	0,8249	(0,8188; 0,8311)	0,0031	0,7442	0,7129	0,6867	0,6995
Все	HGBClassifier	0,8222	(0,8157; 0,8281)	0,0031	0,7423	0,7083	0,6897	0,6989
Все	MLP	0,7165	(0,7090; 0,7233)	0,0037	0,6604	0,6104	0,5993	0,6048
Все	Logistic Regression	0,7913	(0,7844; 0,7975)	0,0033	0,7191	0,6895	0,6406	0,6642

Таблица П4.

Результаты проверки на устойчивость методом бутстрапирования

	Среднее абсолютное значение Шепли	Стандартное отклонение	Нижняя граница ДИ (2,5%)	Верхняя граница ДИ (97,5%)
Целевая сумма (логарифм)	0,5775	0,0036	0,5709	0,5847
Длина описания (логарифм)	0,5009	0,0022	0,4963	0,5048
Год	0,2779	0,0013	0,2756	0,2803
Длительность кампании (логарифм)	0,2127	0,0019	0,2092	0,2165
Категория (Музыка)	0,1453	0,0012	0,1429	0,1477
Длина описания рисков (логарифм)	0,1201	0,0006	0,1189	0,1212
РСА 13	0,0981	0,0005	0,0971	0,0990
РСА 6	0,0757	0,0004	0,0749	0,0765
Категория (Фильмы и Видео)	0,0719	0,0007	0,0706	0,0734
РСА 2	0,0685	0,0004	0,0679	0,0692
Логичность описания (логарифм)	0,0622	0,0005	0,0612	0,0632

Окончание табл. П4.

	Среднее абсолютное значений Шепли	Стандартное отклонение	Нижняя граница ДИ (2,5%)	Верхняя граница ДИ (97,5%)
РСА 1	0,0524	0,0004	0,0516	0,0532
РСА 3	0,0503	0,0004	0,0495	0,0510
Категория (Публицистика)	0,0497	0,0006	0,0486	0,0508
Индекс Ганнинга описания (логарифм)	0,0461	0,0004	0,0454	0,0468
Категория (Поделки)	0,0448	0,0007	0,0434	0,0461
Индекс Ганнинга описания рисков (логарифм)	0,0434	0,0003	0,0427	0,0440
Страна (США)	0,0421	0,0002	0,0416	0,0426
РСА 23	0,0419	0,0002	0,0414	0,0423
РСА 30	0,0392	0,0002	0,0388	0,0397

Приложение 2.

Список причинных и соединительных слов, используемый для определения логичности текста

«and, but, because, therefore, thus, so, however, meanwhile, since, although, besides, furthermore, as a result, for example, for instance, in conclusion, in addition, on the other hand, in fact, due to, owing to, despite, while, as well as, nevertheless, not only, also, consequently, accordingly, though, either, nor, either, both, yet, for, even though, on the contrary, in spite of, that is, such as, this means, so that, thus, hence, if, unless, when, whenever, while, before, after, even if, as long as, provided that, as, in order that, in case, on condition that, so as to, with the result that, by virtue of, considering that, meanwhile, because of, on the grounds that, provided, given that, by the way, so far as, according to, as well as, subsequently, therefore, likewise, similarly, correspondingly»

Список составлен авторами на основе словаря LIWC (Linguistic Inquiry and Word Count, Лингвистический анализ и подсчет слов).

* *
*

СПИСОК ЛИТЕРАТУРЫ

- Вересова А. Индекс туманности Ганнинга // Text.ru. 2024. (<https://text.ru/blog/indeks-tumannosti-ganninga>) (Дата обращения: 09.03.2025).
- Сухановский Ю.А., Миронова Е.М. Сравнение подходов к краудфандингу в России и США // Вестник науки и образования. 2019. 10-1 (64). С. 44–47.
- Agrawal A., Catalini C., Goldfarb A. Crowdfunding: Geography, Social Networks, and the Timing of Investment Decisions // Journal of Economics and Management Strategy. 2015. 24. 2. P. 253–274. DOI: <https://doi.org/10.1111/jems.12093>
- Ahlers G.K.C., Cumming D., Günther C., Schweizer D. Signaling in Equity Crowdfunding // Entrepreneurship Theory and Practice. 2015. 39. 4. P. 955–980. DOI: <https://doi.org/10.1111/etap.12157>
- Aleksina A., Akulenko S., Lublóy Á. Success Factors of Crowdfunding Campaigns in Medical Research: Perceptions and Reality // Drug Discovery Today. 2019. 24. 7. P. 1413–1420. DOI: <https://doi.org/10.1016/j.drudis.2019.05.012>
- Antoun W., Sagot B., Seddah D. ModernBERT or DeBERTaV3? Examining Architecture and Data Influence on Transformer Encoder Models Performance. Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics. 2025. P. 3061–3074. DOI: <https://doi.org/10.18653/v1/2025.ijcnlp-long.164>
- Boeuf B., Darveau J., Legoux R. Financing Creativity: Crowdfunding As a New Approach for Theatre Projects // International Journal of Arts Management. 2014. 16. 3. P. 33–48.
- Spare a Quarter? Crowdfunding Sites Turn Fans into Patrons of the Arts // Knowledge at Wharton. 2012. (<https://knowledge.wharton.upenn.edu/article/can-you-spare-a-quarter-crowdfunding-sites-turn-fans-into-patrons-of-the-arts/>) (Accessed: 31.01.2025).
- clapAI. ModernBERT-base-multilingual-sentiment // Hugging Face. 2024. (<https://huggingface.co/clapAI/modernBERT-base-multilingual-sentiment>) (Accessed: 27.04.2026).
- Colombo M.G., Franzoni C., Rossi-Lamastra C. Internal Social Capital and the Attraction of Early Contributions in Crowdfunding // Entrepreneurship Theory and Practice. 2015. 39. 1. P. 75–100. DOI: <https://doi.org/10.1111/etap.12118>
- Connelly B.L., Certo S.T., Ireland R.D., Reutzel C.R. Signaling Theory: A Review and Assessment // Journal of Management. 2011. 37. 1. P. 39–67. DOI: <https://doi.org/10.1177/0149206310388419>
- Cordova A., Dolci J., Gianfrate G. The Determinants of Crowdfunding Success: Evidence from Technology Projects // Procedia – Social and Behavioral Sciences. 2015. 181. P. 115–124. DOI: <https://doi.org/10.1016/j.sbspro.2015.04.872>
- Di Pietro F., Buttice V. Institutional Characteristics and the Development of Crowdfunding Across Countries // International Review of Financial Analysis. 2020. 71. DOI: <https://doi.org/10.1016/j.irfa.2020.101543>
- Elitzur R., Katz N., Muttath P., Soberman D. The Power of Machine Learning Methods to Predict Crowdfunding Success: Accounting for Complex Relationships Efficiently // Journal of Business Venturing Design. 2024. DOI: <https://doi.org/10.1016/j.jbvd.2024.100022>
- Fan-Osuala O., Zantedeschi D., Jank W. Using Past Contribution Patterns to Forecast Fundraising Outcomes in Crowdfunding // International Journal of Forecasting. 2018. 34. 1. P. 30–44. DOI: <https://doi.org/10.1016/j.ijforecast.2017.07.003>
- Ferreira F., Pereira L. Success Factors in a Reward and Equity Based Crowdfunding Campaign. 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC). 2018. P. 1–8. DOI: <https://doi.org/10.1109/ICE.2018.8436308>
- Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., Bouchachia A. A Survey on Concept Drift Adaptation // ACM Computing Surveys. 2014. 46. 4. P. 1–37. DOI: <https://doi.org/10.1145/2523813>

- Gorn G.J., Chattopadhyay A., Yi T., Dahl D.W. Effects of Color As an Executional Cue in Advertising: They're in the Shade // *Management Science*. 1997. 43. 10. P. 1387–1400. DOI: <https://doi.org/10.1287/mnsc.43.10.1387>
- Hair J.F., Black W.C., Babin B.J., Anderson R.E. *Multivariate Data Analysis*. 2019.
- He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016. P. 770–778. DOI: <https://doi.org/10.1109/CVPR.2016.90>
- Jiang C., Han R., Xu Q., Liu Y. The Impact of Soft Information Extracted from Descriptive Text on Crowdfunding Performance // *Electronic Commerce Research and Applications*. 2020. 43. DOI: <https://doi.org/10.1016/j.elerap.2020.101002>
- Kaminski J.C., Hopp C. Predicting Outcomes in Crowdfunding Campaigns with Textual, Visual, and Linguistic Signals // *Small Business Economics*. 2020. 55. 3. P. 627–649. DOI: <https://doi.org/10.1007/s11187-019-00218-w>
- Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017. P. 3149–3157.
- Kim T., Por M.H., Yang S.B. Winning the Crowd in Online Fundraising Platforms: The Roles of Founder and Project Features // *Electronic Commerce Research and Applications*. 2017. 25. P. 86–94. DOI: <https://doi.org/10.1016/j.elerap.2017.09.002>
- Koch J.A., Siering M. Crowdfunding Success Factors: The Characteristics of Successfully Funded Projects on Crowdfunding Platforms. *Proceedings of the 23rd European Conference on Information Systems (ECIS 2015)*. 2015.
- Labrecque L.I., Milne G.R. Exciting Red and Competent Blue: The Importance of Color in Marketing // *Journal of the Academy of Marketing Science*. 2012. 40. 5. P. 711–727. DOI: <https://doi.org/10.1007/s11747-010-0245-y>
- Li L., Yang L., Zhao M., Liao M., Cao Y. Exploring the Success Determinants of Crowdfunding for Cultural and Creative Projects: An Empirical Study Based on Signal Theory // *Technology in Society*. 2022. 70. DOI: <https://doi.org/10.1016/j.techsoc.2022.102036>
- Liu Z., Hu S. Predicting the Fundraising Performance of Environmental Crowdfunding Projects: An Interpretable Machine Learning Approach // *Information Processing and Management*. 2024. 61. 2. DOI: <https://doi.org/10.1016/j.ipm.2023.103587>
- Lu J., Liu A., Dong F., Gu F., Gama J., Zhang G. Learning under Concept Drift: A Review // *IEEE Transactions on Knowledge and Data Engineering*. 2018. 31. 12. P. 2346–2363.
- Mollick E. The Dynamics of Crowdfunding: An Exploratory Study // *Journal of Business Venturing*. 2014. 29. 1. P. 1–16. DOI: <https://doi.org/10.1016/j.jbusvent.2013.06.005>
- Pieters R., Wedel M. Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects // *Journal of Marketing*. 2004. 68. 2. P. 36–50. DOI: <https://doi.org/10.1509/jmkg.68.2.36.27794>
- Ralcheva A., Roosenboom P. Forecasting Success in Equity Crowdfunding // *Small Business Economics*. 2020. 55. 1. P. 39–56. DOI: <https://doi.org/10.1007/s11187-019-00144-x>
- Smith A.M. The log-0 Problem: Analysis Strategies and Options for Choosing c in log(y + c). *Rbind.io*. 2018. (<https://aosmith.rbind.io/2018/09/19/the-log-0-problem/>) (Accessed: 31.01.2025).
- Snyder J., Crooks V.A., Mathers A., Chow-White P. Appealing to the Crowd: Ethical Justifications in Canadian Medical Crowdfunding Campaigns // *Journal of Medical Ethics*. 2017. 43. 6. P. 364–367. DOI: <https://doi.org/10.1136/medethics-2016-103933>
- Song Y., Berger R., Yosipof A., Barnes B.R. Mining and Investigating the Factors Influencing Crowdfunding Success // *Technological Forecasting and Social Change*. 2019. 148. DOI: <https://doi.org/10.1016/j.techfore.2019.119723>
- Spence M. Job Market Signaling // Diamond P., Rothschild M. (eds.) *Uncertainty in Economics*. Academic Press. 1978. P. 281–306. DOI: <https://doi.org/10.1016/B978-0-12-214850-7.50025-5>
- Statista *Market Insights*. Reward-Based Crowdfunding. *Statista Outlook*. 2025. (<https://www.statista.com/outlook/fmo/capital-raising/digital-capital-raising/reward-based-crowdfunding/worldwide#capital-raised>) (Accessed: 17.04.2026).

- Statista Research Department. Kickstarter Project Funding Success Rate 2025. Statista. 2025. (<https://www.statista.com/statistics/235405/kickstarter-project-funding-success-rate/>) (Accessed: 17.04.2026).
- Steigenberger N., Wilhelm H. Extending Signaling Theory to Rhetorical Signals: Evidence from Crowdfunding // *Organization Science*. 2018. 29. 3. P. 529–546. DOI: <https://doi.org/10.1287/orsc.2017.1195>
- Tang Z., Yang Y., Li W., Lian D., Duan L. Deep Cross-Attention Network for Crowdfunding Success Prediction // *IEEE Transactions on Multimedia*. 2022. 25. P. 1306–1319. DOI: <https://doi.org/10.1109/TMM.2022.3141256>
- Van Duynhoven A., Lee A., Michel R., Snyder J., Crooks V.A., Chow-White P., Schuurman N. Spatially Exploring the Intersection of Socioeconomic Status and Canadian Cancer-Related Medical Crowdfunding Campaigns // *BMJ Open*. 2019. 9. 6. DOI: <https://doi.org/10.1136/bmjopen-2018-026365>
- Wang R., Xu H., Zhang X. Impact of Image Content on Medical Crowdfunding Success: A Machine Learning Approach // *Journal of Medical Internet Research*. 2024. 26. DOI: <https://doi.org/10.2196/58617>
- Xu B., Zheng H., Xu Y., Wang T. Configurational Paths to Sponsor Satisfaction in Crowdfunding // *Journal of Business Research*. 2016. 69. 2. P. 915–927. DOI: <https://doi.org/10.1016/j.jbusres.2015.06.040>
- Zhai Y., Shen W. Signaling Theory in Charity-Based Crowdfunding: Investigating the Effects of Project Creator Characteristics and Text Linguistic Style on Fundraising Performance // *Heliyon*. 2024. 10. 4. DOI: <https://doi.org/10.1016/j.heliyon.2024.e25756>
- Zheng H., Li D., Wu J., Xu Y. The Role of Multidimensional Social Capital in Crowdfunding: A Comparative Study in China and US // *Information and Management*. 2014. 51. 4. P. 488–496. DOI: <https://doi.org/10.1016/j.im.2014.03.003>
- Zhou M., Lu B., Fan W., Wang G.A. Project Description and Crowdfunding Success: An Exploratory Study // *Information Systems Frontiers*. 2018. 20. P. 259–274. DOI: <https://doi.org/10.1007/s10796-016-9723-1>

Multimodal Prediction of Crowdfunding Project Success: The Signaling Role of Text and Visual Content

Valeria Fedorova¹, Marina Zavertyaeva²

¹ National Research University Higher School of Economics.
38, office 311, Studencheskaya St., Perm, Perm Krai, 614000, Russian Federation.
E-mail: VS.Fedorova@hse.ru

² National Research University Higher School of Economics.
38, office 311, Studencheskaya St., Perm, Perm Krai, 614000, Russian Federation.
E-mail: moskolkova@hse.ru

Crowdfunding is a relatively new method of project financing that involves raising funds from a large number of non-professional investors. Despite the rapid growth of this sector, only about 41.98% of crowdfunding campaigns are successful, which highlights the need to identify factors that increase the likelihood of successful fundraising. This study aims to address the question of which textual and visual characteristics of projects most significantly determine their success and to what extent their inclusion improves forecasting accuracy. The study de-

Valeria Fedorova – Junior Research Fellow, International Laboratory of Intangible-Driven Economy.

Marina Zavertyaeva – PhD (Economics), Professor, Department of Economics and Finance; Senior Research Fellow, International Laboratory of Intangible-Driven Economy.

velops a multimodal model for predicting the likelihood of fundraising on the Kickstarter platform. The study hypothesizes that the inclusion of textual (*H1*) and visual (*H2*) characteristics significantly improves the predictive power of the models, and their joint integration (*H3*) ensures the best forecast quality by incorporating additional information. The empirical analysis relies on a dataset of 89,557 projects launched on Kickstarter between 2010 and 2024.

In addition to standard project characteristics commonly used in studies of crowdfunding success (funding goal, campaign duration, category, country, and launch year), the analysis incorporates an extended set of textual and visual factors. Textual features are extracted using natural language processing techniques, including sentiment analysis, readability, and logicity measures applied to project descriptions and risk disclosures. Visual features are obtained using a pretrained ResNet50 neural network, followed by dimensionality reduction via principal component analysis. For predictive purposes, logistic regression, decision trees, ensemble methods (Random Forest, XGBoost, LightGBM), and a multilayer perceptron are employed.

The results demonstrate that the inclusion of textual and visual factors significantly improves predictive performance. The best-performing model (LightGBM) achieves a ROC AUC of approximately 0.82. The findings indicate that funding goals, campaign duration, textual characteristics, and visual content exert the strongest influence on the probability of success. Based on these results, the study concludes that informational signals play a significant role in crowdfunding and offers practical recommendations for improving fundraising strategies on crowdfunding platforms.

Key words: crowdfunding; project success prediction; Kickstarter; project financing; textual characteristics; visual characteristics.

JEL Classification: G23, C53, C45.

* *
*

References

- Agrawal A., Catalini C., Goldfarb A. (2015) Crowdfunding: Geography, Social Networks, and the Timing of Investment Decisions. *Journal of Economics and Management Strategy*, 24, 2, pp. 253–274. DOI: <https://doi.org/10.1111/jems.12093>
- Ahlers G.K.C., Cumming D., Günther C., Schweizer D. (2015) Signaling in Equity Crowdfunding. *Entrepreneurship Theory and Practice*, 39, 4, pp. 955–980. DOI: <https://doi.org/10.1111/etap.12157>
- Aleksina A., Akulenkina S., Lublóy Á. (2019) Success Factors of Crowdfunding Campaigns in Medical Research: Perceptions and Reality. *Drug Discovery Today*, 24, 7, pp. 1413–1420. DOI: <https://doi.org/10.1016/j.drudis.2019.05.012>
- Antoun W., Sagot B., Seddah D. (2025) ModernBERT or DeBERTaV3? Examining Architecture and Data Influence on Transformer Encoder Models Performance. *Proceedings of the 14th International Joint Conference on Natural Language Processing and the 4th Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*, pp. 3061–3074. DOI: <https://doi.org/10.18653/v1/2025.ijcnlp-long.164>
- Boeuf B., Darveau J., Legoux R. (2014) Financing Creativity: Crowdfunding As a New Approach for Theatre Projects. *International Journal of Arts Management*, 16, 3, pp. 33–48.

- clapAI. (2024) ModernBERT-base-multilingual-sentiment. *Hugging Face*. Available at: <https://huggingface.co/clapAI/modernBERT-base-multilingual-sentiment> (Accessed: 27.04.2026).
- Colombo M.G., Franzoni C., Rossi-Lamastra C. (2015) Internal Social Capital and the Attraction of Early Contributions in Crowdfunding. *Entrepreneurship Theory and Practice*, 39, 1, pp. 75–100. DOI: <https://doi.org/10.1111/etap.12118>
- Connelly B.L., Certo S.T., Ireland R.D., Reutzel C.R. (2011) Signaling Theory: A Review and Assessment. *Journal of Management*, 37, 1, pp. 39–67. DOI: <https://doi.org/10.1177/0149206310388419>
- Cordova A., Dolci J., Gianfrate G. (2015) The Determinants of Crowdfunding Success: Evidence from Technology Projects. *Procedia – Social and Behavioral Sciences*, 181, pp. 115–124. DOI: <https://doi.org/10.1016/j.sbspro.2015.04.872>
- Di Pietro F., Buttice V. (2020) Institutional Characteristics and the Development of Crowdfunding Across Countries. *International Review of Financial Analysis*, 71. DOI: <https://doi.org/10.1016/j.irfa.2020.101543>
- Elitzur R., Katz N., Muttath P., Soberman D. (2024) The Power of Machine Learning Methods to Predict Crowdfunding Success: Accounting for Complex Relationships Efficiently. *Journal of Business Venturing Design*. DOI: <https://doi.org/10.1016/j.jbvd.2024.100022>
- Fan-Usuala O., Zantedeschi D., Jank W. (2018) Using Past Contribution Patterns to Forecast Fund-raising Outcomes in Crowdfunding. *International Journal of Forecasting*, 34, 1, pp. 30–44. DOI: <https://doi.org/10.1016/j.ijforecast.2017.07.003>
- Ferreira F., Pereira L. (2018) Success Factors in a Reward and Equity Based Crowdfunding Campaign. *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, pp. 1–8. DOI: <https://doi.org/10.1109/ICE.2018.8436308>
- Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., Bouchachia A. (2014) A Survey on Concept Drift Adaptation. *ACM Computing Surveys*, 46, 4, pp. 1–37. DOI: <https://doi.org/10.1145/2523813>
- Gorn G.J., Chattopadhyay A., Yi T., Dahl D.W. (1997) Effects of Color As an Executional Cue in Advertising: They're in the Shade. *Management Science*, 43, 10, pp. 1387–1400. DOI: <https://doi.org/10.1287/mnsc.43.10.1387>
- Hair J.F., Black W.C., Babin B.J., Anderson R.E. (2019) *Multivariate Data Analysis*.
- He K., Zhang X., Ren S., Sun J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778. DOI: <https://doi.org/10.1109/CVPR.2016.90>
- Jiang C., Han R., Xu Q., Liu Y. (2020) The Impact of Soft Information Extracted from Descriptive Text on Crowdfunding Performance. *Electronic Commerce Research and Application*, 43. DOI: <https://doi.org/10.1016/j.elerap.2020.101002>
- Kaminski J.C., Hopp C. (2020) Predicting Outcomes in Crowdfunding Campaigns with Textual, Visual, and Linguistic Signals. *Small Business Economics*, 55, 3, pp. 627–649. DOI: <https://doi.org/10.1007/s11187-019-00218-w>
- Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 3149–3157.
- Kim T., Por M.H., Yang S.B. (2017) Winning the Crowd in Online Fundraising Platforms: The Roles of Founder and Project Features. *Electronic Commerce Research and Applications*, 25, pp. 86–94. DOI: <https://doi.org/10.1016/j.elerap.2017.09.002>
- Koch J.A., Siering M. (2015) Crowdfunding Success Factors: The Characteristics of Successfully Funded Projects on Crowdfunding Platforms. *Proceedings of the 23rd European Conference on Information Systems (ECIS 2015)*.
- Labrecque L.I., Milne G.R. (2012) Exciting Red and Competent Blue: The Importance of Color in Marketing. *Journal of the Academy of Marketing Science*, 40, 5, pp. 711–727. DOI: <https://doi.org/10.1007/s11747-010-0245-y>
- Li L., Yang L., Zhao M., Liao M., Cao Y. (2022) Exploring the Success Determinants of Crowdfunding for Cultural and Creative Projects: An Empirical Study Based on Signal Theory. *Technology in Society*, 70. DOI: <https://doi.org/10.1016/j.techsoc.2022.102036>

- Liu Z., Hu S. (2024) Predicting the Fundraising Performance of Environmental Crowdfunding Projects: An Interpretable Machine Learning Approach. *Information Processing and Management*, 61, 2. DOI: <https://doi.org/10.1016/j.ipm.2023.103587>
- Lu J., Liu A., Dong F., Gu F., Gama J., Zhang G. (2018) Learning under Concept Drift: A Review. *IEEE Transactions on Knowledge and Data Engineering*, 31, 12, pp. 2346–2363.
- Mollick E. (2014) The Dynamics of Crowdfunding: An Exploratory Study. *Journal of Business Venturing*, 29, 1, pp. 1–16. DOI: <https://doi.org/10.1016/j.jbusvent.2013.06.005>
- Pieters R., Wedel M. (2004) Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects. *Journal of Marketing*, 68, 2, pp. 36–50. DOI: <https://doi.org/10.1509/jmkg.68.2.36.27794>
- Ralcheva A., Roosenboom P. (2020) Forecasting Success in Equity Crowdfunding. *Small Business Economics*, 55, 1, pp. 39–56. DOI: <https://doi.org/10.1007/s11187-019-00144-x>
- Smith A.M. (2018) The log-0 Problem: Analysis Strategies and Options for Choosing c in $\log(y + c)$. *Rbind.io*. Available at: <https://aosmith.rbind.io/2018/09/19/the-log-0-problem/> (Accessed: 31.01.2025).
- Snyder J., Crooks V.A., Mathers A., Chow-White P. (2017) Appealing to the Crowd: Ethical Justifications in Canadian Medical Crowdfunding Campaigns. *Journal of Medical Ethics*, 43, 6, pp. 364–367. DOI: <https://doi.org/10.1136/medethics-2016-103933>
- Song Y., Berger R., Yosipof A., Barnes B.R. (2019) Mining and Investigating the Factors Influencing Crowdfunding Success. *Technological Forecasting and Social Change*, 148. DOI: <https://doi.org/10.1016/j.techfore.2019.119723>
- Spence M. (1978) Job Market Signaling. *Uncertainty in Economics*. (eds. P. Diamond, M. Rothschild). Academic Press. P. 281–306. DOI: <https://doi.org/10.1016/B978-0-12-214850-7.50025-5>
- Statista Market Insights (2025) Reward-Based Crowdfunding. *Statista Outlook*. Available at: <https://www.statista.com/outlook/fmo/capital-raising/digital-capital-raising/reward-based-crowdfunding/worldwide#capital-raised> (Accessed: 17.04.2026).
- Statista Research Department (2025) Kickstarter Project Funding Success Rate 2025. *Statista*. Available at: <https://www.statista.com/statistics/235405/kickstarter-project-funding-success-rate/> (Accessed: 17.04.2026).
- Steigenberger N., Wilhelm H. (2018) Extending Signaling Theory to Rhetorical Signals: Evidence from Crowdfunding. *Organization Science*, 29, 3, pp. 529–546. DOI: <https://doi.org/10.1287/orsc.2017.1195>
- Sukhanovskiy Yu.A., Mironova E.M. (2019) Comparison of Crowdfunding Approaches in Russia and the USA. *Bulletin of Science and Education*, 10–1 (64), pp. 44–47. (in Russia).
- Tang Z., Yang Y., Li W., Lian D., Duan L. (2022) Deep Cross-Attention Network for Crowdfunding Success Prediction. *IEEE Transactions on Multimedia*, 25, pp. 1306–1319. DOI: <https://doi.org/10.1109/TMM.2022.3141256>
- Van Duynhoven A., Lee A., Michel R., Snyder J., Crooks V.A., Chow-White P., Schuurman N. (2019) Spatially Exploring the Intersection of Socioeconomic Status and Canadian Cancer-Related Medical Crowdfunding Campaigns. *BMJ Open*, 9, 6. DOI: <https://doi.org/10.1136/bmjopen-2018-026365>
- Veresova A. (2024) Gunning fog index. *Text.ru*. Available at: <https://text.ru/blog/indeks-tumannosti-ganninga> (Accessed: 09.03.2025). (in Russia).
- Wang R., Xu H., Zhang X. (2024) Impact of Image Content on Medical Crowdfunding Success: A Machine Learning Approach. *Journal of Medical Internet Research*, 26. DOI: <https://doi.org/10.2196/58617>
- Xu B., Zheng H., Xu Y., Wang T. (2016) Configurational Paths to Sponsor Satisfaction in Crowdfunding. *Journal of Business Research*, 69, 2, pp. 915–927. DOI: <https://doi.org/10.1016/j.jbusres.2015.06.040>
- Zhai Y., Shen W. (2024) Signaling Theory in Charity-Based Crowdfunding: Investigating the Effects of Project Creator Characteristics and Text Linguistic Style on Fundraising Performance. *Heliyon*, 10, 4. DOI: <https://doi.org/10.1016/j.heliyon.2024.e25756>
- Zheng H., Li D., Wu J., Xu Y. (2014) The Role of Multidimensional Social Capital in Crowdfunding: A Comparative Study in China and US. *Information and Management*, 51, 4, pp. 488–496. DOI: <https://doi.org/10.1016/j.im.2014.03.003>
- Zhou M., Lu B., Fan W., Wang G.A. (2018) Project Description and Crowdfunding Success: An Exploratory Study. *Information Systems Frontiers*, 20, pp. 259–274. DOI: <https://doi.org/10.1007/s10796-016-9723-1>