

# Обзор литературы по теме "Построение скоринговой модели"

Максименко Елизавета

БЭАД 223

## Введение

В современном мире одной из важнейших задач финансовых организаций является определение кредитоспособности клиента. Банкам очень важно минимизировать риски от выдачи кредитов, в связи с чем возникает необходимость выявления вероятности дефолта каждого конкретного субъекта. Для этого финансовые организации разрабатывают скоринговые модели, позволяющие делить заёмщиков на "хороших" и "плохих".

Моей задачей стала разработка такой модели, в связи с чем появилась необходимость изучения доступной литературы из различных источников. Данная работа представляет собой обзор статей, значимых для более полного понимания проблематики и отдельных этапов создания и проверки скоринговых карт. Моей целью на данный момент является выявление недостатков существующих скоринговых моделей и сравнение полученных результатов других исследователей. Также для меня важно выделить положительные аспекты рассматриваемых статей, чтобы в будущем учесть это в своей работе.

Непосредственно выбор темы обусловлен тем, что кредитный скоринг захватывает сразу две сферы: экономическую и машинного обучения. Для меня, как для студента программы "Экономика и анализ данных объединяющей факультеты экономических и компьютерных наук, это наиболее актуально, поскольку данная работа даёт возможность объединить знания, полученных в ходе обучения.

Теоретическая значимость данного литературного обзора заключается в выявлении недостатков существующих моделей и изучении пробелов в исследованиях по этой теме.

Практическая значимость заключается в дальнейшем применении данного обзора для построения скоринговой модели на примере существующих разработок, но с учётом их недостатков и внедрением новых, ещё не рассмотренных методов.

Выбирая материал для данного обзора я столкнулась с явным недостатком статей по разработке кредитного скоринга на русском языке за последние годы. Соответственно, новизна моей работы будет заключаться не только в учёте ошибок предыдущих разработчиков, но и в анализе иностранных работ и нацеленности на русскоязычную публику.

## Статья №1

### "An ensemble credit scoring model based on logistic regression with heterogeneous balancing and weighting effects"

Авторы: Zhang Runchi, Xue Liguu, Wang Qin

#### Оценка содержания:

Статья представляет собой исследование в области кредитного скоринга. Основная цель статьи - представить новую модель Logistic-BWE, основанную на логистической регрессии с разнообразными эффектами балансировки и взвешивания для улучшения способности модели распознавать дефолтные случаи в условиях дисбаланса выборки.

#### Положительные аспекты:

- Чёткое описание модели Logistic-BWE
- Хорошая структурированность
- Использование различных наборов данных для тестирования и сравнения с эталонными моделями.

#### Оценка результатов:

##### Положительные аспекты:

- Разработка новой модели Logistic-BWE, которая демонстрирует превосходство в распознавании дефолтных случаев и общей точности прогнозирования на различных наборах данных.

##### Отрицательные аспекты:

- Ограничения, связанные с применением только одной методологии балансировки

#### Сравнение результатов:

Исследование предоставляет результаты сравнения производительности модели Logistic-BWE с эталонными моделями на различных наборах данных. Оно показывает превосходство Logistic-BWE в распознавании дефолтных случаев и общей точности прогнозирования на большинстве данных.

#### Ссылка на статью :

<https://www.sciencedirect.com/science/article/abs/pii/S0957417422017511>

## Статья №2

### "Interpretable machine learning for imbalanced credit scoring datasets"

Автор: Yujia Chen

#### Оценка содержания:

- Все шаги подкреплены таблицами и схемами для более наглядного понимания
- Присутствует обзор литературы, где представлены работы, использующие различные методы машинного обучения для прогнозирования дефолтов по кредитам
- Важным аспектом было замечание о том, что недостаточно исследований уделяется влиянию дисбаланса классов на методы интерпретации, не привязанные к модели
- Для проверки используются реальные данные из открытых источников, что делает исследование более применимым на практике

#### Оценка результатов:

##### Положительные аспекты:

- Новизна полученных результатов
- Подтверждено, что дисбаланс класса оказывает негативное влияние на интерпретационную производительность как LIME, так и SHAP
- были введены различные несбалансированные методы обучения для улучшения эффективности классификации
- Предлагаемая экспериментальная структура может быть использована для исследований в других сферах

##### Отрицательные аспекты:

- Когда признаки ранжируются в соответствии с величиной их абсолютного значения коэффициентов, их порядок становится более неопределенным и ближе к случайной перестановке. Таким образом, признаки, занимающие верхние позиции в списках, могут получить более высокое значение согласования ранжирования
- Методы повторной выборки могут вызвать такие проблемы, как переобучение (методы избыточной выборки) или потеря информации (методы недостаточной выборки)
- Методы повторной выборки могут увеличить случайность и добавить шума к исходным входным данным, что нежелательно в финансовых операциях и ограничивает их распространение в корпоративной среде

#### Сравнение результатов:

Это первое исследование, которое измеряет стабильность LIME и SHAP с точки зрения классового дисбаланса. Также преимуществом является широкий обзор предшествующих работ и анализ их методологий в контексте кредитного скоринга.

#### Ссылка на статью :

<https://www.sciencedirect.com/science/article/pii/S0377221723005088>

## Статья №3

### "Feature Selection in a Credit Scoring Model"

Автор: Juan Laborda

#### Оценка содержания:

Статья посвящена методам выбора признаков в моделях кредитного скоринга. Она охватывает различные методы отбора признаков, включая фильтры, обертки и встроенные методы, и применяет их к разным алгоритмам классификации, таким как логистическая регрессия, метод опорных векторов, метод ближайших соседей и случайный лес.

#### Оценка результатов:

##### Положительные аспекты:

- Представлены различные методы классификации, что позволяет сравнивать эффективность отбора признаков на различных моделях
- Исследование предлагает методологию, которая может быть использована в практике для улучшения кредитных скоринговых моделей
- Отчетливо демонстрирует важность отбора признаков для улучшения производительности моделей кредитного скоринга.
- Объясняет влияние количества выбранных признаков на производительность моделей
- Методы отбора признаков помогли сделать модели кредитного скоринга более точными

##### Отрицательные аспекты:

- Необходимо больше исследований на других наборах данных и использование других методов отбора признаков для подтверждения результатов
- Не оценивается влияние времени обучения моделей и других практических аспектов
- Обобщение результатов на другие наборы данных может потребовать дополнительных исследований

#### Сравнение результатов:

В данной статье проводится сравнение различных методов построения модели кредитного скоринга с целью выбора наиболее эффективного. Предоставлена методология выбора ключевых переменных для создания модели оценки кредитоспособности.

#### Допущения в исследовании:

- В работе использовались классические методы отбора признаков, что может ограничивать обобщение результатов на более сложные датасеты или на применение в других областях, требующих более сложных моделей
- Оценка моделей производилась с использованием MAE, что, хоть и является показательным метрикой, не учитывает все возможные аспекты производительности модели

#### Ссылка на статью :

<https://www.mdpi.com/2227-7390/9/7/746>

## Статья №4

### "Разработка скоринговой модели оценки кредитного риска"

Автор: Нозимов Э.А.

Оценка содержания:

Исследовалась кредитоспособность заемщиков на основе клиентской базы банка. Было сформировано факторное пространство, позволяющее формализовать и объяснить различия между классами прежних заемщиков. Выявлены факторы, вносящие наибольший вклад в различение классов. В результате проведения анализа разработана скоринговая модель оценки кредитного риска, позволяющая адекватно и быстро оценить кредитоспособность потенциального клиента банка, что способствует росту конкурентоспособности на рынке кредитования.

Оценка результатов:

Положительные аспекты:

- Построенная дискриминантная функция обладает хорошей разделяющей способностью
- Данная функция может быть использована для оценки кредитоспособности клиента
- Предложенная модель может стать эффективным инструментом снижения рисков

Отрицательные аспекты:

- Недостаточно объяснена суть модели
- Отсутствует описание самого построения скоринговой модели
- Нет проверки модели на конкретных данных

Ссылка на статью :

<https://cyberleninka.ru/article/n/razrabotka-skoringovoy-modeli-otsenki-kreditnogo-riska-1>

## Статья №5

### "Внедрение IT-технологий в процесс формирования кредитного рейтинга заёмщика"

Авторы: Морозова Г.В., Жогина К.А.

#### Оценка содержания:

В статье рассмотрены проблемы внедрения информационных технологий в процесс формирования кредитного рейтинга заемщика. исследованы механизмы построения системы кредитного скоринга на основе нейросетевых моделей и их использование в процессе формирования кредитного рейтинга заемщика - физического лица.

#### Оценка результатов:

Положительные аспекты:

- Широкая методологическая база
- Статья объясняет сложности при построении скоринговой модели и работы с ней
- С теоретической точки зрения объяснена необходимость усовершенствования баз данных, предусмотренных для определения кредитного рейтинга заёмщиков

Отрицательные аспекты:

- Отсутствие подтверждения исследованием
- Отсутствие работы с конкретными данными
- Отсутствие практической значимости статьи
- Не предложены способы решения проблем, связанных с работой скоринговых карт

#### Ссылка на статью :

<https://cyberleninka.ru/article/n/vnedrenie-it-tehnologiy-v-protsess-formirovaniya-kreditnogo-reytinga-zayomshchika>

## Статья №6

### "Скоринговые модели как интеллектуальная собственность банка"

Авторы: Мешкова Е.И., Данилова Е.Н.

#### Оценка содержания:

В статье представлена широкая теоретическая база, позволяющая читателю ознакомиться со скоринговыми моделями и особенностями их построения. Объясняется процесс выбора зависимых переменных, формирования выборки, других коэффициентов и показателей. В целом статья может быть полезна для тех, кто хочет разобраться в значении скоринговых моделей для оценки банковских рисков, однако не даёт практических результатов.

#### Оценка результатов:

Положительные аспекты:

- Объяснена выгода банков от скоринговых карт
- Исследованы способы передачи прав на ПО компании, нанявшей разработчика

Отрицательные аспекты:

- Не рассказано о различных методах построения скоринговых карт
- Отсутствует практическая значимость

#### Сравнение результатов:

Скоринговые модели рассмотрены с позиции интеллектуальной собственности. Объяснено, за кем в различных ситуациях закрепляются права на ПО.

#### Ссылка на статью :

<https://cyberleninka.ru/article/n/skoringovye-modeli-kak-intellektualnaya-sobstvennost-banka>

## Статья №7

### "Построение скоринговых карт с использованием модели логистической регрессии"

Автор: Сорокин А.С.

#### Оценка содержания:

Данная статья представляет методику построения скоринговой модели на основе логистической регрессии с использованием различных данных

Положительные аспекты:

- Цели и проблематика чётко изложены.
- Процесс подготовки данных для моделирования подробно описан
- Шаги по построению модели логистической регрессии изложены и подкреплены формулами
- Проведён анализ результатов моделирования, а также оценка точности

Отрицательные аспекты:

- Выбор модели логистической регрессии не обоснован. Указано лишь, что он наиболее часто используется.
- Приведён пример расчета скоринговых баллов лишь для одного атрибута скоринговой карты
- Недостаточная актуальность. Статья была опубликована в 2014 году.

#### Оценка результатов:

Поскольку в статье приведён пример расчета скоринговых баллов лишь для одного атрибута скоринговой карты, невозможно оценить точность результатов, так как как таковых результатов нет. В работе содержится информация о том, как построить скоринговую карту, однако нет приведённого примера скоринговой карты, из-за чего сложно оценить достоверность.

#### Сравнение результатов:

При построении модели логистической регрессии в кредитном скоринге возникает низкая степень точности предсказания отрицательных исходов, т.е. дефолтов.

#### Ссылка на статью :

<https://cyberleninka.ru/article/n/postroenie-skoringovyh-kart-s-ispolzovaniem-modeli-logisticheskoy-regressii>



## Статья №8

### "Особенности построения скоринговой модели на основе аналитической платформы Deductor"

Авторы: Лагереv Д.Г., Бондарева И.В

#### Оценка содержания:

Данная статья представляет методику построения скоринговой модели на основе аналитической платформы Deductor

Положительные аспекты:

- Сформулированы все значимые для понимания термины
- Широкая теоретическая база
- Перечислены этапы построения модели
- Проведён анализ результатов моделирования, а также оценка точности

Отрицательные аспекты:

- В данной статье построение скоринговой модели на основе аналитической платформы Deductor сравнивается исключительно с построением методом логистической регрессии

#### Оценка результатов:

В ходе исследования было выявлено, что данная методика даёт сравнительно точные результаты. Однако точность зависит от количества данных и точности выбора множества значений.

#### Сравнение результатов:

При построении скоринговой модели на основе аналитической платформы Deductor точность результатов не страдает, однако значительно ускоряется процесс вычислений, делая процесс анализа данных практически автоматическим.

#### Ссылка на статью :

<https://cyberleninka.ru/article/n/osobennosti-postroeniya-skoringovoy-modeli-na-osnove-analiticheskoy-platformy-deductor>

## Статья №9

### "Разработка скоринговой модели. Методы классификации заемщиков"

Авторы: Котляр В.П., Антипова Е.А.

#### Оценка содержания:

Положительные аспекты:

- В статье рассмотрен процесс построения скоринговой модели «с нуля»
- Описаны примеры подобных методов, с их подробным описанием
- Приведены варианты решения проблемы реализации скоринговой модели на практике, с приведением конкретной компании разработчика в качестве примера
- Приведены все формулы, используемые в расчётах
- Перечислены различные методы построения скоринговой модели с их подробным описанием

Отрицательные аспекты:

- Недостаточная актуальность. Статья была опубликована в 2014 году.

#### Оценка результатов:

В данной статье перечислены и описаны различные методы построения скоринговой модели, однако не присутствует их непосредственное сравнение, что могло бы добавить статье практическую значимость.

#### Сравнение результатов:

Не проведён сравнительный анализ скоринговых моделей

#### Ссылка на статью :

<https://cyberleninka.ru/article/n/razrabotka-skoringovoy-modeli-metody-klassifikatsii-zaemshikov>

## Статья №10

### "Методологические аспекты построения скоринговых моделей"

Автор: Гринь Н.В.

Оценка содержания:

Положительные аспекты:

- Цели и проблематика чётко изложены.
- Сформулированы все значимые для понимания термины
- Широкая теоретическая база
- Рассматриваются способы преодоления проблем анализа реальных статистических данных

Отрицательные аспекты:

- Недостаточная актуальность. Статья была опубликована в 2012 году.

Оценка результатов:

Достигнут уровень корректных классификаций 83–88%

Сравнение результатов:

- Для повышения точности классификации на основе скоринговых моделей используется алгоритм корректировки обучающей выборки за счет исключения многомерных аномальных наблюдений и алгоритм формирования области неопределенности в принятии решений на основе композиции алгоритмов статистической классификации, а также способ задания границ области неопределенности на основе анализа апостериорных вероятностей классов.
- Проведён сравнительный анализ скоринговых моделей

Ссылка на статью :

<https://elib.grsu.by/katalog/176678-405824.pdf>

## Статья №11

### "Использование нейросетевых моделей в поведенческом скоринге"

Авторы: Сорокин С.В., Сорокин А.С.

#### Оценка содержания:

Положительные аспекты:

- Цели и проблематика чётко изложены.
- Исходные данные для наглядности приведены в табличном виде
- Приведены все формулы, используемые в расчётах
- Присутствуют графики и зарисовки для объяснения работы модуля нейросетевого анализа данных
- Представлены решения проблем переобучения сети
- Объяснена необходимость разделения данных на выборки
- Приведен пример использования одной из программ — SPSS

Отрицательные аспекты:

- Данный метод влечёт переобучение сети, которое, в свою очередь, несёт значительное увеличение ошибок в дальнейшей работе
- Недостаточная актуальность. Статья была опубликована в 2015 году.

#### Оценка результатов:

В статье рассмотрена нейросетевая модель является альтернативой практически ставшей стандартом при построении скоринговых систем логистической регрессии. Объяснено, что использование нелинейных функций в узлах нейронной сети позволяет нейронным сетям моделировать сложные нелинейные зависимости.

#### Сравнение результатов:

Переобучение сети несёт значительное увеличение ошибок, что влечёт дополнительную работу по исправлению этих ошибок

#### Ссылка на статью :

<https://cyberleninka.ru/article/n/ispolzovanie-neyrosetevyh-modeley-v-povedencheskom-skoringe>

## Статья №12

### "Обзор методов кредитного скоринга"

Авторы: Кочеткова В.В., Ефремова К.Д.

#### Оценка содержания:

В представленной статье рассмотрены основные методы кредитного скоринга

Положительные аспекты:

- Дано определение кредитного скоринга
- Описана роль кредитных отношений в экономике
- Рассматривается риск невозврата суммы кредита и его связь с оценкой кредитных рисков через кредитный скоринг
- Описаны четыре основных категории оценки кредитных рисков
- Описывается использование математических и статистических методов для разделения кредитных операций на группы риска с целью предсказания возможных невозвратов и нарушений обязательств заемщиков
- Указывается на разнообразие факторов, влияющих на решение о выдаче кредита

Отрицательные аспекты:

- Отсутствие конкретных примеров
- Некоторые из описанных методов, такие как нейронные сети или метод опорных векторов, могли бы быть более подробно рассмотрены с упоминанием их особенностей и возможных ограничений

#### Оценка результатов:

Статья предоставляет обзор различных методов кредитного скоринга и упоминает специализированное программное обеспечение, предназначенное для оценки кредитных рисков. Однако, эта статья скорее представляет обзорный материал, чем исследование, поэтому оценка конкретных результатов в контексте новых открытий или экспериментов ограничена.

#### Сравнение результатов:

В данной статье отсутствуют конкретные результаты исследования или анализа эффективности этих методов в реальных условиях

#### Допущения в статье:

- Обобщение описанных методов: В статье представлен обзор различных методов кредитного скоринга, но без конкретных контекстуальных примеров или подробных случаев их применения.
- Упрощенное представление о программном обеспечении: Предоставление перечня продуктов без детального анализа их функциональности и возможностей.
- Статья не описывает конкретные сценарии, в которых те или иные методы могут быть более или менее эффективны

#### Ссылка на статью :

<https://cyberleninka.ru/article/n/obzor-metodov-kreditnogo-skoringa>