

# Прогнозирование банкротств контрагентов на основе данных платежной дисциплины

**Ахrameев Максим Дмитриевич,**

*аспирант Российской академии народного хозяйства и государственной  
службы при Президенте Российской Федерации*

*E-mail: akhrameevmd@gmail.com*

**Стефановский Дмитрий Владимирович,**

*кандидат технических наук, доцент кафедры системного анализа и информатики Российской  
академии народного хозяйства и государственной службы при Президенте Российской Федерации*

*E-mail: stefanovskiy-dv@ranepa.ru*

**Сенько Олег Валентинович,**

*доктор физико-математических наук, ведущий научный сотрудник, Федеральный  
исследовательский центр «Информатика и управление» Российской Академии Наук*

*E-mail: senkoov@mail.ru*

В данной статье исследуется проблема прогнозирования банкротств фирм с использованием данных о платежной дисциплине. Большинство предыдущих исследователей использовали в качестве источника данных бухгалтерский баланс, тогда как данные о платежной дисциплине позволяют сократить время до принятия решения по фирме, а также получить оценки благонадежности, основываясь на данных другого типа. Для прогнозирования банкротств фирм была предложена новая методика работы с сильно несбалансированными данными, заключающаяся в обучении классификаторов на специально сформированных подвыборках и усреднении полученных результатов. В роли классификатора на подвыборках выступал случайный лес (Random forest), а для проверки качества модели использовался AUC-score, который показал хорошие результаты.

**Akhrameev Maxim D.,**

*postgraduate student, The Russian Presidential Academy of National Economy and Public Administration*

*E-mail: akhrameevmd@gmail.com*

**Stefanovsky Dmitry V.,**

*candidate of Sciences (Technical), associate Professor of the Department of system analysis and  
informatics of the Russian Presidential Academy of National Economy and Public Administration*

*E-mail: stefanovskiy-dv@ranepa.ru*

**Senko Oleg V.,**

*Doctor of physical and mathematical Sciences, leading researcher,*

*Federal research center «Informatics and management» of the Russian Academy of Sciences*

*E-mail: senkoov@mail.ru*

## FORECASTING BANKRUPTCIES OF COUNTERPARTIES BASED ON PAYMENT DISCIPLINE DATA

In this article, we study the problem of forecasting bankruptcy of firms using data on payment discipline. Most previous researchers used the balance sheet as a data source, while data on payment

discipline will reduce the time before making a decision on the firm, as well as obtain reliability ratings based on other types of data. To predict bankruptcy of the firms proposed a new method of work with highly unbalanced data, which consists in training the classifiers on the automatically generated sub-sample and averaging the obtained results. Random forest served as a classifier for subsamples, and AUC-score was used to check the quality of the model, which showed good results.

**Ключевые слова:** классификация, случайный лес, прогнозирование банкротств, несбалансированная выборка.

**Keywords:** classification, random forest, bankruptcy forecasting, unbalanced sample.

## ПОДХОДЫ ПРЕДЫДУЩИХ ИССЛЕДОВАТЕЛЕЙ

На сегодняшний день не существует единой общепризнанной методологии прогнозирования банкротства. В большинстве существующих практик установление факта несостоятельности предприятия происходит, когда все признаки банкротства уже слишком явно проявили себя. Традиционно, для прогнозирования предбанкротного состояния используют данные бухгалтерской отчетности, на их основе рассчитываются коэффициенты, анализ которых и позволяют сделать вывод о финансовом состоянии. Однако все эти данные описывают уже случившееся с большим временным лагом, поэтому дать достоверную текущую оценку финансовому состоянию предприятия порой очень сложно. К тому же, бухгалтерские данные не отражают всей информации, характеризующей деятельность предприятия, не учитывают окружение предприятия и его тип.

Обилие существующих методик приводит к тому, что одно и то же исследуемое предприятие может быть отнесено к банкротам по результатам одной модели, а по результатам другой модели — нет. И в зарубежной, и в отечественной экономической литературе, изучением данной проблемы занимались такие ученые, как Э. Альтман, Д. Конан, М. Голдер и многие другие. Было предложено большое количество методов, в основном основанных на дискриминантном анализе; статистической информации о финансовом состоянии некоторого количества предприятий, часть которых обанкротилась, а другая смогла улучшить результаты деятельности и восстановить свое финансовое положение.

Для развивающихся рынков, Э. Альтманом в 1968 г. была предложена скоринго-

вая модель, которая впоследствии применялась и в России. Модель получила название «Z-Score» [10]. Модель использует следующие переменные:

- $x_1$ : Отношение оборотного капитала к активам — **Working Capital/Total Assets**
- $x_2$ : Отношение нераспределенной прибыли к активам — **Retained Earnings/Total Assets**
- $x_3$ : Отношение прибыли до вычета налогов и процентов к активам — **EBIT/Total Assets**
- $x_4$ : Отношение капитала к обязательствам — **Equity Book Value/Total Liabilities**.

Для конкретных случаев ее применение зависит от особенности компании, выделяются три большие категории компаний: производственные; производственные, акции которых котируются на биржах; производственные, акции которых не котируются на биржах.

Модель для непроизводственных фирм в развивающихся экономиках выглядит следующим образом:

$$EMZ = 6,56x_1 + 3,26x_2 + 6,72x_3 + 1,05x_4 + 3,25$$

При  $EMZ \geq 2,9$  — вероятность банкротства незначительна, компания финансово устойчива;  $1,2 < EMZ < 2,9$  — ситуация не определена;  $EMZ \leq 1,2$  — ситуация критична, с высокой долей вероятности предприятие обанкротится в ближайшей перспективе.

Дальнейшие исследования проводились в направлении использования эконометрики для прогнозирования. Например, Змиевский в работе [15] использовал данные о 840 компаниях, из которых 40 были банкротами, рабочей моделью была пробит-модель с 3 переменными.

Олсон в своей работе [9] использовал для прогнозирования банкротств данные о 2163 фирмах за период с 1970 по 1976 гг.,

из них 105 фирм были банкротами. Он получил логит-модель с 9 переменными, в которой банкротства прогнозировались за год:

$$z = -1,32 - 0,40x_1 + 6,03x_2 - 1,43x_3 + 0,08x_4 - 2,37x_5 - 1,83x_6 + 0,28x_7 - 1,72x_8 - 0,52x_9$$

- $x_1$ : Логарифм отношения совокупных активов к индексу дефлятору — **TotalAssets/GNP price-level index**,

- $x_2$ : Отношение совокупных обязательств к активам — **Total Liabilities/Total Assets**,

- $x_3$ : Отношение оборотного капитала к активам — **Working Capital/Total Assets**,

- $x_4$ : Отношение краткосрочных обязательств к оборотным активам — **Current Liabilities/Current Assets**,

- $x_5$ : 1, если совокупные обязательства больше активов, иначе 0,

- $x_6$ : Отношение чистой прибыли к активам — **Net income/Total Assets**,

- $x_7$ : Отношение средств от операций к совокупным обязательствам — **Funds from operations/Total Liabilities**,

- $x_8$ : 1, если компания убыточна в течение 2-х последних лет, иначе 0,

- $x_9$ :  $(NI_t - NI_{t-1}) / (|NI_t| + |NI_{t-1}|)$ , где  $NI_t$  — чистая прибыль периода  $t$ .

При решающем правиле, равным 0,5, значение **Accuracy** ( $Accuracy = P/N$ , где  $P$  — число верно классифицированных контрагентов,  $N$  — общее число контрагентов) составило 96,12%, тогда как если бы все фирмы были классифицированы как не банкроты, то значение **Accuracy** было бы равно 91,15%.

Модели, основанные на статистике и дискриминантном анализе, просты и интуитивно понятны, что вполне подходит для потенциальных пользователей, но у них есть свои недостатки, которые проявляются, например, при маленьком размере выборки при моделировании. Исследователи продолжают искать иные способы прогнозирования предбанкротного состояния, и направляют свои усилия на развитие предложенных алгоритмов, вместе с применением современных подходов. В обзорной работе [8] была предложена классификация подходов по видам алгоритмов, например, основан ли алгоритм на статистике, или же использует методы искусственного интеллекта. К статистическим алгоритмам были

отнесены линейный дискриминантный анализ (LDA), многофакторный дискриминантный анализ (MDA), квадратичный дискриминантный анализ (QDA), логистическая регрессия (logit) и факторный анализ (FA).

К «искусственному интеллекту» были отнесены различные нейронные сети с разной архитектурой, как, например, многослойный перцептрон (MLP), решающие деревья (Decision tree), рассуждения на основе прецедентов (case-based reasoning, CBR), эволюционные алгоритмы, метод опорных векторов (SVM), и алгоритмы работы с нечеткими множествами. Подробное описание используемых алгоритмов классификации можно найти в обзорных статьях [7] и [6].

В обзорах [6; 7], авторы также предложили классифицировать исследования по предметным областям, в которых проводилось изучение потенциального банкротства. Было отмечено, что чаще всего предсказывалось банкротство фирм реального сектора экономики, а другим видам бизнеса уделялось меньше внимания. Авторы описывают источники данных и классифицируют переменные, которые используются для прогноза, а также анализируют период, за который собирались данные. Среди 128 исследований, приведенных в обзоре, временной интервал составляет от двух до десяти лет.

В качестве финансовых переменных, исследователи обычно выбирают совокупные активы компании, долговую нагрузку, коэффициенты ликвидности, EBIT (прибыль до вычета процентов и налогов), ROE (коэффициент рентабельности собственного капитала), денежные потоки, и различные соотношения этих показателей.

Все вышеперечисленные модели обладают одним существенным недостатком — используемые данные о бухгалтерском балансе могут появляться не чаще, чем раз в квартал, а некоторые фирмы и вовсе не предоставляют такую отчетность, тогда как бизнес интересуется более оперативная информация о всех потенциальных банкротах, что заставляет исследователей продолжать свою работу.

Другие авторы [1] предлагают альтернативный подход, основанный на платежной дисциплине, то есть добавляют еще один ис-

точник данных для анализа платежеспособности и прогнозирования потенциальных банкротов, в отличие от P. Ravi Kumar, V. Ravi [8], которые в своем обзоре не приводят статьи, основанные на анализе платежной дисциплины. Традиционно, под платежной дисциплиной понимается своевременное и полное осуществление платежей и расчетов. Примером таких данных (их еще называют транзакционными) являются платежные операции фирмы. Возможно, отсутствие статей в обзоре связано с закрытым характером информации о платежной дисциплине, тогда как бухгалтерский баланс обязан находиться в открытом доступе.

Данные о платежной дисциплине позволяют иначе взглянуть на проблему банкротства, а также проводить комплексные исследования и применять различные композиции алгоритмов с использованием данных баланса и платежной дисциплины.

Использование данных платежной дисциплины компании напрямую без предобработки невозможно, поэтому необходимо создать набор агрегатов и уже их использовать в различных задачах классификации. Zeng S. с соавторами [11] использовал данные о платежной дисциплине клиентов четырех разных компаний для прогнозирования просроченных счетов. Каждый счет был описан 54 характеристиками, из которых были отброшены характеристики, уникальные для каждого счета, а также те, которые давали информацию о целевой переменной, в момент времени, в котором этой информации быть не должно. Целевая переменная представляет собой категориальный признак, зависящий от срока просрочки, и для прогнозирования было создано 17 переменных. Также авторы выделяли переменные, которые отображали агрегированную историю каждого счета.

Предложенный авторами список переменных дает достаточную информацию о счете и истории операций до этого. Конечно, для первой операции с новым контрагентом, такие переменные получить нельзя, поэтому все первые операции были удалены из обучения. Для обучения использовалось 3 алгоритма — дерево решений C4.5, наивный байесовский классификатор, а также деревья ре-

шений PART. Для оценки качества применялся критерий *Accuracy*, согласно которому алгоритм PART показал наилучший результат. В работе также изменяются матрицы штрафов за ошибки при обучении таким образом, чтобы достичь лучший результат для классификации счетов с просрочкой более 90 дней, которые составляют только 3% от общего числа счетов. Все эти техники могут оказаться полезными не только при решении задачи классификации потенциально просроченных счетов, но и выявлению потенциальных банкротов.

В следующей научной работе автор [5] исследовал платежную дисциплину клиентов 500 IT-компаний по всему миру, за период 6 месяцев. В выборке было более 4 миллионов счетов, из которых 18,8% были оплачены позже положенного. Большое внимание было уделено препроцессингу, как и в статье [11], не информативные переменные удалялись, созданы новые переменные, добавлены дополнительные переменные, учитывающие историю операций контрагента. Выборка была разделена на обучающую и тестовую в пропорции 80:20, и на обучающей выборке непосредственно применялись алгоритмы машинного обучения: логистическая регрессия, метод k-ближайших соседей (KNN), решающее дерево, случайный лес (Random forest), нейронная сеть. Значение *Accuracy* для случайного леса составило 75%, и так как для других алгоритмов значение *Accuracy* не было указано, случайный лес признавался авторами наилучшим алгоритмом.

Исследование [14] посвящено прогнозированию банкротств компаний на основе их платежной дисциплины. Автор выясняет, насколько информация о платежной дисциплине релевантна для решения задачи прогнозирования банкротств, насколько велик горизонт прогнозирования, какие алгоритмы показывают наилучшие результаты в условиях несбалансированности данных. Использовались данные о более, чем 1,3 млн операций 32562 компаний с пенсионным фондом, из которых 650 компаний стали банкротами (2%).

Переменные для прогноза банкротств выбраны аналогично переменным в работах по прогнозированию просрочек платежей, но так как счета открываются и закрываются



с определенным промежутком времени, переменные усреднялись по месяцам с помощью метода Piecewise Aggregate Approximation (РАА). Данный метод применяется для сокращения размерности, например, для временных рядов [4]. Сравнивая по метрике AUC-score, наилучший результат показал алгоритм случайного леса, а оптимальный горизонт прогнозирования составил 3 месяца.

## ЭКСПЕРИМЕНТ

Для прогнозирования использовались данные по операциям с контрагентами за период более 5 лет, и список должников с указанием дат банкротства. Под должником принимается контрагент, с которым произошла хозяйственная операция, и который не оплачивает счета на протяжении более 90 дней. По экспертной оценке, такая ситуация приравнена к банкротству контрагента.

Для получения переменных, которые будут использованы в модели, данные по операциям обрабатывались следующим образом:

- определены условия включения операций по каждому контрагенту, например, документы с суммой меньше 100 руб. не включаются в анализ;
- для каждой отобранной операции определяется дата погашения, как число дней с момента осуществления операции до погашения;
- формируется признак, показывающий просрочен ли платеж или нет. Этот признак принимает значение 1 в случае, если число дней с момента создания операции до ее оплаты превышало число дней отсрочки более, чем на 5 дней, иначе признак принимает значение 0.

Полная схема прогнозирования банкротств изображена на рис. 1. Стоит отметить возможность параллельной обработки данных, которая возникает на этапе формирования признаков и обучения моделей.

В результате обработки был сформирован набор переменных:

- дебет счета — из исходных данных;
- сумма в документе — из исходных данных;
- сумма проводки — из исходных данных;
- отсрочка по счету — из исходных данных;

• кредитный лимит — из исходных данных;

• *days* — число дней, прошедших с момента отгрузки товара до погашения (может быть отрицательным);

• *late* — просрочен ли платеж, или нет;

• *days.late* — число просроченных дней оплаты по счету;

• *av.late* — сумма переменной «*late*» по всем предыдущим счетам контрагента;

• *av.debet* — сумма переменной «Дебет» по всем предыдущим счетам;

• *av.days* — сумма переменной «*days*» среди всех предыдущих счетов;

• *av.days.late* — среднее число дней просрочки среди всех предыдущих счетов;

• *av.sum* — среднее значение переменной «Сумма в документе» среди всех предыдущих счетов;

• *share\_late.debet* — доля счетов с дебиторской задолженностью среди всех предыдущих;

• *share\_av.late* — доля просроченных платежей среди всех предыдущих.

После создания, переменные по счетам усредняются по месяцам с помощью РАА, и образуются новые переменные:

• *count* — число счетов в этом месяце;

• *der.count* — изменение числа счетов по сравнению с предыдущим месяцем;

• *der.days* — изменение числа дней, прошедших с момента отгрузки товара, по сравнению с прошлым месяцем;

• *der.days.late* — изменение числа просроченных дней по сравнению с прошлым месяцем;

• *der.debet* — изменение числа счетов по дебету по сравнению с прошлым месяцем;

• *der.sum* — изменение средней суммы операции по сравнению с прошлым месяцем;

• *der.late* — изменение средней доли просроченных платежей в этом месяце по сравнению с предыдущим месяцем;

• *der.av.days* — изменение кумулятивной суммы числа дней с момента отгрузок товара по сравнению с прошлым месяцем;

• *der.av.days.late* — изменение кумулятивной суммы числа просроченных дней по сравнению с прошлым месяцем;

• *der.av.sum* — изменение среднего значения переменной «Сумма в документе» по сравнению с прошлым месяцем;

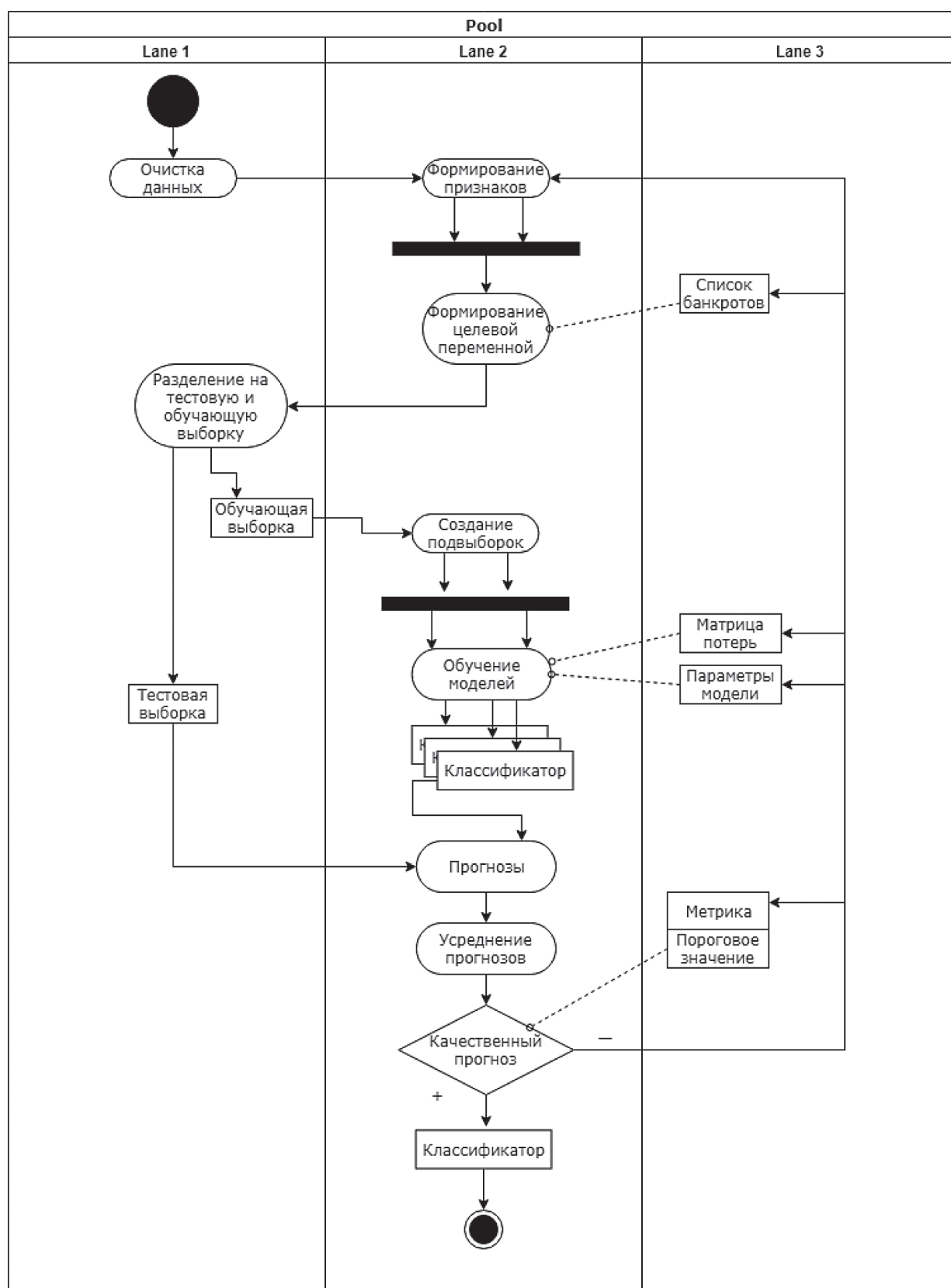


Рисунок 1. Полная схема прогнозирования банкротств

- *der.share\_late.debet* — изменение доли счетов с дебиторской задолженностью среди всех предыдущих счетов по сравнению с прошлым месяцем;

- *der.share\_av.late* — изменение доли просроченных платежей среди всех предыдущих счетов по сравнению с прошлым месяцем.

Все рассчитанные значения переменных сведены в матрицу, в которой столбцы — переменные, а строки — характеристика контрагента в каком-либо месяце. На основе матрицы можно создавать обучающую выборку. В обучающую выборку выбирались только контрагенты с задолженностями более 10000 руб., даты были вручную проверены и исправлены экспертами. После всех манипуляций осталось около 300 контрагентов в списке исторических банкротов. Из них 94 окончательно попали в обучающую выборку следующим способом:

- Список банкротов был отсортирован по датам.

- Из даты первого банкротства вычиталось три месяца. Таким образом, горизонт прогнозирования составляет три месяца, что предполагается оптимальным для принятия превентивных мер в расчетах с потенциальным банкротом. В обучающую выборку включались все наблюдения за этот месяц, значение 1 целевого класса присваивалось банкроту, всем остальным присваивался 0.

- Данные о банкротстве исключались из следующего набора. Процедура повторялась для следующего банкрота из отсортированного списка. После выполнения процедуры были исключены все дублированные наблюдения. Конечный размер обучающей выборки составил более 700 тысячи наблюдений, и, так как на более, чем 700 тысяч наблюдений приходится только 94 наблюдений из целевого класса, данные получились несбалансированными.

В результате обучение происходило на подвыборке — 94 наблюдений целевого класса, и 1000 случайных наблюдений из 700 тысяч. В литературе данный способ работы с несбалансированными данными называется «Under-sampling» [3], который заключается в принудительном уменьшении числа наблюдений, относящихся к нецелевому классу. Еще один способ работы с обучающей выбор-

кой, направленной на борьбу с несбалансированностью данных — «Over-sampling», заключающийся в сэмплировании наблюдений целевого класса, может привести к переобучению при столь малом соотношении целевого и нецелевого класса.

Тысяча случайных наблюдений из 700 тысяч может не содержать в себе всех зависимостей, помогающих определить целевой класс. Для решения этой проблемы предлагается увеличить число обучений на подвыборках до некоторого числа, например, до 100 или 200, а затем усреднить полученные оценки за классы. Тестирование алгоритма осуществляется путем кросс-валидации, с разделением исходной выборки на обучающую и тестовую. Так как при создании переменных теряется временная структура, то никакой сложности с кросс-валидацией не возникает, в отличие от временных рядов, и тестовая выборка может быть выбрана случайно, в размере, например, 20% от исходной. Данный процент выборки никак не принимает участия в обучении.

Проверка алгоритма на тестовой выборке может заключаться в выборе оптимальных значений метрик, таких как *accuracy*, *precision*, *recall*, *f1-score*, который является средним гармоническим между *precision* и *recall*, метрикой AUC, которая представляет собой площадь под кривой ROC [2]. Подробное описание метрик приведено в обзорной статье [13].

Алгоритм — решающий лес, деревья решений которого построены с помощью алгоритма *c50* [12], реализованного в языке R с матрицей потерь  $2 \times 2$  (матрицей ошибок). За наблюдение, ложно отнесенное к классу 0 назначается штраф в 1000, а за наблюдение, ложно отнесенное к классу 1, назначается штраф 1. Такой подход позволяет обучать модель на все еще несбалансированных данных. В нашем исследовании использовался случайный лес, сгенерированный с помощью AdaBoost с максимальным количеством решающих деревьев в комитете, равным 100, причем реальное число деревьев определяется самостоятельно алгоритмом, и может быть меньше 100.

Использование 20% выборки для тестирования на больших данных приводит к большой неопределенности в результатах. Для опре-

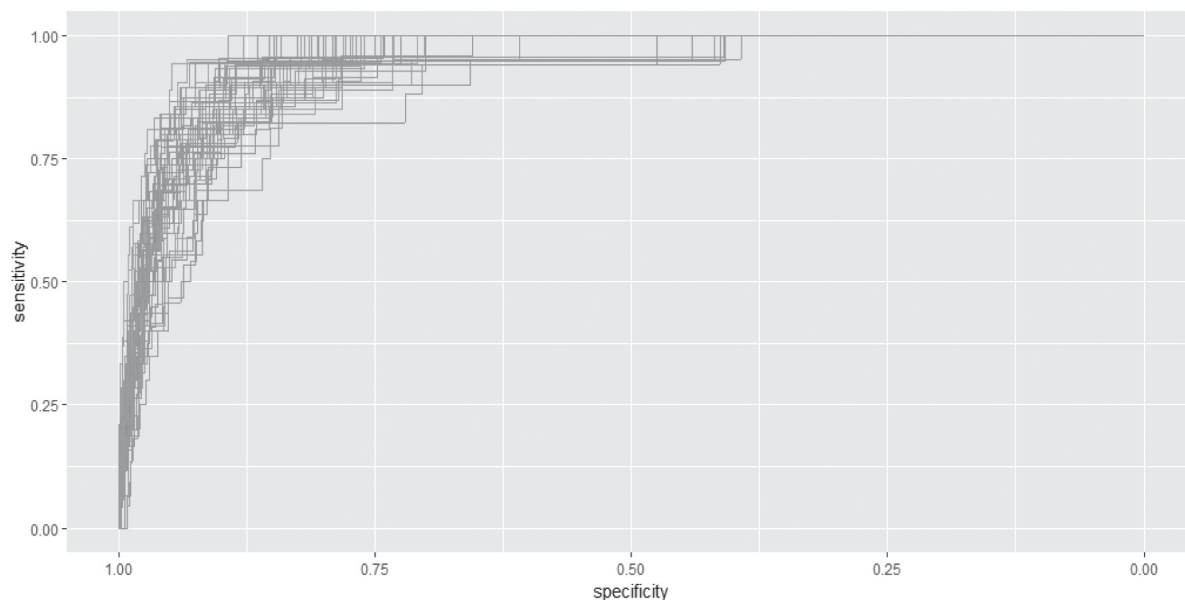


Рисунок 2. ROC-кривые моделей со 200 классификаторами

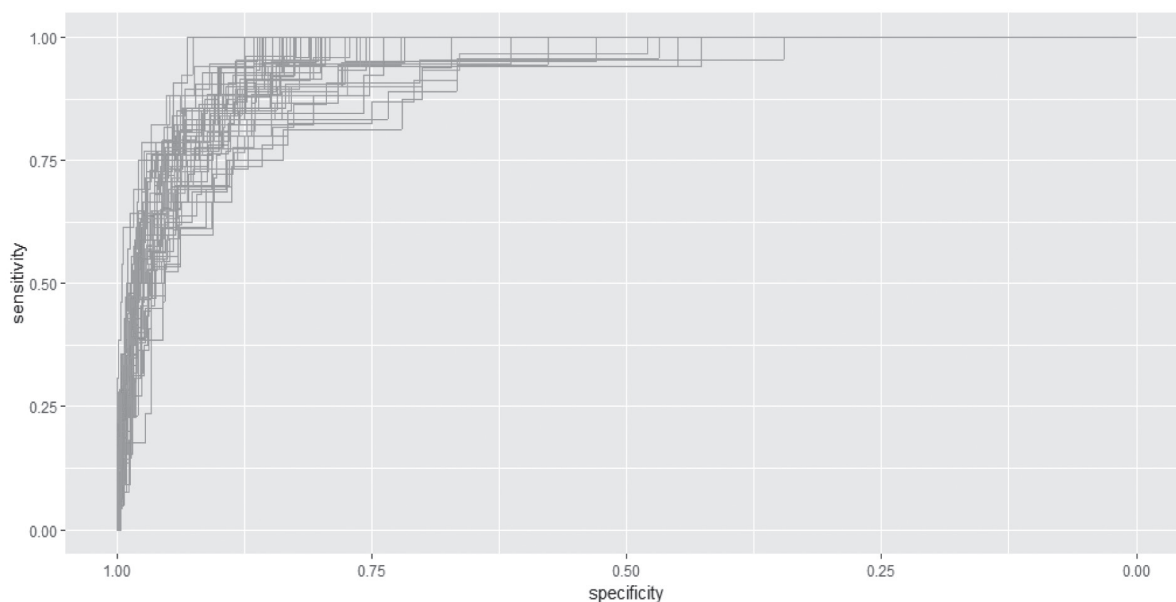


Рисунок 3. ROC-кривые моделей со 100 классификаторами

деления стабильности алгоритма, набор тестовой выборки и проведение обучений повторяется некоторое достаточное количество раз. Число повторений было выбрано равным 50, что обусловлено долгим временем работы алгоритма (несколько часов). Затем проводится построение наборов классификаторов для каждой реализации тестовой и обучающей выборок.

Качество общего прогноза характеризуется метрикой AUC. На *рис. 1* показано 50 ROC-кривых для разных реализаций обучающей и тестовой выборок при числе классификаторов в каждой реализации равном 200.

На *рис. 2* показаны 50 ROC-кривых для разных реализаций обучающей и тестовой выборок при числе классификаторов в каждой реализации равном 100.



Таблица 1. Описательная статистика метрик AUC

Число классификаторов	Число выборок	Среднее	Стандартное отклонение
200	50	0,9462	0,0150
100	50	0,9482	0,01793

Как видно на рис. 2 и 3, а также исходя из табл. 1, большее число классификаторов приводит к меньшему стандартному отклонению метрики AUC, но и чуть меньшему среднему значению метрики.

### ЗАКЛЮЧЕНИЕ

Целью этого исследования являлось моделирование поведения контрагентов для определения потенциальных банкротов с использованием данных о платежной дисциплине.

Результаты исследования показали, что данные о платежной дисциплине могут успешно использоваться для прогнозирования и классификации, несмотря на несбалансированность данных, поскольку средний AUC-score при проверке модели равен 0,948.

Предложен новый способ работы с несбалансированными данными, заключающийся в обучении на специально сформированных выборках и усреднении результатов классификаторов.

Различные преобразования показателей платежной дисциплины, например, кумулятивные суммы и изменения показателей к предыдущему периоду, использовавшиеся в качестве дополнительных переменных, были очень важны в формировании классификаторов. Использование ряда преобразований также позволяет уйти от временной структуры, но в тоже время учитывает историю контрагента.

Внедрение данной модели в существующие системы принятия решений позволит эффективнее взаимодействовать с неплатежеспособными контрагентами и снизит долю просроченных платежей. В будущих исследованиях стоит продолжить развитие данной модели, использовать дополнительные показатели качества, а также использовать другие источники данных вместе с платежной дисциплиной.

**Благодарность.** Статья подготовлена при финансовой поддержке Российского фонда фундаментальных исследований (Грант РФФИ 18-29-03151. Развитие и использование алгебраических, логических и логико-статистических методов обучения по прецедентам для повышения эффективности организаций, оказывающих услуги населению, в цифровой экономике).

### ЛИТЕРАТУРА

1. Andrjovská A., Bánociová A. Payment discipline in business environment// *Procedia Economics and Finance*. 2014. Т. 15. С. 1217–1224.
2. Bekkar M., Djemaa H.K., Alitouche T.A. Evaluation measures for models assessment over imbalanced data sets // *J Inf Eng Appl*. 2013. Т. 3. №10.
3. C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling / C. Drummond, R.C. Holte [и др.] // *Workshop on learning from imbalanced datasets II*. Т. 11. Cite-seer. 2003. С. 1–8.
4. Dimensionality reduction for fast similarity search in large time series databases / E. Keogh [и др.] // *Knowledge and information Systems*. 2001. Т. 3. №3. С. 263–286.
5. Hu W. Overdue invoice forecasting and datamining: дис....канд./ Hu Weikun. Massachusetts Institute of Technology, 2016.
6. Kiang M. Y. A comparative assessment of classification methods // *Decision Support Systems*. 2003. Т. 35. №4. С. 441–454.
7. Kotsiantis S.B., Zaharakis I., Pintelas P. Supervised machine learning: A review of classification techniques// *Emerging artificial intelligence applications in computer engineering*. 2007. Т. 160. С. 3–24.
8. Kumar P.R., Ravi V. Bankruptcy prediction in banks and firms via statistical and intelligent techniques — A review // *European journal of operational research*. 2007. Т. 180. №1. С. 1–28.

9. *Ohlson J.A.* Financial ratios and the probabilistic prediction of bankruptcy // *Journal of accounting research*. 1980. C. 109–131.

10. Predicting financial distress of companies: revisiting the Z-score and ZETA models/ E.I. Altman [и др.] // *Stern School of Business, New York University*, 2000.

11. Predictive modeling for collections of accounts receivable / S. Zeng [и др.] // *Proceedings of the 2007 international workshop on Domain driven data mining*. ACM. 2007. C. 43–48.

12. Quinlan R. C5.0: An Informal Tutorial. 2017 [Электронный ресурс]. — Режим доступа: [www.rulequest.com/see5-unix.html](http://www.rulequest.com/see5-unix.html). (дата обращения: 01.11.2018).

13. *Sokolova M., Lapalme G.* A systematic analysis of performance measures for classification tasks // *Information Processing & Management*. 2009. T. 45. №4. C. 427–437.

14. *Wagenmans F.* Machine Learning in Bankruptcy Prediction: диссертация / *Wagenmans Frank*, 2017.

15. *Zmijewski M.E.* Methodological issues related to the estimation of financial distress prediction models // *Journal of Accounting research*. 1984. C. 59–82.

## REFERENCES

1. *Andrjovská A., Bánociová A.* Payment discipline in business environment// *Procedia Economics and Finance*. 2014. Vol. 15. pp. 1217–1224.

2. *Bekkar M., Djemaa H.K., Alitouche T.A.* Evaluation measures for models assessment over imbalanced data sets // *J Inf Eng Appl*. 2013. Vol. 3. №10.

3. C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling / C. Drummond, R.C. Holte [et al.] // *Workshop on learning from imbalanced datasets II*. Vol. 11. Citeseer. 2003. pp. 1–8.

4. Dimensionality reduction for fast similarity search in large time series databases / E. Ke-

ogh [и др.] // *Knowledge and information Systems*. 2001. Vol. 3. №3. pp. 263–286.

5. *Hu W.* Overdue invoice forecasting and datamining/ *Hu Weikun*. Massachusetts Institute of Technology, 2016.

6. *Kiang M.Y.* A comparative assessment of classification methods // *Decision Support Systems*. 2003. Vol. 35. №4. pp. 441–454.

7. *Kotsiantis S.B., Zaharakis I., Pintelas P.* Supervised machine learning: A review of classification techniques// *Emerging artificial intelligence applications in computer engineering*. 2007. Vol. 160. pp. 3–24.

8. *Kumar P.R., Ravi V.* Bankruptcy prediction in banks and firms via statistical and intelligent techniques -A review // *European journal of operational research*. 2007. Vol. 180. №1. pp. 1–28.

9. *Ohlson J.A.* Financial ratios and the probabilistic prediction of bankruptcy // *Journal of accounting research*. 1980. pp. 109–131.

10. Predicting financial distress of companies: revisiting the Z-score and ZETA models/ E.I. Altman [et al.] // *Stern School of Business, New York University*, 2000.

11. Predictive modeling for collections of accounts receivable / S. Zeng [et al.] // *Proceedings of the 2007 international workshop on Domain driven data mining*. ACM. 2007. pp. 43–48.

12. Quinlan R. C5.0: An Informal Tutorial. 2017 [Electronic resource]. — Access mode: [www.rulequest.com/see5-unix.html](http://www.rulequest.com/see5-unix.html). (date of access: 01.11.2018).

13. *Sokolova M., Lapalme G.* A systematic analysis of performance measures for classification tasks // *Information Processing & Management*. 2009. Vol. 45. №4. pp. 427–437.

14. *Wagenmans F.* Machine Learning in Bankruptcy Prediction: MS thesis / *Wagenmans Frank*, 2017.

15. *Zmijewski M.E.* Methodological issues related to the estimation of financial distress prediction models // *Journal of Accounting research*. 1984. pp. 59–82.