

DOI: <https://doi.org/10.17323/j.jcfr.2073-0438.17.1.2023.64-77>

JEL classification: G33, G21, C51



Модель прогнозирования дефолта для компаний сферы услуг на развивающемся рынке

Владислав Афанасьев

Преподаватель, Санкт-Петербургская школа экономики и менеджмента, Департамент финансов, Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербург, Россия
vvafanasev@hse.ru, [ORCID](#)

Аннотация

Автор протестировал гипотезу о том, что прогнозирование дефолтов с помощью финансовых данных может быть неприменимо для организаций сферы услуг из России, проанализировав различия в точности моделей предсказания дефолта, построенных на основе исключительно финансовых данных, для организаций сферы услуг из России и развитых европейских стран.

В качестве инструментов моделирования использовались методы машинного обучения: логистическая регрессия, random forest (случайный лес), k-nearest neighbors (k-ближайших соседей) на выборке из 404 российских фирм и 304 фирм из развитых европейских стран.

Результаты моделирования показали, что ошибка прогнозирования значительно выше для российских организаций относительно организаций из развитых европейских стран. Таким образом, использование исключительно финансовых коэффициентов для прогнозирования банкротства для организаций сферы услуг в России представляется недостаточным.

Результаты исследования могут быть использованы кредитными организациями, организациями, осуществляющими профессиональную оценку кредитного риска, и прочими участниками рынка, заинтересованными в оценке финансового состояния контрагентов.

Ключевые слова: сфера услуг, прогнозирование дефолта, кредитный риск, алгоритмы машинного обучения

Цитирование: Afanasev V. (2023) Default Prediction Model for Emerging Capital Market Service Companies. *Journal of Corporate Finance Research*. 17(1): 64-77. <https://doi.org/10.17323/j.jcfr.2073-0438.17.1.2023.64-77>

Введение

Традиционный подход к прогнозированию дефолта подразумевает применение финансовых коэффициентов в качестве детерминант дефолтов. С 1960-х гг. многочисленные исследования показывали, что с помощью финансовых коэффициентов можно прогнозировать дефолт с высокой точностью, начиная со знаменитого труда Э. Альтмана [1] и заканчивая несколькими недавними работами как иностранных [2], так и российских исследователей [3; 4].

На протяжении этих 60 лет прогнозирование дефолта с применением финансовых коэффициентов развивалось наряду со статистическими методами, лежащими в его основе. Такие простые алгоритмы линейной классификации, как множественный дискриминантный анализ [1], логистическая регрессия [5–8] или пробит-регрессия [9; 10], сегодня заменяют нелинейные алгоритмы машинного обучения [11–17].

Кроме того, расширился набор финансовых коэффициентов, используемых в качестве факторов для прогнозирования дефолта. Исследователи включают нетривиальные факторы для прогнозирования, такие как темпы роста дохода [18] или стандартное отклонение доходности акций [19]. Некоторые исследователи также доказывают, что включение нефинансовых переменных может повысить точность прогнозирования [20–26]. Тем не менее работ, посвященных нефинансовым факторам для прогнозирования дефолта в целом и касающихся российских фирм в частности, очень мало. Одним из возможных объяснений этому может быть высокая предсказательная способность традиционных моделей прогнозирования дефолта (на основе финансовых коэффициентов).

В то же время в случае развивающихся экономик использование только финансовых коэффициентов для прогнозирования дефолта кажется неэффективным. Представляется, что в России финансовая отчетность организаций сферы услуг не всегда отражает реальное состояние бизнеса. Во-первых, некоторые операции могут не раскрываться либо могут иметь место фальсификации. Во-вторых, бизнес может быть раздроблен на несколько юридических лиц, и руководство имеет возможность распределять прибыль, расходы, долговую нагрузку и капитал между юридическими лицами по собственному усмотрению. Данные факторы могут привести к необъективности финансовой отчетности и, следовательно, ее непригодности для прогнозирования дефолта. Таким образом, точность прогнозирования может оказаться низкой.

В статье сравнивается точность прогнозирования алгоритмов классификации логистической регрессии, k -ближайших соседей и случайного леса, обученных на выборке российских организаций сферы услуг, а также на выборке организаций сферы услуг из развитых европейских стран. Алгоритмы были обучены на финансовых коэффициентах фирм из сферы услуг, попавших в ситуацию дефолта, которые были представлены в отчетности за год, предшествующий году дефолта, и на финансовых коэффициентах фирм, не попадавших в ситуацию дефолта. Фирмы из развитых европейских стран использовались в качестве контрольной группы. Ожидалось, что из-за вероятной необъективности финансовой отчетности, вызванной теневыми операциями и дроблением бизнеса, для российских организаций сферы услуг точность прогнозирования будет ниже, чем для фирм из развитых европейских стран, которые, по всей видимости, не обладают указанными характеристиками. Следовательно, целью настоящего исследования является оценка

точности прогнозирования вероятности дефолта для российских фирм из сферы услуг, если в качестве факторов для прогнозирования используются только финансовые данные, и сравнение ее с оценкой для фирм с развитых европейских рынков. Результат такого анализа – оценка возможности использования финансовых коэффициентов для прогнозирования дефолта российских организаций сферы услуг.

В следующем разделе представлен обзор литературы по прогнозированию дефолта, затем объясняется, почему для данного анализа выбрана сфера услуг. В разделе «Теоретическая база» более подробно раскрывается, почему финансовые коэффициенты не кажутся надежными факторами для прогнозирования дефолта российских организаций сферы услуг, а в разделе «Методы исследования» описываются используемые данные и применяемые алгоритмы. В завершение представляются результаты моделирования.

Обзор литературы

Модели прогнозирования дефолта для фирм разрабатываются уже более 50 лет, начиная с первой модели оценки кредитного риска, разработанной У. Бивером [27]. С целью увеличения точности прогнозирования исследователи совершенствовали модели в двух основных направлениях: методы и объясняющие переменные.

Во-первых, исследователи начали использовать более современные методы моделирования, начиная с Э. Альтмана [1], применившего для своей модели множественный дискриминантный анализ. Дж. Олсен [5], возможно, первым использовал логистическую регрессию для создания модели оценки вероятности дефолта. Логистическая регрессия (Logit) и подобный ей алгоритм – пробит-регрессия – широко использовались исследователями XX в. и все еще применяются в наши дни [4; 8; 10; 17] в большей мере благодаря их простоте (линейные алгоритмы). Однако в настоящее время основой для исследований прогнозирования дефолта являются алгоритмы машинного обучения. Существует множество различных алгоритмов машинного обучения, применяемых для целей прогнозирования дефолта, однако исходя из проанализированной литературы самыми популярными являются искусственные нейронные сети [14; 28] и метод опорных векторов [18; 25].

Одним из вкладов статьи является внедрение алгоритма случайного леса в качестве метода, лежащего в основе прогнозирования дефолта. Данный алгоритм нечасто используется в исследованиях прогнозирования дефолта несмотря на его высокую эффективность, показанную предыдущими исследователями [29; 30].

Отдельной областью исследования в рамках прогнозирования дефолта является моделирование кредитного рейтинга [31; 32]. Модели основаны на финансовых данных для корпораций и макроэкономических данных и применимы в основном к публичным компаниям благодаря существенному влиянию рыночной капитализации на кредитный рейтинг.

Вторым направлением развития прогнозирования дефолта является расширение набора объясняющих переменных – выход за рамки использования лишь финансовых данных. Данное направление развития относительно новое, «новый тренд в данной сфере» [21]. Согласно Э. Альтману [23] до 2010 г. исследования в данной области для малых и средних предприятий не проводились вовсе.

В отношении прогнозирования дефолта не существует ограничений по использованию любых данных, доступных для

анализируемых фирм, и исследователи начинают применять эти доступные данные. Примерами таких переменных являются показатели текста, опубликованного в новостях или в раскрываемой информации фирмы (например, уровень эмоциональности или использование определенных слов) [25; 26], показатели, связанные с судебными исками [21], показатели корпоративного управления [20], показатели корпоративной социальной ответственности [22], показатели аудиторских заключений (например, уровень эмоциональности, количество комментариев аудитора и т.д.) [24].

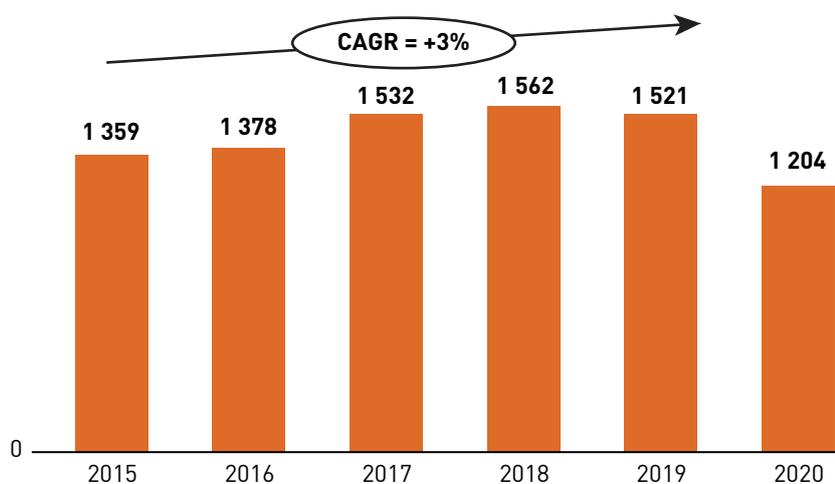
На основании проанализированной литературы представляется, что применение нефинансовых данных пока что не заменяет традиционный подход (основанный на финансовых данных), особенно в части исследований по России.

Этот факт можно объяснить высокой точностью прогнозирования дефолта на основании финансовых данных. Однако, как показано ниже, подход, основанный на финансовых данных, может быть низкоэффективным для российских фирм, и в этом случае использование нефинансовых данных может оказаться удачным решением.

Дефолты в сфере услуг России

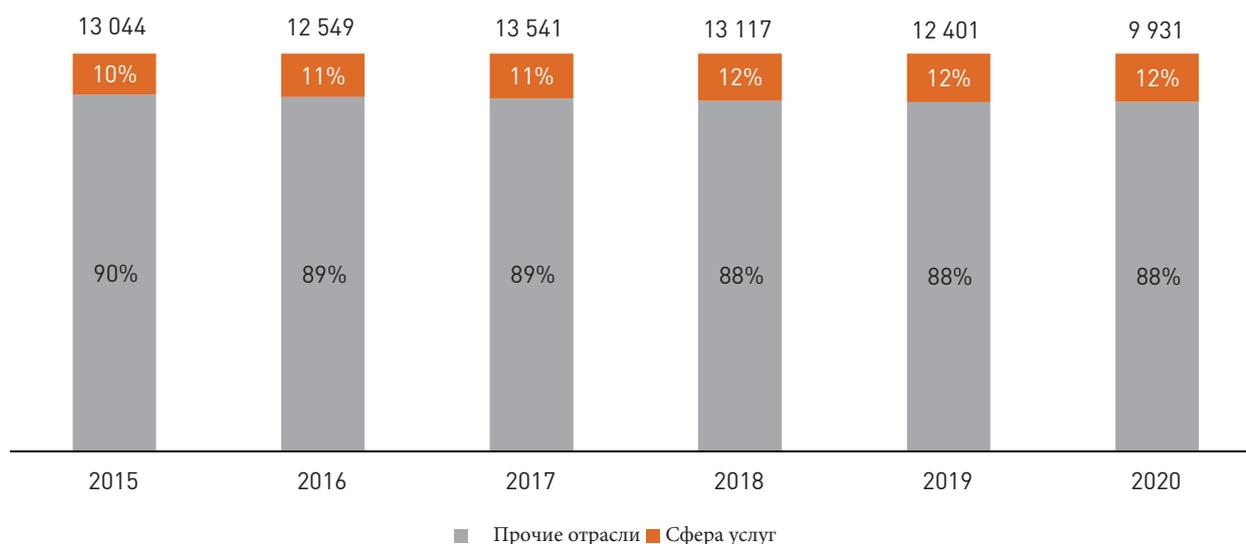
Для данного исследования выбрана сфера услуг России, потому что именно в ней особенно необходимо точное прогнозирование дефолта. Во-первых, в 2015–2020 гг. общее количество банкротств в России снизилось, в то время как доля банкротств в сфере услуг в общем количестве случаев увеличилась (Рисунки 1 и 2).

Рисунок 1. Банкротства в сфере услуг России, 2015–2020 гг., количество случаев



Источник: Федресурс. URL: <https://fedresurs.ru/news/5343e0f4-bf32-4fef-b293-cc752e65f491> (дата обращения: 15.06.2021).

Рисунок 2. Структура банкротств по секторам, 2015–2020 гг., %, количество случаев



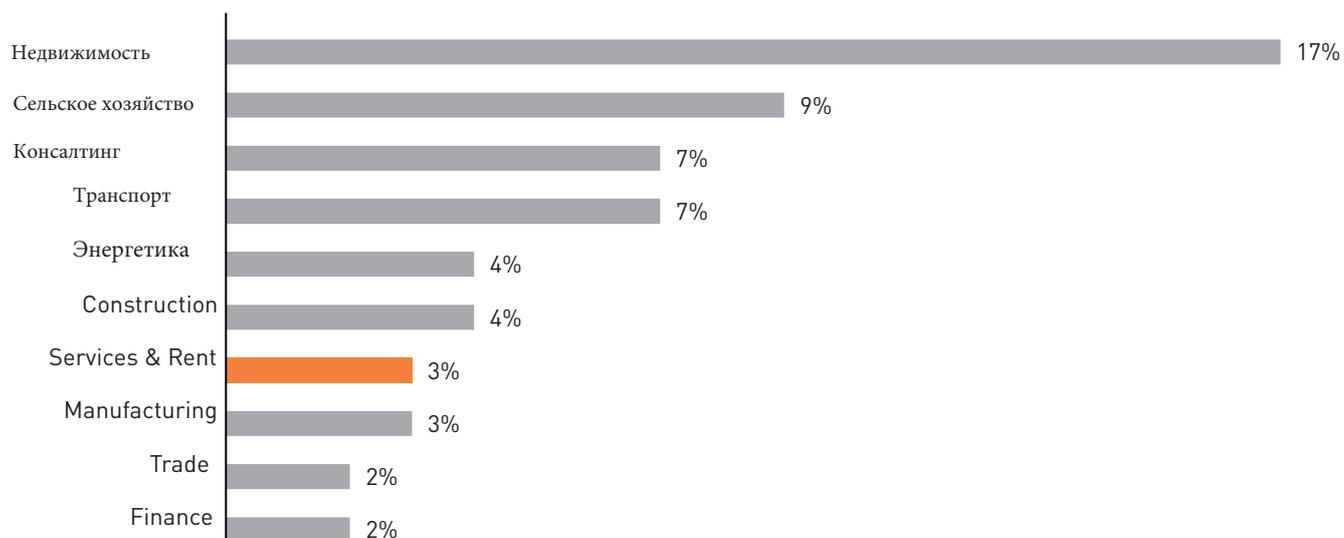
Источник: Федресурс. URL: <https://fedresurs.ru/news/5343e0f4-bf32-4fef-b293-cc752e65f491> (дата обращения: 15.06.2021).

Далее 2020 г. не принимается в расчет из-за моратория на банкротство в России, установленного по причине пандемии COVID-19.

Во-вторых, доля задолженности, выплачиваемая кредиторам в рамках процедур банкротства, в сфере услуг является

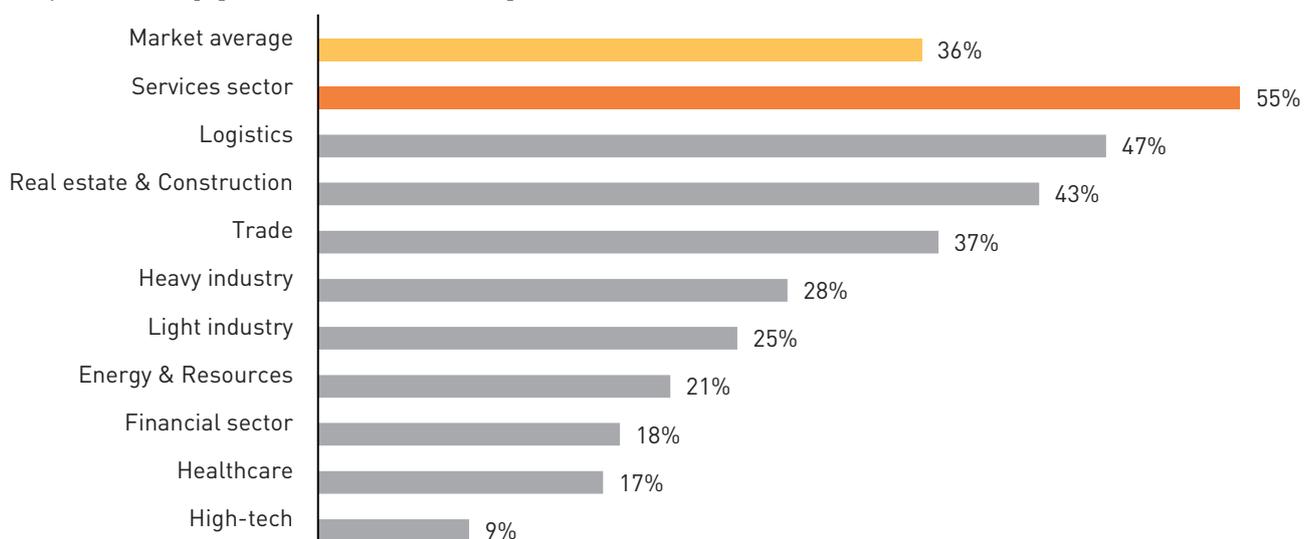
одной из самых низких среди всех отраслей. В 2019 г. этот показатель составлял лишь 3.4% (ниже среднего показателя 4.7%) (Рисунок 3). Это означает, что в случае банкротства ожидаемая сумма возврата долга на каждые занятые 100 руб. составляет только 3.4 руб.

Рисунок 3. Доля задолженности, выплачиваемая в случае банкротства в первой десятке отраслей по количеству случаев банкротства, 2019 г., % от общей задолженности



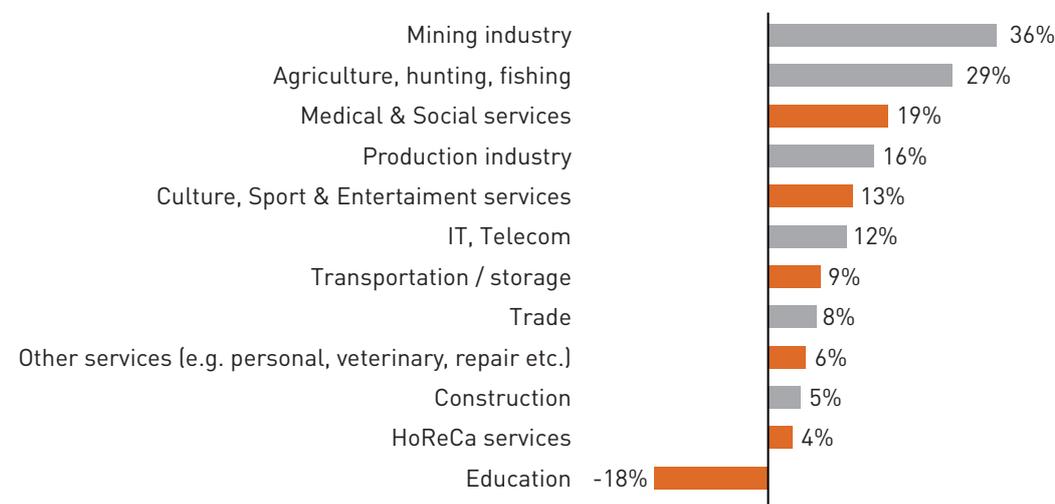
Источник: Федресурс. URL: <https://fedresurs.ru/news/5343e0f4-bf32-4fef-b293-cc752e65f491> (дата обращения: 15.06.2021).

Рисунок 4. Доля фирм с задолженностью по отраслям, 2020 г., %



Источник: Центр стратегических разработок.

Рисунок 5. Доходность продаж по отраслям, 2021 г., %



Источник: Росстат. URL: <https://www.fedstat.ru/indicator/58261> (дата обращения: 17.03.2023).

Фирмы в сфере услуг также более склонны к наличию задолженности, чем фирмы любой другой отрасли. Согласно исследованию, проведенному Центром стратегических разработок¹, 55% фирм из сферы услуг имеют задолженность, в то время как средний показатель по рынку составляет 36% (Рисунок 4). Это свидетельствует о более высоком кредитном риске сферы услуг по сравнению с другими отраслями.

Растущее количество банкротств и низкий показатель возврата долга в случае дефолта объясняются спецификой сферы услуг. Эта сфера в основном состоит из предприятий, работающих по схеме бизнес-клиент (B2C), что подразумевает высокий уровень конкуренции, а следовательно, низкую рентабельность. По сравнению с другими отраслями, такими как производство, сельское хозяйство или добыча полезных ископаемых, средняя рентабельность сферы услуг ниже, или даже отрицательная (Рисунок 5). Это утверждение менее применимо к медицинским услугам, но весьма актуально для таких крупных рынков, как отельно-ресторанные услуги и личные услуги (включающие повседневные услуги, т.е. ремонт, услуги парикмахера и т.д.).

И последним, но не менее важным аргументом в пользу необходимости уделить внимание такому специфическому сектору экономики, как сфера услуг, является пробел в исследованиях, связанных с моделированием кредитного риска для конкретных отраслей экономики [31]. Настоящее исследование направлено на заполнение этого пробела для сферы услуг.

Теоретическая база

Существуют две основных причины для утверждения, что финансовая отчетность российских организаций сферы услуг не отражает реальное состояние бизнеса.

Искусственное дробление бизнеса приводит к необъективности финансовых коэффициентов

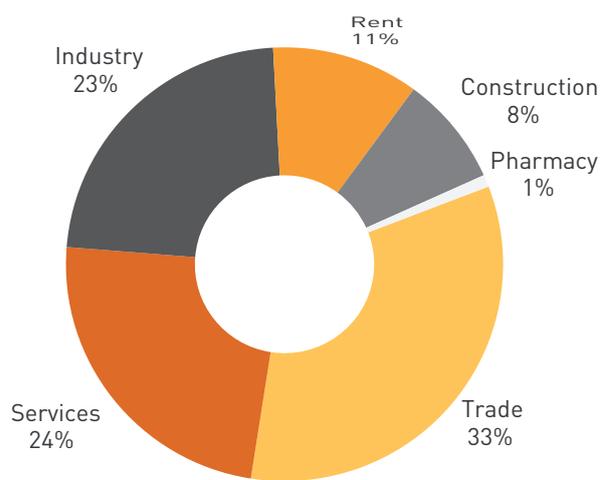
Если фирма разделена на несколько юридических лиц, это означает, что необходимо иметь консолидированную финансовую отчетность, чтобы судить о состоянии всего бизнеса. С одной стороны, не всегда возможно получить отчетность по группе юридических лиц, с другой стороны, некоторые части группы могут быть представлены в форме ИП или юридических лиц на упрощенной системе налогообложения, которые не обязаны предоставлять полную отчетность. Поэтому обычно для анализа фирмы приходится использовать данные по одному юридическому лицу, и в этом случае существует вероятность, что данные могут быть необъективными.

Проблема дробления бизнеса весьма актуальна для российского рынка. Малые юридические лица имеют возможность снизить налоговую нагрузку при помощи упрощенной системы налогообложения. По этой причине владельцы зачастую разбивают свой бизнес на несколько

малых юридических лиц, таким образом, снижая налоговую нагрузку [33]. Актуальность проблемы дробления бизнеса подтверждается активными упреждающими мерами правительства. С 2017 г. Федеральная налоговая служба и Следственный комитет России активно проводят разоблачительную политику, включающую постоянную разработку и обновление критериев дробления. [34].

Проблема дробления бизнеса актуальна для всех секторов экономики России, включая и сферу услуг. Согласно обзору, проведенному TaxCoach², 24% исковых заявлений по дроблению бизнеса в 2020 г. были связаны с организациями сферы услуг (Рисунок 6).

Рисунок 6. Исковые заявления по дроблению бизнеса по отраслям, 2020 г., %



Источник: TaxCoach.

Теневые операции приводят к необъективности в финансовых коэффициентах

Во времена СССР в России не было законных частных фирм, которые могли бы оказывать услуги населению. В то же время государственные организации не предоставляли повседневные услуги. Таким образом, необходимые услуги оказывали частные лица, включая ремонт, транспортировку, репетиторство и т.д. Это был незаконный, но единственный способ получить требуемые услуги. Длительное существование в условиях теневой экономики повлияло на представление граждан России о деловой культуре [35].

Согласно обзору, проведенному Форумом по изучению стран Восточной Европы и развивающихся рынков (FREE Network)³, объем теневой экономики в России составляет почти 45% от ВВП. Два основных вида теневых операций – это занижение прибыли в отчетности и «зарплаты в конвертах» (как отмечает Татьяна Голикова⁴, заместитель

¹ Папченкова Е. (24 декабря 2020 г.). *Бизнес-климат России. Итоги 2020 года. Банкротство.* (Business climate in Russia. 2020 year summary. Bankruptcy). URL: <https://www.youtube.com/watch?v=cF98nMjWSbs> (дата обращения: 29.05.2021).

² TaxCoach. (2021). *Остаться в живых. Гид по обвинениям в искусственном дроблении бизнеса на основе анализа 450 арбитражных дел* (Stay alive. A guide for legal claims for artificial business separation, based on 450 legal proceedings). URL: https://www.taxcoach.ru/taxbook/droblenie_biznesa (дата обращения: 01.06.2021).

³ Putniņš T., Sauka A. (2020). *The Shadow Economy in Russia: New Estimates and Comparisons with Nearby Countries.* (Путниньш Т. и Саука А. (2020). *Теневая экономика в России: новые оценки и сравнение с соседними странами.*) URL: <https://freepolicybriefs.org/2020/03/16/shadow-economy-russia/> (дата обращения: 03.06.2021).

⁴ Голикова Т. (июнь 2019 г.). *Интервью с Татьяной Голиковой для «ИЗВЕСТИЙ».* URL: <https://iz.ru/886870/elena-lorija-elena-likhomanova/deistvie-sotckontrakta-ne-dolzno-ogranichivatsia-mesiatcem-ili-godom> (дата обращения: 03.06.2021).

председателя Правительства Российской Федерации, около 15 млн граждан России получают зарплату «в конверте»). Согласно данным Российского многолетнего мониторинга (ВШЭ) 2020 г.⁵, 16% граждан России признаются, что их труд оплачивается неофициально, а 51% из них получают всю зарплату «в конверте».

Если в деятельности организации присутствуют теневые операции, официальная финансовая отчетность юридического лица может быть необъективной: доходы могут быть занижены, затраты – преувеличены и т. д.

Дополнительным косвенным доказательством необъективности финансовой отчетности российских фирм являются недостаточно строгие стандарты аудита и бухгалтерского учета. Согласно данным Глобального индекса конкурентоспособности Всемирного банка⁶ Российская Федерация занимает 100-е место из 137 стран по строгости стандартов аудита и бухгалтерского учета (4 из 7 баллов за вопрос «Насколько строгие стандарты финансового аудита и бухгалтерской отчетности в вашей стране?» (1 – очень мягкие; 7 – очень строгие)).

Таким образом, учитывая приведенные тезисы, финансовые коэффициенты российских организаций сферы услуг могут быть необъективными, следовательно, одной только финансовой информации недостаточно для оценки кредитного риска в случае с российскими организациями сферы услуг.

Методы исследования

Описание данных

Поскольку не существует единого определения сферы услуг, необходимо указать отрасли, которые к ней относятся. Согласно данным *Большой российской энциклопедии*⁷ в сферу услуг включаются услуги в области культуры, образования и домашнего хозяйства. Федеральная служба государственной статистики Российской Федерации⁸ к сфере услуг относит услуги почтовой связи, телекоммуникационные, коммунальные, медицинские, туристические услуги, услуги системы образования и юридические. В настоящем исследовании рассмотрены фирмы из следующих отраслей, которые определены являются частью сферы услуг:

- туризм, гостиничный бизнес и пассажирские перевозки;
- общественное питание и доставка продуктов;
- образование;
- медицинские и социальные услуги;
- культура, спорт и развлечения;
- прочие услуги (персональные услуги, ветеринарные услуги).

При отборе российских фирм для анализа использовалась классификация ОКВЭД-2, при отборе европейских фирм – Классификация видов экономической деятельности ЕС. Количество фирм по категориям услуг приведено в Приложении 1.

Было подготовлено два набора данных. Первый набор данных содержит информацию по российским организациям сферы

услуг, столкнувшихся с ситуацией дефолта с 2017 по 2020 г. Год, в котором кредитор направил уведомление о намерении кредитора обратиться в суд с заявлением о банкротстве, использовался как год определения банкротства. Данные были собраны из базы данных СПАРК-Интефакс⁹. Набор данных состоит из 202 фирм, нарушивших платежные обязательства. Для каждой фирмы в «дефолтной» выборке была подобрана контрольная «недефолтная» фирма. В качестве критерия соответствия была выбрана стоимость активов фирмы. Исследователи широко используют этот критерий соответствия [8].

Зависимая переменная является дамми переменной: 1 – есть факт дефолта, 0 – нет факта дефолта. Независимые переменные – финансовые коэффициенты фирм (рассчитанные за год, предшествующий году дефолта для «дефолтных» фирм и за тот же год для «недефолтных» фирм).

Обычно финансовые коэффициенты, используемые исследователями для создания моделей прогнозирования дефолта, включают:

- коэффициенты оборачиваемости;
- коэффициенты рентабельности;
- коэффициенты ликвидности;
- коэффициенты активов, концентрации собственного капитала или структуры задолженности, коэффициенты покрытия долга [36].

Оказалось невозможным включить коэффициенты покрытия долга, потому что информация о сумме выплат по процентам для большинства российских фирм в наборе данных отсутствует. В Таблице 1 приведен окончательный список используемых независимых переменных.

Таблица 1. Список независимых переменных

	Оборачиваемость чистых активов
Коэффициенты оборачиваемости	Оборачиваемость запасов
	Период погашения дебиторской задолженности
	Период погашения кредиторской задолженности
Коэффициенты рентабельности	Рентабельность по чистой прибыли
	Рентабельность активов
Коэффициенты ликвидности	Коэффициент текущей ликвидности
	Коэффициент быстрой ликвидности
Коэффициенты активов, концентрации собственного капитала или структуры задолженности	Акционерный капитал / Активы

Источник: составлено автором.

⁵ Российский многолетний мониторинг (ВШЭ) 2020 г. URL: <https://www.hse.ru/rims/spss> (дата обращения: 19.02.2022).

⁶ Рейтинги конкурентоспособности. (2017). Глобальный индекс конкурентоспособности 2017–2018. URL: <http://wef.ch/2wcVUt8> (дата обращения: 21.09.2021).

⁷ Большая российская энциклопедия. URL: <https://bigenc.ru/economics/text/3546082> (дата обращения: 10.06.2021).

⁸ Росстат. URL: <https://rosstat.gov.ru> (дата обращения: 10.06.2021).

⁹ СПАРК Интерфакс. URL: <https://spark-interfax.ru> (дата обращения: 12.06.2021).

Таблица 2. Описательная статистика переменных по двум базам данных

Переменная	Набл.	Среднее	Ст. откл.	Мин.	Макс.
Европейские данные, дефолты					
Рентабельность по чистой прибыли, %	145	-11.382	21.112	-97.23	48.447
Рентабельность активов по чистой прибыли, %	141	-13.629	20.021	-89.48	28.26
Оборачиваемость чистых активов, X	112	9.984	22.574	.042	183.346
Оборачиваемость запасов, X	104	90.301	138.537	2.566	924.534
Период погашения дебиторской задолженности, дни	148	51.385	85.619	0	688.013
Период погашения кредиторской задолженности, дни	149	61.669	92.285	0	654.728
Коэффициент текущей ликвидности, X	151	.833	1.069	.005	12.263
Коэффициент быстрой ликвидности, X	148	.755	1.078	.005	12.263
Акционерный капитал / Общие активы, X	152	-.117	.921	-9.207	.899
Европейские данные, отсутствие дефолта					
Рентабельность по чистой прибыли, %	152	4.194	15.772	-83.884	94.162
Рентабельность активов по чистой прибыли, %	152	5.459	9.087	-20.883	35.116
Оборачиваемость чистых активов, X	152	5.325	8.147	.06	70.962
Оборачиваемость запасов, X	93	122.626	146.115	1.87	845.975
Период погашения дебиторской задолженности, дни	152	29.849	36.799	0	213.023
Период погашения кредиторской задолженности, дни	152	18.533	19.88	0	108.371
Коэффициент текущей ликвидности, X	152	2.228	7.108	.014	80.151
Коэффициент быстрой ликвидности, X	152	2.104	7.1	.014	80.151
Акционерный капитал / Активы, X	152	.36	.263	-.607	.987
Российские данные, дефолты					
Рентабельность по чистой прибыли, %	201	-937.1	8705.9	-102 815.1	100
Рентабельность активов по чистой прибыли, %	202	-462.1	5971	-84 837.1	1907.9
Оборачиваемость чистых активов, X	190	7.961	55.255	-352.55	400.299
Оборачиваемость запасов, X	176	372.208	1478.422	0	14 753.5
Период погашения дебиторской задолженности, дни	199	6600.915	50 664.473	1	579 366
Период погашения кредиторской задолженности, дни	195	26 474.502	330 690.63	2.57	4 618 755.6
Коэффициент текущей ликвидности, X	200	5.654	19.806	.005	180.6
Коэффициент быстрой ликвидности, X	200	4.635	15.212	.003	122.56
Акционерный капитал / Активы, X	200	-5.684	53.369	-750.114	1
Российские данные, отсутствие дефолта					
Рентабельность по прибыли, %	202	.051	.302	-2.902	100
Рентабельность активов по чистой прибыли, %	200	-.182	10.05	-132.3	4765.2
Оборачиваемость чистых активов, X	184	304.016	1554.744	-620.513	16 101.7

Переменная	Набл.	Среднее	Ст. откл.	Мин.	Макс.
Оборачиваемость запасов, X	172	1042.513	9181.763	.2	117 718
Период погашения дебиторской задолженности, дни	190	182.779	1472.473	1	20 240
Период погашения кредиторской задолженности, дни	194	1436.098	18 422.058	.42	256 678.65
Коэффициент текущей ликвидности, X	202	4.169	11.505	.012	140.883
Коэффициент быстрой ликвидности, X	200	3.361	11.143	.007	140.883
Акционерный капитал / Активы, X	202	-1.259	9.951	-81.2	.993

Источник: составлено автором.

Второй набор данных – это контрольная группа. Он содержит ту же информацию, но по организациям сферы услуг из развитых стран ЕС (152 фирмы, нарушившие платежные обязательства, и 152 финансово устойчивые фирмы). Дата начала процедуры банкротства использовалась для определения года дефолта. Данные были взяты из базы данных Amadeus¹⁰.

В качестве контрольной группы были выбраны фирмы из развитых стран Европейского союза (ЕС), потому что для них существенно менее актуальны проблемы теневых операций и дробления бизнеса. В то время как объем теневого рынка в европейских развивающихся странах и странах с переходной экономикой оценивается примерно в 27%, тот же коэффициент для ЕС в два раза ниже (лишь около 14%)¹¹. Страны с самым низким коэффициентом теневой экономики – Австрия, Люксембург, Великобритания, Нидерланды, Франция, Ирландия, Исландия, Германия, Дания, Швеция, Словакия, Финляндия, Испания, Норвегия¹². Контрольный набор данных сформирован из фирм этих стран.

Что касается дробления бизнеса, по ЕС статистика отсутствует, но можно утверждать, что данная проблема менее актуальна для европейского рынка. Учитывая, что дробление бизнеса – это инструмент уменьшения налоговой нагрузки, отношение бизнес-сообщества к налоговым ставкам может использоваться в качестве прокси для уровня дробления. Согласно данным Всемирного банка¹³ 22.6% российских фирм рассматривают налоговые ставки как самую серьезную помеху для бизнеса. Тот же показатель для Австрии составляет лишь 20.6%, для Дании – 6.4, Люксембурга – 5.7, Нидерландов – 7.4, Ирландии – 13.6, Швеции – 13.4, Словакии – 17.7, Финляндии – 9.5%. Данные по другим европейским странам отсутствуют, но можно предположить, что поскольку они находятся на более высоком «уровне развития», проблема дробления бизнеса для

них менее актуальна. ВВП на душу населения используется как прокси для «уровня развитости» стран. ВВП на душу населения в остальных странах, по которым нет данных касательно отношения бизнеса к налогам, намного выше, чем в России¹⁴.

Описательная статистика переменных по двум базам данных приведена в Таблице 2. Можно заметить, что данные российской финансовой отчетности обладают своей спецификой, например, чрезвычайно низкий коэффициент рентабельности или исключительно длинные период погашения дебиторской задолженности и срок кредитования по дефолтам. Данная специфика также может быть показателем смещенности финансовой отчетности. Принято решение не рассматривать фирмы с крайними значениями как выбросы, потому что эти крайние значения взяты из реально существующей финансовой отчетности (отчетность таких фирм проверена вручную).

Алгоритмы машинного обучения

Для обучения данных были использованы три алгоритма машинного обучения: логистическая регрессия, k-ближайших соседей и случайный лес. Логистическая регрессия является линейным алгоритмом классификации, часто используемым для прогнозирования дефолта [5–8]. Одним из преимуществ логистической регрессии является способность интерпретировать вклад в прогноз каждой независимой переменной. Алгоритм k-ближайших соседей был выбран как, возможно, самый простой алгоритм машинного обучения, часто используемый в исследованиях, посвященных прогнозированию дефолта [37]. Классификатор на основе случайного леса выбран как один из самых эффективных алгоритмов, применяемых для прогнозирования дефолта и скоринга, что подтверждают предыдущие исследования [29; 30].

¹⁰ База данных Amadeus. (2021). URL: <https://amadeus.bvdinfo.com> (дата обращения: 15.07.2021).

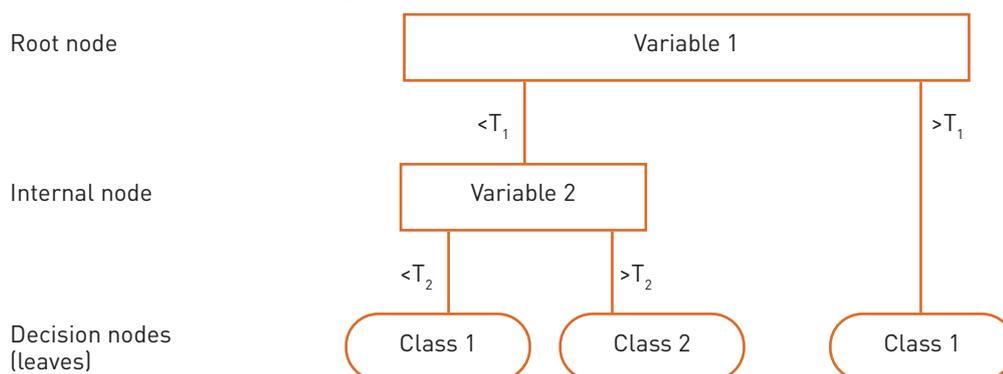
¹¹ Boumans D., Schneider F. (2019). Ifo World Economic Survey (No. 18; p. 2)]. Leibniz Institute for Economic Research at the University of Munich. URL: https://www.ifo.de/DocDL/WES_4_19_0.pdf (дата обращения: 08.08.2021).

¹² Kelmanson B., Kirabaeva K., Medina L., Mircheva B., Weiss J. (2019). Explaining the Shadow Economy in Europe: Size, Causes and Policy Options. Келмансон Б., Кирабаева К., Медина Л., Мирчева Б. и Вейсс Д. (2019). Объясняя теневую экономику в Европе: размер, причины и варианты политики. Рабочий документ МВФ. Международный валютный фонд. URL: <https://www.google.com/url?sa=t&trct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjHiM3o4dvzAhVFpIsKHaXvDLoQFnoECAgQAQ&url=https%3A%2F%2Fwww.imf.org%2F~%2Fmedia%2FFiles%2FPublications%2FWP%2F2019%2Fwpia2019278-print-pdf&usq=AOvVaw3112V7M9BqTYQpO-Xh1z> (дата обращения: 10.08.2021).

¹³ Обзоры предприятий (Данные Всемирного банка). URL: <https://www.enterprisesurveys.org/en/custom-query> (дата обращения: 15.09.2021).

¹⁴ ВВП (Данные Всемирного банка). URL: <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD> (дата обращения: 15.09.2021).

Рисунок 7. Пример простого дерева решений (CART)



Источник: составлено автором.

Логистическая регрессия – это алгоритм, сходный с обыкновенной линейной регрессией. Разница заключается в том, что прогнозируемая зависимая переменная может меняться лишь в диапазоне от 0 до 1, в то время как в случае с обыкновенной линейной регрессией она может принимать любое значение. Для прогнозов используется логистическая функция (логистическая кривая):

$$P(x) = \frac{e^{B_0 + B_1 X_1 + \dots + B_n X_n}}{1 + e^{B_0 + B_1 X_1 + \dots + B_n X_n}}$$

$P(x)$ в случае данного исследования является оценкой вероятности дефолта, а $B_0 - B_n$ – линейными коэффициентами независимых переменных (финансовых коэффициентов). Чтобы трансформировать регрессию в алгоритм классификации, устанавливается предельное значение вероятности (50% в данном случае) для присваивания класса наблюдению. Наблюдения классифицируются как «дефолтные», если оценка вероятности по модели выше 50%.

Логистическая регрессия оценивается при помощи метода максимального правдоподобия. Коэффициенты регрессии определяются так, чтобы максимизировать функцию правдоподобия:

$$LF = \prod \left(P(x_i)^{y_i} (1 - P(x_i))^{1-y_i} \right), i \in (1; n),$$

которая является произведением всех рассчитанных оценок вероятностей принадлежности наблюдения к истинному классу для наблюдения [38].

Регуляризация Лассо L1 используется для ограничения количества переменных. Сумма абсолютных значений коэффициентов прибавляется к минимизируемой функции.

Классификатор k-ближайших соседей является одним из простейших алгоритмов классификации. Классификация основана на классах нескольких (k) наиболее сходных фирм из набора данных для обучения. Наблюдение классифицируется на основании голосования большинством голосов. Процедура классификации состоит из следующих трех этапов.

Выбор количества «соседей». Количество «соседей» не должно быть очень низким (может снизить точность) или очень высоким (большинство наблюдений в тестовом наборе будут отнесены к одному классу, который имеет наи-

большее количество представителей в наборе данных для обучения). В качестве k использован квадратный корень из количества наблюдений согласно признанному исследователями подходу [39].

Оценка расстояний между данными для обучения и тестирования.

Для определения ближайших «соседей», применено евклидово расстояние:

$$\sqrt{\sum \left(\begin{array}{l} \text{Значение переменной } i \text{ для наблюдения в тестовом наборе} - \\ \text{Значение переменной } i \text{ для наблюдения в наборе для обучения} \end{array} \right)^2}$$

Классификация тестового наблюдения на основании голосования большинством голосов, другими словами, отнесение к классу, исходя из самого популярного класса среди «соседей»¹⁵.

Из-за использования евклидова расстояния перед моделированием данные необходимо нормализовать.

Классификатор случайного леса является алгоритмом машинного обучения, использующий ансамбль деревьев регрессии и классификации (CART). Пример простого CART представлен на Рисунке 7.

При обучении дерева данные для обучения разбивают на две подвыборки в каждом узле. Разбивку производят на основании значения определенной переменной. Индекс Джини используется для выбора переменных (переменная 1, переменная 2 на Рисунке 7) и порогового значения для разбивки (T_1 и T_2 на Рисунке 7). Основополагающая идея заключается в минимизации данного индекса. Индекс Джини отражает обратную точность разбивки:

$$\begin{aligned} \text{Индекс Джини} &= \frac{\text{Количество наблюдений в } L}{\text{Общее количество наблюдений}} \cdot \\ &\cdot \left(1 - \sum \left(\frac{\text{Количество наблюдений класса } i \text{ в } L}{\text{Общее количество наблюдений в } L} \right) \right) + \\ &+ \frac{\text{Количество наблюдений в } R}{\text{Общее количество наблюдений}} \cdot \\ &\cdot \left(1 - \sum \left(\frac{\text{Количество наблюдений класса } i \text{ в } R}{\text{Общее количество наблюдений в } R} \right) \right). \end{aligned}$$

¹⁵ Laszlo K. (2008). *K Nearest Neighbors algorithm (kNN)*. *Special Course in Computer and Information Science*. (Лазло К. (2008). Алгоритм k-ближайших соседей. *Специальный курс вычислительной техники и информатики*). URL: <http://www.lkozma.net/knn2.pdf> (дата обращения: 15.08.2021).

L и R – подвыборка 1 и подвыборка 2 (справа и слева); i – класс (1 – дефолт, 0 – нет дефолта) [40].

«Лес» означает сочетание простых деревьев решений, «случайный» – тот факт, что каждое дерево обучено на случайно выбранной подвыборке из обучающей выборки, а переменные для разбивки выбраны произвольным образом. Подвыборки формируются при помощи бутстрапа. Идея, лежащая в основе данного метода, заключается в том, что повторные выборки берут из начальной обучающей выборки. По каждому дереву переменные (переменная 1 и переменная 2 на Рисунке 7) выбирают из случайного списка k -переменных, взятых из общего списка. Благодаря этому деревья отличаются друг от друга.

Таблица 3. Доли отсутствующих данных в наборах данных (%)

	С произошедшим дефолтом		Без дефолта	
	Данные по России	Данные по Европе	Данные по России	Данные по Европе
Рентабельность по чистой прибыли, %	6	26	9	0
Рентабельность активов по чистой прибыли, %	13	32	15	39
Оборачиваемость чистых активов, X	1	3	6	0
Оборачиваемость запасов, X	3	2	4	0
Период погашения дебиторской задолженности, дни	0	5	0	0
Период погашения кредиторской задолженности, дни	0	7	1	0
Коэффициент текущей ликвидности, X	1	1	0	0
Коэффициент быстрой ликвидности, X	1	3	1	0
Акционерный капитал / Активы, X	1	0	0	0

Источник: составлено автором.

Полученные наборы данных разделены на обучающие и тестовые выборки. Затем алгоритмы классификации были обучены на обучающих наборах, после чего обученные алгоритмы были применены к тестовым выборкам и была рассчитана точность прогнозирования. Чтобы удостовериться, что результат по конкретной выборке не является выбросом, произошедшим из-за специфического разделения на тестовую и обучающую выборки, было осуществлено 100 произвольных разделений на тестовую и обучающую выборки по каждому набору данных, а затем алгоритмы были обучены на каждой обучающей выборке и была рассчитана точность на каждой соответствующей тестовой выборке.

Основная гипотеза заключается в том, что средняя точность прогнозирования дефолта для российских организаций сферы услуг будет ниже, чем у европейских организаций сферы услуг. Данная гипотеза была протестирована при помощи теста Манна – Уитни.

Результаты

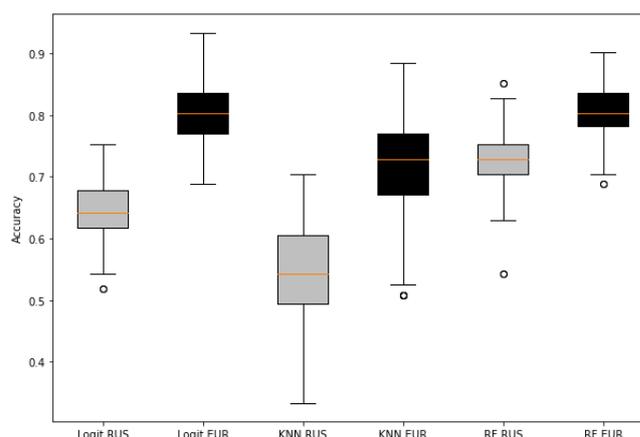
Результаты свидетельствуют о том, что точность прогнозирования для российских фирм существенно ниже. Результаты по трем алгоритмам классификации представлены на Рисунке 8.

Необходимо ограничивать количество деревьев и внутренних узлов в каждом дереве. Принято решение обучить 100 деревьев для каждого набора обучения и установить максимальное количество уровней дерева в размере 2.

Подготовка и моделирование данных

В наборах данных имелись пропущенные значения. Они заменены на средние значения соответствующей переменной. В Таблице 3 приведены доли пропущенных значений по каждой переменной двух наборов данных. Имеются некоторые различия, но представляется, что качество собранных данных для российских и европейских фирм сходно.

Рисунок 8. Результаты классификации по логистическому алгоритму, алгоритму k -ближайших соседей и случайного леса

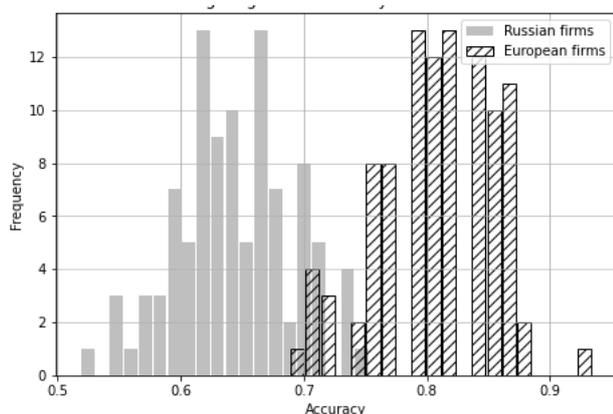


Источник: составлено автором.

¹⁶ Steorts R. (2014). Bagging and Random Forests. (Среортс Р. (2014). Бэггинг и случайный лес).

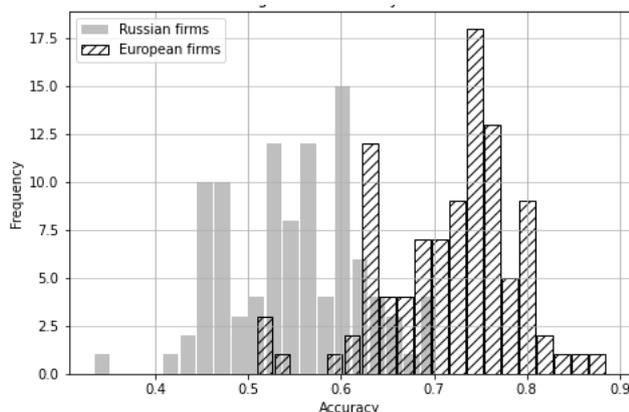
URL: http://www2.stat.duke.edu/~rcs46/lectures_2015/random-forest/slides_lecture15.pdf (дата обращения: 15.08.2021).

Рисунок 9. Распределение точности алгоритма логистической регрессии



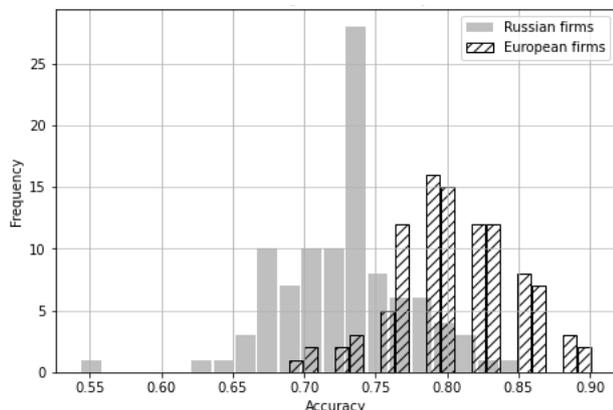
Источник: составлено автором.

Рисунок 10. Распределение точности алгоритма k-ближайших соседей



Источник: составлено автором.

Рисунок 11. Распределение точности алгоритма случайного леса



Источник: составлено автором.

В первую очередь была оценена логистическая регрессия. Средняя точность классификации составляет 64.4% для российских организаций сферы услуг и 80.7% для набора

данных по европейским фирмам. Рисунок 9 показывает распределение точности логистического алгоритма, рассчитанной на произвольно сформированных тестовых наборах для российских и европейских фирм из сферы услуг. Как для российских, так и для европейских фирм распределение визуально близко к нормальному, но результат теста Шапиро – Уилка на нормальность говорит о том, что точность для европейских фирм распределена ненормально (р-значения для наборов российских и европейских фирм составляют 0.386 и 0.04 соответственно). Поэтому был применен непараметрический критерий Манна – Уитни вместо традиционного критерия Стьюдента, чтобы понять, есть ли статистически значимые отличия в средней точности. Р-значение критерия Манна – Уитни близко к нулю ($1.35 \cdot 10^{-33}$), а это означает, что вероятность получения такой статистики критерия очень низка, если средняя точность для российских и европейских фирм одинакова.

Точность алгоритма k-ближайших соседей ниже в обоих случаях: 54.8% для российских фирм и 71.7% – для европейских. Точность классификации для европейских фирм можно посчитать недостаточной, но она все еще значительно выше, чем средняя точность для российских компаний. На Рисунке 10 показано распределение точности алгоритма k-ближайших соседей, рассчитанной на случайно сформированных тестовых наборах для российских и европейских организаций сферы услуг. Распределение точности нормально для российских фирм, но не для европейских (р-значения теста Шапиро – Уилка составляют 0.389 и 0.008 соответственно), поэтому для оценки значимости разницы между показателями средней точности использовался критерий Манна – Уитни. Р-значение критерия Манна – Уитни близко к нулю ($4.40 \cdot 10^{-28}$), а это означает, что вероятность получения такого значения очень низка, если средняя точность для российских и европейских фирм одинакова.

Алгоритм случайного леса оказался наиболее точным классификатором как для российских, так и для европейских фирм. Средняя точность классификации составляет 72.7 и 80.6% для российских и европейских фирм соответственно. Рисунок 11 показывает распределение точности алгоритма случайного леса, рассчитанной на произвольно сформированных тестовых наборах для российских и европейских организаций сферы услуг. Результаты теста Шапиро – Уилка указывают на то, что распределение точности ненормально для российских фирм (р-значения составляют 0.019 для российских фирм и 0.18 – для европейских), поэтому критерий Манна – Уитни использовался для оценки значимости разницы между средней точностью. Р-значение теста близко к нулю ($6.14 \cdot 10^{-23}$), а из этого следует, что вероятность получения такого значения критерия очень низка, если средняя точность для российских и европейских фирм одинакова.

Помимо общей точности моделей полезно проанализировать ошибки I и II рода. В Таблице 4 представлены средние значения ошибок I и II рода по наборам данных для российских и европейских фирм в соответствии с используемым алгоритмом. Результаты согласуются с выводами, полученными выше: ошибки как I, так и II рода выше в случае российских организаций сферы услуг по сравнению с европейскими фирмами из этой же сферы.

Таблица 4. Чувствительность, специфичность и ошибки классификации I и II родов (%)

		Чувствительность	Ошибка I рода	Специфичность	Ошибка II рода
Логистическая регрессия	Набор данных российских фирм	72.3	27.7	57.2	42.8
	Набор данных европейских фирм	74.9	25.1	86.3	13.7
k-ближайших соседей	Набор данных российских фирм	60.5	39.5	51.0	49.0
	Набор данных европейских фирм	76.7	23.3	68.7	31.3
Случайный лес	Набор данных российских фирм	73.4	26.6	72.5	27.5
	Набор данных европейских фирм	76.6	23.4	84.6	15.4
Среднее значение	Набор данных российских фирм	68.7	31.3	60.2	39.8
	Набор данных европейских фирм	76.1	23.9	79.9	20.1

Источник: составлено автором.

Заключение

Учитывая полученные результаты, можно утверждать, что прогнозирование дефолта, основанное на финансовых данных, менее эффективно для российских организаций сферы услуг, чем для фирм этой сферы, работающих на развитых европейских рынках. Точность в зависимости от алгоритма для российских фирм составляет 55–73% по сравнению точностью в 72–81% для фирм с развитых европейских рынков. Результаты для наборов данных по европейским фирмам с точки зрения общей точности согласуются с результатами предыдущих исследований [23], в то время как результаты для набора данных по российским фирмам сильно отстают.

Таким образом, в случае с российскими фирмами можно ожидать более высокой вероятности ошибки при прогнозировании дефолта на основании финансовых показателей. Иначе говоря, результаты указывают на то, что финансовые коэффициенты являются менее качественными показателями будущего дефолта для российских фирм, чем для фирм с развитых рынков.

Финансовая отчетность российских юридических лиц не отражает реальное состояние фирм по двум возможным причинам, приведенным в статье: дробление бизнеса и теневые операции. Использование нефинансовых показателей может быть полезным для увеличения точности классификации, что может быть отправной точкой для дальнейших исследований, связанных с прогнозированием дефолта в России.

Более того, результаты настоящего исследования можно обобщить в том смысле, что традиционный подход к прогнозированию дефолта может оказаться неприменимым не только для российских фирм из сферы услуг, но также для фирм из других развивающихся экономик, сталкивающихся с проблемой необъективной финансовой отчетности.

Дополнительным результатом данного исследования является сравнение предсказательной способности алгоритмов

классификации. Алгоритм случайного леса показывает самую высокую эффективность, подтверждая результаты предыдущих исследований [29; 30]. Несмотря на то, что это алгоритм линейной классификации, классификатор логистической регрессии также можно использовать для прогнозирования дефолта (81% точности в среднем по европейским фирмам). Однако алгоритм k-ближайших соседей представляется наименее точным (лишь 72% точности в среднем по европейским фирмам и только 55% – в среднем по российским фирмам, а это означает, что предсказательная сила алгоритма для российских фирм близка к нулю).

Список литературы

1. Altman E.I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*. 1968;23(4):589-609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
2. Matenda F.R., Sibanda M., Chikodza E., Gumbo V. Corporate default risk modeling under distressed economic and financial conditions in a developing economy. *Journal of Credit Risk*. 2021;17(1):89-115. <https://doi.org/10.21314/JCR.2020.267>
3. Fedorova E., Musienko S., Fedorov F. Analysis of the external factors influence on the forecasting of bankruptcy of Russian companies. *Vestnik Sankt-Peterburgskogo universiteta. Ekonomika = St Petersburg University Journal of Economic Studies*. 2020;36(1):117-133. (In Russ.). <https://doi.org/10.21638/spbu05.2020.106>
4. Grigoriev A., Tarasov K. Corporate bankruptcy prediction using the principal components method. *Journal of Corporate Finance Research*. 2019;13(4):20-38. <https://doi.org/10.17323/j.jcfr.2073-0438.13.4.2019.20-38>
5. Ohlson J.A. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*. 1980;18(1):109-131. <https://doi.org/10.2307/2490395>

6. Hunter J., Isachenkova N. Failure risk: A comparative study of UK and Russian firms. *Journal of Policy Modeling*. 2001;23(5):511-521. [https://doi.org/10.1016/S0161-8938\(01\)00064-3](https://doi.org/10.1016/S0161-8938(01)00064-3)
7. Gruszczyński M. Financial distress of companies in Poland. *International Advances in Economic Research*. 2004;10(4):249-256. <https://doi.org/10.1007/BF02295137>
8. Sirirattanaphonkun W., Pattarathammas S. Default prediction for small-medium enterprises in emerging market: Evidence from Thailand. *Seoul Journal of Business*. 2012;18(2):25-54. <https://doi.org/10.35152/snusbj.2012.18.2.002>
9. Ahmadpour Kasgari A., Divsalar M., Javid M.R., Ebrahimian S.J. Prediction of bankruptcy Iranian corporations through artificial neural network and Probit-based analyses. *Neural Computing and Applications*. 2013;23(3-4):927-936. <https://doi.org/10.1007/s00521-012-1017-z>
10. Kovacova M., Klietnik T. Logit and Probit application for the prediction of bankruptcy in Slovak companies. *Equilibrium*. 2017;12(4):775-791. <https://doi.org/10.24136/eq.v12i4.40>
11. Odom M., Sharda R. A neural network model for bankruptcy prediction. In: 1990 IJCNN International joint conference on neural networks (San Diego, CA, 17-21 June 1990). Piscataway, NJ: IEEE; 163-168. <https://doi.org/10.1109/IJCNN.1990.137710>
12. Coats P.K., Fant L.F. Recognizing financial distress patterns using a neural network tool. *Financial Management*. 1993;22(3):142-155. <https://doi.org/10.2307/3665934>
13. Altman E.I., Marco G., Varetto F. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance*. 1994;18(3):505-529. [https://doi.org/10.1016/0378-4266\(94\)90007-8](https://doi.org/10.1016/0378-4266(94)90007-8)
14. Zhang G., Hu M.Y., Patuwo B.E., Indro D.C. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*. 1999;116(1):16-32. [https://doi.org/10.1016/S0377-2217\(98\)00051-4](https://doi.org/10.1016/S0377-2217(98)00051-4)
15. Kumar P.R., Ravi V. Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research*. 2007;180(1):1-28. <https://doi.org/10.1016/j.ejor.2006.08.043>
16. Cao Y., Liu X., Zhai J., Hua S. A two-stage Bayesian network model for corporate bankruptcy prediction. *International Journal of Finance & Economics*. 2022;27(1):455-472. <https://doi.org/10.1002/ijfe.2162>
17. Mselmi N., Lahiani A., Hamza T. Financial distress prediction: The case of French small and medium-sized firms. *International Review of Financial Analysis*. 2017;50:67-80. <https://doi.org/10.1016/j.irfa.2017.02.004>
18. Xie C., Luo C., Yu X. Financial distress prediction based on SVM and MDA methods: The case of Chinese listed companies. *Quality & Quantity*. 2011;45(3):671-686. <https://doi.org/10.1007/s11135-010-9376-y>
19. Shumway T. Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*. 2001;74(1):101-124. <https://doi.org/10.1086/209665>
20. Fernando J.M.R., Li L., Hou G. Financial versus non-financial information for default prediction: Evidence from Sri Lanka and the USA. *Emerging Markets Finance and Trade*. 2020;56(3):673-692. <https://doi.org/10.1080/1540496X.2018.1545644>
21. Blanco-Oliver A., Irimia-Diéguez A., Oliver-Alfonso M., Wilson N. Improving bankruptcy prediction in micro-entities by using nonlinear effects and non-financial variables. *Finance a úvěr – Czech Journal of Economics and Finance*. 2015;65(2):144-166. URL: http://journal.fsv.cuni.cz/storage/1321_blanco_oliver.pdf
22. Boubaker S., Cellier A., Manita R., Saeed A. Does corporate social responsibility reduce financial distress risk? *Economic Modelling*. 2020;91:835-851. <https://doi.org/10.1016/j.econmod.2020.05.012>
23. Altman E.I., Sabato G., Wilson N. The value of non-financial information in SME risk management. *Journal of Credit Risk*. 2010;6(2):95-127. <https://doi.org/10.21314/JCR.2010.110>
24. Muñoz-Izquierdo N., Laitinen E.K., Camacho-Miñano M del-Mar, Pascual-Ezama D. Does audit report information improve financial distress prediction over Altman's traditional Z-Score model? *Journal of International Financial Management & Accounting*. 2020;31(1):65-97. <https://doi.org/10.1111/jifm.12110>
25. Makeeva E., Sinilshchikova M. News sentiment in bankruptcy prediction models: Evidence from Russian retail companies. *Journal of Corporate Finance Research*. 2020;14(4):7-18. <https://doi.org/10.17323/j.jcfr.2073-0438.14.4.2020.7-18>
26. Feng M., Shaonan T., Chihoon L., Ling M. Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*. 2019;274(2):743-758. <https://doi.org/10.1016/j.ejor.2018.10.024>
27. Beaver W.H. Financial ratios as predictors of failure. *Journal of Accounting Research*. 1966;4:71-111. <https://doi.org/10.2307/2490171>
28. Zhu Y., Xie C., Wang G.-J., Yan X.-G. Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SME credit risk in supply chain finance. *Neural Computing and Applications*. 2017;28(1):41-50. <https://doi.org/10.1007/s00521-016-2304-x>
29. Barboza F., Kimura H., Altman E. Machine learning models and bankruptcy prediction. *Expert Systems with Applications*. 2017;83:405-417. <https://doi.org/10.1016/j.eswa.2017.04.006>
30. Brown I., Mues C. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*. 2012;39(3):3446-3453. <https://doi.org/10.1016/j.eswa.2011.09.033>
31. Karminsky A. Corporate rating models for emerging markets. *Korporativnye finansy = Journal of Corporate Finance Research*. 2011;5(3):19-29. (In Russ.). <https://doi.org/10.17323/j.jcfr.2073-0438.5.3.2011.19-29>
32. Grishunin S., Egorova A. Comparative analysis of the predictive power of machine learning models for forecasting the credit ratings of machine-building companies. *Journal*

- of *Corporate Finance Research*. 2022;16(1):99-112. <https://doi.org/10.17323/j.jcfr.2073-0438.16.1.2022.99-112>
33. Kachalin D. Analysis of Russian models of splitting (reorganization) of business that ensure compliance of its scale with special taxation regime. *Finansovaya analitika: problemy i resheniya = Financial Analytics: Science and Experience*. 2011;(5):52-63. (In Russ.).
 34. Donich S.R. Novelties in the tax administration: The concept of splitting a business. *Vestnik Sibirskogo gosudarstvennogo universiteta putei soobshcheniya: Gumanitarnye issledovaniya = The Siberian Transport University Bulletin: Humanitarian Research*. 2021;(1):39-44. (In Russ.).
 35. Williams C.C., Nadin S., Newton S., Rodgers P., Windebank J. Explaining off-the-books entrepreneurship: A critical evaluation of competing perspectives. *International Entrepreneurship and Management Journal*. 2013;9(3):447-463. <https://doi.org/10.1007/s11365-011-0185-0>
 36. Jaki A., Ćwiąg W. Bankruptcy prediction models based on value measures. *Journal of Risk and Financial Management*. 2021;14(1):6. <https://doi.org/10.3390/jrfm14010006>
 37. Jandaghi G., Saranj A., Rajaei R., Ghasemi A., Tehrani R. Identification of the most critical factors in bankruptcy prediction and credit classification of companies. *Iranian Journal of Management Studies*. 2021;14(4):817-934. <https://doi.org/10.22059/IJMS.2021.285398.673712>
 38. Hosmer D.W. Jr., Lemeshow S., Sturdivant R.X. Introduction to the logistic regression model. In: *Applied logistic regression*. Hoboken, NJ: John Wiley & Sons, Inc.; 2013:1-33. (Wiley Series in Probability and Statistics). <https://doi.org/10.1002/9781118548387.ch1>
 39. Hassanat A.B., Abbadi M.A., Altarawneh G.A., Alhasanat A.A. Solving the problem of the K parameter in the KNN classifier using an ensemble learning approach. *International Journal of Computer Science and Information Security*. 2014;12(8):33-39. <https://doi.org/10.48550/arXiv.1409.0919>
 40. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32. <https://doi.org/10.1023/A:1010933404324>

Приложение 1

Количество фирм в наборах данных по категории услуг

Категория услуг	Данные по европейским фирмам		Данные по российским фирмам	
	Количество фирм	Доля фирм, %	Количество фирм	Доля фирм, %
Общественное питание и доставка продуктов	91	30	172	43
Прочие услуги	52	17	45	11
Медицинские и социальные услуги	49	16	58	14
Туризм, гостиничный бизнес и пассажирские перевозки	72	24	53	13
Культура, спорт и развлечения	21	7	66	16
Образование	19	6	10	2

Статья была представлена 25.12.2022; одобрена после рецензирования 23.01.2023; принята для публикации 10.02.2023.