

АКТУАЛЬНЫЕ ВОПРОСЫ ФИНАНСОВО-ЭКОНОМИЧЕСКОГО РАЗВИТИЯ

Международный научно-практический журнал "Вестник МИРБИС" ISSN 2411-5703 <http://journal-mirbis.ru/>
№ 4 (12) 2017 http://cs.journal-mirbis.ru/-/wF9hjEzYe7C2vJ88Ih1pw/sv/document/78/d5/56/521295/226/4_2017_VM.pdf?1517909660
URL статьи: http://cs.journal-mirbis.ru/-/wF9hjEzYe7C2vJ88Ih1pw/sv/document/78/d5/56/521295/226/4_2017_VM.pdf?1517909660#page=65

Статья получена: 20.12.2017

Ссылка для цитирования этой статьи: Егоров Н. В., Павлова И. В. Методы определения подлинности финансовых отчетов [Электронный ресурс] // Вестник Московской международной высшей школы бизнеса (МИРБИС). – 2017. № 4 (12). С. 65-72. – Библиогр.: с. 71 (16 назв.). – URL: http://cs.journal-mirbis.ru/-/wF9hjEzYe7C2vJ88Ih1pw/sv/document/78/d5/56/521295/226/4_2017_VM.pdf?1517909660#page=65 (доступ свободный). Загл. с экрана. Яз. рус., англ.

УДК 338.012

Николай Егоров¹, Ирина Павлова²

МЕТОДЫ ОПРЕДЕЛЕНИЯ ПОДЛИННОСТИ ФИНАНСОВЫХ ОТЧЕТОВ

Аннотация. Целью данного исследования является построение достоверной модели определения фиктивной финансовой отчетности. Объектами исследования являются компании, предоставлявшие фиктивные финансовые отчеты за период с 2010 по 2015 год. Анализ проводится в два этапа. На первом этапе анализа для выбора используемых признаков объекта используются алгоритм классификации и регрессии с помощью построения дерева решений (CART) и метод автоматического детектирования с использованием распределения хи-квадрат (CHAID). На втором этапе для решения задачи классификации применяются алгоритмы CART, CHAID, байесовские сети доверия (BBN), метод опорных векторов (SVM), искусственная нейронная сеть (ANN). На финальной стадии оценивается точность каждого метода и предлагается оптимальный вариант. В качестве предикатов используются финансовые и нефинансовые показатели из финансовой отчетности законопослушных компаний. Показатели выбирались таким образом, чтобы исключить возможность манипуляции данными показателями. Построенная модель позволяет обнаружить фиктивные финансовые отчеты с общей точностью 87,97 %. Данная модель может использоваться как дополнительный инструмент при аудиторских проверках. Также она может быть полезна собственникам бизнеса во избежание случаев, способных привести к банкротству компании.

Целью исследования является анализ предпосылок формирования рынка биотехнологий в России и определение роли и места российских биотехнологий на мировом рынке.

Методы: анализ, синтез и аналогия развития отраслей, подотраслей, организаций и предприятий, вовлеченных в процесс создания биотехнологий, применение графического изображения статистических данных.

Результаты: обоснование развития российского рынка биотехнологий для формирования организационно-экономического механизма развитием этой отрасли.

Ключевые слова: фиктивная финансовая отчетность, алгоритмы деревьев решений, байесовские сети доверия, метод опорных векторов, искусственные нейронные сети.

1 **Егоров Николай Валерьевич** – студент магистратуры. Финансовый Университет при Правительстве Российской Федерации, Россия, 125993, Москва, Ленинградский проспект, 49. E-mail: n.v.egorov@yandex.ru

2 **Павлова Ирина Владимировна** – д-р экон наук, профессор. Академия труда и социальных отношений (Россия, 119454, Москва, ул. Лобачевского, 90); Финансовый Университет при Правительстве Российской Федерации (Россия, 125993, Москва, Ленинградский проспект, 49). E-mail: IVPavlova@fa.ru

Финансовая отчетность это основной документ компании, отражающий ее финансовое состояние. Финансовая отчетность является основой для принятия решения со стороны большого количества инвесторов, кредиторов и других лиц. Финансовая отчетность отражает эффективности бизнеса, финансовое положение. В последние годы случаи участились случаи мошенничества при предоставлении финансовой отчетности.

Наступившая эпоха информационных технологий диктует новые «правила игры» для финансовых менеджеров. Применение программных алгоритмов для повышения эффективности работы стало повсеместным.

В данном исследовании ставилась задача построить такую

модель «признаки-объект-ответ», чтобы точность обнаружения фиктивных финансовых отчетов была максимальной. В качестве признаков будут выступать различные финансовые переменные, а ответов в данной задаче классификации будет два (0 – анализируемая финансовая отчетность является подлинной, 1 – анализируемая финансовая отчетность является подлинной фиктивной).

Актуальность исследования объясняется тем, что стоимость предотвращения подачи фиктивных финансовых отчетов очень высока, для примера в Соединенных Штатах эта стоимость оценивается в миллиарды долларов каждый год, поэтому построение достоверной модели обнаружения фиктивной финансовой отчетности имеет огромное значение

[6. С. 570–575]. Убытки от подачи фиктивных финансовых отчетов могут обходиться компании в миллиарды долларов, поэтому превентивные меры очень важны.

Американская ассоциация сертифицированных экспертов по мошенничеству (ACFE) классифицирует мошенничество на шесть типов:

- 1) предоставление ложной финансовой информации;
- 2) неправильное использование или незаконное присвоение корпоративных активов;
- 3) ненадлежащая поддержка или займы;
- 4) ненадлежащее приобретение активов;
- 5) обход затрат или сборов;
- 6) неправильная манипуляция финансированием руководителей или членов правления.

Поддельные финансовые отчеты представляют собой преднамеренные и незаконные действия, которые вводят в заблуждение всех участников финансовых операций. Интеллектуальный анализ данных (data mining) является ключевым инструментом для комплексного анализа и классификации данных. Интеллектуальный анализ данных идентифицирует события, которые скрыты в больших данных для анализа, и объединяет данные в структурированные модели, предоставляя сведения для принятия решений. Интеллектуальный анализ данных имеет множество различных функций, таких как классификация, объединение, кластеризация и прогнозирование. Функция классификации используется наиболее часто. Проблему обнаружения фиктивности финансовых отчетов можно рассматривать как типичную проблему классификации. Задача классификации включает в себя выполнение вычислений с использованием переменных характеристик известных данных классификации, чтобы получить классификационное правило. Впоследствии неизвестные данные классификации вводятся в правила, чтобы получить окончательные результаты классификации.

Использование метода интеллектуального анализа данных из-за его превосходства с точки зрения прогнозирования после ввода больших данных для машинного обучения, позволяет получить точность намного выше, чем у обычного регрессионного анализа. Существующие сейчас статистические модели имеют относительно высокий коэффициент ошибок. Большинство из используемых методов используют также одностадийную обработку. В данном исследовании применяется двух стадийная обработка и пять методов интеллектуального анализа данных. Обычные статистические методы требуют соблюдения с ограничениями конкретных гипотез, например, избегание коллинеарности независимых переменных и определенного распределения данных. Однако, эмпирические финансовые переменные часто не соответствуют статистическим условиям, таким как нормальное распределение. Метод машинного обучения не требует статистических гипотез комбинаций данных, поэтому может использоваться как классификатор. Эмпирические результаты показывают, что метод машинного обучения имеет положи-

тельный классификационный эффект. Данное исследование старается обобщить те модели, которые раньше использовались для определения фиктивных финансовых отчетов в работах Чжоу, Капура, Йеха [5. С. 1535–1541], Рависанкара [12. С. 491–500].

Одно из применений интеллектуального анализа данных – это обнаружение фальсифицированной финансовой отчетности. В этом исследовании используются несколько методов интеллектуального анализа данных: DT, BBN, SVM и ANN.

Деревья решений (DT) – самый простой метод машинного обучения [13. С. 97–106]. Процесс принятия решения по методу дерева решений следующий: 1) рисуется диаграмма дерева; 2) определяются значения предикатов; 3) выборка делится на две части; 4) выбирается ветвь с наибольшей ожидаемой стоимостью, остальные ветви обрезаются.

Один из самых известных алгоритмов для построения деревьев алгоритм ID3, разработанный в 1986 г. Данный алгоритм хорошо работает во многих задачах, но имеет свойство усложнять структуру дерева. Поэтому сейчас используются более современные алгоритмы, например CART [7. С. 24–30] и CHAID, которые и применяются в данном расчете. Ветвление деревьев может осуществляться по следующим критериям:

- 1) Индекс Джини;
- 2) D-критерий В. И. Донского;
- 3) Энтропийный критерий.

Далее используется критерий отсекающего, в котором используется обрезание на основе ошибок (EBP). Он вычисляет частоту ошибок каждого узла и определяет узлы, которые вызывают повышение в частоте ошибок DT, прежде чем включать соответствующую обрезку этих узлов, чтобы повысить точность DT.

Байесовская сеть доверия (BBN) играет важную роль в вопросах неопределенности и широко применяется во многих случаях, таких как оценка медицинских диагнозов и программного обеспечения. Это очень хороший метод для создания модели, способной отражать неопределенные факторы. Значения вероятности состояния соответствующих узлов корректируются согласно теореме Байеса. Структура диаграммы представляет причинные отношения и дает окончательные результаты путем вычисления вероятности. При получении новой информации BBN может регулировать вероятность. А именно, когда значение вероятности корректируется, все соответствующие узлы сети могут быть скорректированы в соответствии с условными вероятностями. BBN – направленный ациклический граф, состоящий из ряда узлов и стрелок и включающий в себя ряд переменных принятия решений. Эти переменные, принимающие решения, связаны в одном направлении, чтобы сформировать отношение родитель-потомство.

Метод опорных векторов (SVM) – один из самых известных методов в машинном обучении [4. С. 127–135]. Классификация происходит по знаку дискриминантной функции. Метод опорных векторов – это контролируемый алгоритм

машинного обучения, который может использоваться как для задач классификации, так и для регрессии. Однако он в основном используется в задачах классификации. В этом алгоритме мы записываем каждый элемент данных как точку в n -мерном пространстве (где n – количество функций, которые у нас есть) со значением каждой функции, являющейся значением конкретной координаты. Затем мы выполняем классификацию путем нахождения гиперплоскости, которая очень хорошо дифференцирует два класса. Гиперплоскость, разделяющую классы, можно построить разными способами. Максимизация расстояний между ближайшей точкой данных (либо класса) и гиперплоскостью поможет нам построить правильную гиперплоскость. Это расстояние называется предельным. SVM также имеет технику, выполняющую некоторые сложные преобразования данных, а затем выясняет процесс разделения данных на основе обозначений или выходов, которые вы определили.

Искусственные нейронные сети (ANN) обычно организованы слоями. Слои состоят из нескольких взаимосвязанных узлов, которые содержат функцию активации. Шаблоны представлены в сети через входной уровень, который связывается с одним или несколькими скрытыми слоями, где фактическая обработка выполняется через систему взвешенных соединений. Затем скрытые слои ссылаются на выходной уровень. Большинство ANN содержат некоторую форму правила обучения, которая изменяет веса соединений в соответствии с входными шаблонами, которые представлены. В некотором смысле, ANN учатся на примере. Хотя существует множество различных правил обучения, используемых нейронными сетями, наиболее часто используется правило дельты. В правиле дельты обучение – это контролируемый процесс, который возникает с каждым циклом или посредством

прямого потока активации выходов, а в случае обратного распространения корректируются веса ошибок. Проще говоря, когда нейронная сеть изначально представлена шаблоном, она делает случайную догадку относительно того, чем объект может быть. Затем сеть видит, как далеко ответ был от фактического и делает соответствующую корректировку весов объекта [13. С. 118-146].

В этом исследовании рассматриваются 47 азиатских компаний, которые выпустили фиктивные финансовые отчеты за период 2013–2015 гг. Эти компании включают одну компанию в строительной отрасли, две в пищевой промышленности, по одной в текстильной и волоконной промышленности, семь в полупроводниковой промышленности, девять в электронике, пять в фотоэлектрической промышленности, одну в телекоммуникационной отрасли, семь в других отраслях электроники, две в сталелитейной промышленности, две в резиновой промышленности, одну в судоходной отрасли, три в отрасли программного обеспечения, по одной в электромеханической промышленности и кабельной промышленности, и три в других отраслях. В исследовании выбираются законопослушные компании с аналогичными суммами активов в той же отрасли в предыдущем году и мошеннические компании в качестве сопоставимой выборки. Выбрано 178 компаний, в том числе 47 мошеннических компаний и 131 законопослушная компания.

Зависимая переменная используется для классификации компаний: для законопослушных компаний ее значение равно 0, для мошеннических 1. В исследовании используется 30 независимых переменных, включая двадцать три финансовые переменные и семь нефинансовых переменных. Исследуются следующие финансовые переменные:

- 1) коэффициент дебиторской задолженности $X_1 = \frac{\text{Дебиторская задолженность}}{\text{Валюта Баланса}}$
- 2) доля оборотных активов $X_2 = \frac{\text{Оборотные активы}}{\text{Валюта Баланса}}$
- 3) коэффициент основных средств $X_3 = \frac{\text{Основные средства}}{\text{Валюта Баланса}}$
- 4) коэффициент операционной прибыли $X_4 = \frac{\text{Операционный доход}}{\text{Валюта Баланса}}$
- 5) доля чистой прибыли по совокупным активам $X_5 = \frac{\text{Чистая прибыль}}{\text{Валюта Баланса}}$
- 6) доля чистой прибыли по основным средствам $X_6 = \frac{\text{Чистая прибыль}}{\text{Основные средства}}$
- 7) доля денежных средств от общей суммы активов $X_7 = \frac{\text{Денежные средства}}{\text{Валюта Баланса}}$
- 8) коэффициент общих активов $X_8 = \frac{\ln(\text{Активы})}{\text{Валюта Баланса}}$
- 9) коэффициент совокупных обязательств $X_9 = \frac{\ln(\text{Обязательства})}{\text{Обязательства}}$
- 10) коэффициент валовой прибыли $X_{10} = \frac{\text{Валовая прибыль}}{\text{Чистые продажи}}$
- 11) коэффициент операционных расходов $X_{11} = \frac{\text{Операционные расходы}}{\text{Чистые продажи}}$

- 12) коэффициент заемных средств $X_{12} = \frac{\text{Обязательства}}{\text{Валюта Баланса}}$
- 13) коэффициент текущей ликвидности $X_{13} = \frac{\text{Оборотные активы}}{\text{Текущие обязательства}}$
- 14) коэффициент текущих активов $X_{14} = \frac{\text{Мгновенные активы}}{\text{Текущие обязательства}}$
- 15) коэффициент оборота запасов $X_{15} = \frac{\text{Себестоимость проданных товаров}}{\text{Средний запас}}$
- 16) коэффициент денежного потока $X_{16} = \frac{\text{Операционный денежный поток}}{\text{Текущие обязательства}}$
- 17) коэффициент прибыли до налогообложения $X_{17} = \frac{\text{Прибыль до налогообложения}}{\text{Чистые продажи}}$
- 18) коэффициент оборота дебиторской задолженности $X_{18} = \frac{\text{Чистые продажи}}{\text{Средняя дебиторская задолженность}}$
- 19) темп роста продаж $X_{19} = \frac{\text{Продажи в текущем году} - \text{Прошлогодние продажи}}{\text{Средняя дебиторская задолженность}}$
- 20) доля заемного капитала к собственному $X_{20} = \frac{\text{Обязательства}}{\text{Собственный капитал}}$
- 21) коэффициент прибыли до налогообложения, уплаты процентов и амортизации $X_{21} = \frac{\text{EBITDA}}{\text{Активы}}$
- 23) доля текущих обязательств $X_{22} = \frac{\text{Текущие обязательства}}{\text{Валюта баланса}}$
- 24) общий оборот активов $X_{23} = \frac{\text{Чистые продажи}}{\text{Активы}}$

Нефинансовые переменные: доля участия в капитале основных акционеров (X24), двойственность директора совета директоров и генерального директора правления (X25), размер совета директоров (X26), коэффициент заложенных акций, удерживаемых директорами и руководителями (X27), коэффициент акций, удерживаемых директорами и руководителями (X28), проверка BIG4 (четыре крупные фирмы CPA) значение 1 для компаний, прошедших аудит BIG4, в противном случае, значение 0 (X28), количество внешних проверок (X30). В этом исследовании выбираются 30 переменных с наибольшим влиянием на фиктивные финансовые отчеты. Для этого применяют алгоритмы дерева решений (DT) CART и CHAID. Затем применяются CART, CHAID, BBN, SVM и ANN для построения модели обнаружения, а попарное сравнение тестовых групп проводится в терминах точности обнаружения (ошибки первого и второго типа).

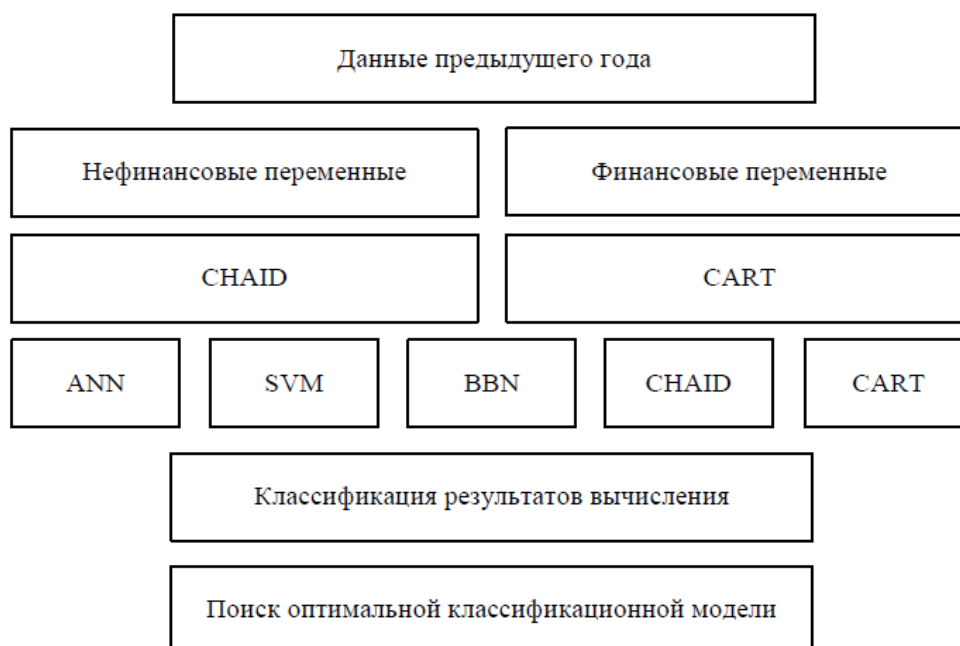


Рис. 4 – Структура исследования

Выбор условий сегментации определяется классами данных и их атрибутами. Условия сегментации определяются правилами Джини. Сегментация делит данные на два подмножества, а условия для следующей сегментации находятся в подмножествах. В таблице 1 показаны результаты использования алгоритма CART при выборе переменных. Только 4 из 30 переменных остались после выбора алгоритмом CART.

CHAID – это метод, который применяет критерий Хи-квадрат при вычислении значения вероятности узлов ветвей

и листьев дерева решений (DT) для определения следует ли продолжить сегментацию или нет. CHAID может предотвратить чрезмерное использование данных и позволить DT прекратить сегментацию. Таблица 1 иллюстрирует результаты применения алгоритма CHAID при выборе. Только 6 из 30 переменных остались после выбора алгоритма CHAID. Переменные представлены в таблице по степени важности. После отбора переменных проводится случайная выборка без повторения. Применяется десятикратная перекрестная проверка для проверки точности классификации.

Таблица 1 – Результаты применения алгоритма CHAID

Назначение весов переменным при помощи алгоритма CART	
Переменная	Вес переменной
Коэффициент денежного потока (X16)	0,496
Оборотные активы (X2)	0,48
Темпы роста продаж (X19)	0,016
Натуральный логарифм совокупных обязательств (X9)	0,008
Назначение весов переменным при помощи алгоритма CHAID	
Коэффициент заемных средств (X12)	0,2347
Коэффициент денежного потока (X16)	0,2181
Коэффициент текущих активов (X14)	0,2119
Оборотные активы (X2)	0,1337
Коэффициент прибыли до налогообложения (X21)	0,1282
Коэффициент операционных расходов (X11)	0,0734

Источник: таблица составлена авторами на основе источников – *Applied Predictive Modeling – M. Kuhn, K. Johnson (2013) [14]; Bayesian Reasoning and Machine Learning – D. Barber (2015) [15]; Core Concepts in Data Analysis: Summarization, Correlation and Visualization – Boris Mirkin [16].*

Как показано в таблице 2, точность определения фальсифицированной финансовой отчетности и общая точность модели CART-CART была самой высокой – 88,59 и 83,19 % соответственно. В этом исследовании также раскрываются ошибки каждой модели первого и второго типа, как показано в таблице 3.

Ошибки первого типа указывают на фиктивные финансовые отчеты, которые были ошибочно помечены как подлинны. Ошибки второго типа указывают на действительные финансовые отчеты, которые были классифицированы как

фиктивные финансовые отчеты. Логически, ошибка первого типа является большей ошибкой. Поэтому, помимо сравнения с точностью тестирования, также учитывается частота ошибок первого типа. Модель CART-CART имеет самый низкий показатель ошибок первого типа на уровне 11,41 % и общий коэффициент ошибок 16,81 %.

Как показано в таблице 2, точность определения фиктивной финансовой отчетности и общая точность модели CHAID-CART самая высокая 92,69 и 87,97 %. Ошибки первого и второго типа модели CHAID показаны в таблице 3.

Таблица 2 – Точность определения фиктивной финансовой отчетности и общая точность модели CHAID-CART

Модель	Фиктивная финансовая отчетность (%)	Законная финансовая отчетность (%)	Общая точность (%)
Точность обнаружения моделей CART – десятикратная перекрестная проверка			
CART-CART	88,59	77,78	83,19
CART-CHAID	81,88	79,51	80,70
CART-BBN	77,18	73,22	75,20
CART-SVM	75,17	74,19	74,68
CART-ANN	75,84	74,16	75,00
Точность обнаружения моделей CHAID – десятикратная перекрестная проверка			
CHAID-CART	92,69	83,24	87,97
CHAID-CHAID	79,19	71,37	75,28

Модель	Фиктивная финансовая отчетность (%)	Законная финансовая отчетность (%)	Общая точность (%)
CHAID-BBN	81,88	80,13	81,01
CHAID-SVM	79,87	78,23	79,05
CHAID-ANN	83,20	81,59	82,40

Источник: таблица составлена авторами на основе источников – *Applied Predictive Modeling* – M. Kuhn, K. Johnson (2013) [14]; *Bayesian Reasoning and Machine Learning* – D. Barber (2015) [15]; *Core Concepts in Data Analysis: Summarization, Correlation and Visualization* – Boris Mirkin [16].

Таблица 2 – Ошибки первого и второго типа модели CHAID

Модель	Ошибки первого типа (%)	Ошибки второго типа (%)	Общий коэффициент ошибок (%)
Ошибки первого и второго типа модели CART			
CART-CART	11,41	22,22	16,81
CART-CHAID	18,12	20,49	19,30
CART-BBN	22,82	26,78	24,80
CART-SVM	24,83	25,81	25,32
CART-ANN	24,16	25,84	25,00
Ошибки первого и второго типа модели CHAID			
CHAID-CART	7,31	16,76	12,03
CHAID-CHAID	20,81	28,63	24,72
CHAID-BBN	18,12	19,87	18,99
CHAID-SVM	20,13	21,77	20,95
CHAID-ANN	16,80	18,41	17,60

Источник: таблица составлена авторами на основе источников – *Applied Predictive Modeling* – M. Kuhn, K. Johnson (2013) [14]; *Bayesian Reasoning and Machine Learning* – D. Barber (2015) [15]; *Core Concepts in Data Analysis: Summarization, Correlation and Visualization* – Boris Mirkin [16].

Финансовый отчет компании является основой для суждений. Если руководство пытается удержать информацию, даже когда речь идет о независимых аналитиках инвестиционно-банковских услуг и ценных бумаг, инвесторы могут испытывать значительные потери. Самый известный скандал - дело о банкротстве Enron. Топ-менеджмент намеренно ввел в заблуждение инвесторов, чтобы получить прибыль свыше одного млрд. долл. США. Это вызвало банкротство многих инвесторов и сокращение сотрудников и повлияло на все бизнес сообщество. Дело Enron заставило инвесторов потерять уверенность в финансовой отчетности и привело к формированию аудиторских комитетов, возглавляемых независимыми директорами и реформам стандартов бухгалтерского учета и реконструкция механизмов регулирования.

На самом деле, необычные признаки в финансовой отчетности часто возникают до начала скандала. Иррациональность можно найти в финансовой отчетности от нескольких кварталов до одного года до момента подачи. Фиктивные финансовые отчеты могут выглядеть презентабельно, и многие инвесторы могут быть обмануты. Однако лучше предотвратить мошенничество, защитить инвесторов от обмана и обеспечить наказание преступников. Вопрос защиты инвесторов нельзя однозначно решить с помощью правовых норм и жестких требований к надзору. Все

большее число случаев фиктивной финансовой отчетности может нанести ущерб компаниям, и приводят к крупным потерям для инвесторов. Собственники платят тяжелую цену, чтобы компенсировать этот ущерб. Поэтому создание эффективной модели обнаружения фиктивного финансового отчета очень важно.

Данное исследование направлено на предоставление метода нетрадиционного анализа с использованием нескольких алгоритмов данных, включая: DT, BBN, SVM и ANN. На первом этапе исследования применяется алгоритмы CART и CHAID для выбора важных переменных. CART, CHAID, BBN, SVM и ANN затем объединяются для построения классификационной модели для сравнения. Согласно результатам исследования, эффективность обнаружения (общая точность) модели CHAID-CART является лучшей на 87,97 % (точность обнаружения фиктивная финансовой отчетности составляет 92,69 %). Данная модель также имеет самый низкий показатель ошибок первого типа 7,31 %. Для сравнения общая точность других моделей: 83,19 % для модели CART-CART, 82,40 % для модели CHAID-ANN, 81,01% для модели CHAID-BBN, 80,70 % для модели CART-CHAID, 79,05 % для модели CHAID-SVM, 75,28 % для модели CHAID-CHAID, 75,20 % для CART-BBN, 75,00 % для модели CART-ANN и 74,68 % для CART-SVM модели. Основываясь на эмпирических результатах этого исследования, точность алгоритма

CHAID дерева решений, объединенного с алгоритмом CART при обнаружении фиктивных финансовых отчетов довольно высока. Данные алгоритмы поэтому могут использоваться в качестве инструмента для оказания помощи аудиторам в выявлении фиктивных финансовых отчетов.

Результаты исследования могут быть полезны для инвесторов, акционеров, руководителей компаний, кредитных рейтинговых учреждений, аудиторов, сертифицированных государственных бухгалтеров, аналитиков по ценным бумагам, органам финансового регулирования.

Литература

1. Ронова Г. Н., Ронова Л. А. Анализ финансовой отчетности : учебное пособие. — М: Евразийский открытый институт, 2009. — 179 с.
2. Пожидаева Т. А. Анализ финансовой отчетности : учебное пособие — М: КноРус, 2016. — 320 с.
3. Сапожникова Н. Г. Международные стандарты финансовой отчетности : учебное пособие — М: КноРус, 2016. — 367 с.
4. Чубукова И. А. Data Mining : учебное пособие. — М: Интуит НОУ, 2016. — 471 с.
5. Yeh C. C., Chi D. J., Hsu M. F. A hybrid approach of DEA, rough set and support vector machines for business failure prediction, *Expert Syst Appl*, 2011, 37:1535–1541.
6. Zhou W., Kapoor G., Detecting evolutionary financial statement fraud, *Decis Support Syst*, 2011, 50:570–575.
7. Salehi M., Fard F. Z., Data mining approach to prediction of going concern using classification and regression tree (CART), *Glob J Manag Bus Res*, 2013, 13(3):24–30.
8. Shin K. S., Lee T. S., Kim H. J. An application of support vector machines in bankruptcy prediction model. *Expert Syst Appl*, 2005, 28:127–135.
9. Humpherys S. L., Moffitt K. C., Burns M.B., Burgoon J. K., Felix W. F. (2011) Identification of fraudulent financial statements using linguistic credibility analysis. *Decis Support Syst* 50: 585–594.
10. Kamarudin KA, Ismail WAW, Mustapha WAHW (2012) Aggressive financial reporting and corporate fraud. *Procedia Soc Behav Sci* 65:638–643. DOI: 10.1016/j.sbspro.2012.11.177.
11. Pai P. F., Hsu M. F., Wang M. C. (2011) A support vector machine-based model for detecting top management fraud. *Knowl Based Syst* 24:314–321
12. Ravisankar P., Ravi V., Rao G. R., Bose I. (2011) Detection of financial statement fraud and feature selection using data mining techniques. *Decis Support Syst* 50:491–500
13. Kirkos S., Spathis C., Manolopoulos Y. (2007) Data mining techniques for the detection of fraudulent financial statements. *Expert Syst Appl* 32(4): 995–1003
14. Kuhn M., Johnson K. (2013) *Applied Predictive Modeling*. Springer, New York, NY. doi: 10.1007/978-1-4614-6849-3
15. Barber D. (2015) *Bayesian Reasoning and Machine Learning*. — URL: <http://web4.cs.ucl.ac.uk/staff/D.Barber/text-book/181115.pdf>
16. Boris Mirkin. *Core Concepts in Data Analysis: Summarization, Correlation and Visualization*. — URL: https://www.hse.ru/data/2010/10/14/1223126254/Mirkin_All.pdf

Nikolay Egorov¹, Irina Pavlova²

METHODS OF FINDING AUTHENTICITY OF FINANCIAL STATEMENTS

Abstract. Objective of this research is creation of reliable model of detection of fictitious financial statements. Objects of a research are the companies providing fictitious financial statements from 2010 for 2015. The analysis is carried out in two steps. At the first analysis stage for the choice of the used signs of an object are used an algorithm of classification and regression by means of creation of a tree of decisions (CART) and a method of automatic detecting with distribution use a chi-square (CHAID). At the second stage algorithms of CART, CHAID, Bayesian networks of trust (BBN), a method of basic vectors (SVM), artificial neural network (ANN) are applied

1 **Egorov Nikolay Valerievich** – student of magistracy Financial University under the Government of the Russian Federation. 49, Leningradsky prospect, Moscow, 125993, Russia. E-mail: n.v.egorov@yandex.ru

2 **Pavlova Irina Vladimirovna** – Doctor of Economics, Professor Academy of Labor and Social Relations (90, Lobachevsky st., 119454, Russia); Financial University under the Government of the Russian Federation (Leningradsky prospect, Moscow, 125993, Russia). E-mail: IVPavlova@fa.ru

to the solution of a problem of classification. At a final stage the accuracy of each method is estimated and the optimal variant is offered. As predicates a financial and non-financial performance from financial statements of the law-abiding companies is used. Indicators were chosen so that to exclude a possibility of manipulation with these indicators. The constructed model allows to find fictitious financial statements with the general accuracy of 87,97%. This model can be used as the additional tool at the audit inspections. Also it can be useful to owners of business to avoiding of the cases capable to result in bankruptcy of the company.

Key words: fictitious financial statements, algorithms of trees of decisions, Bayesian networks of trust, method of basic vectors, artificial neural networks.

References

1. Ronova G. N., Ronova L. A. *Analiz finansovoy otchetnosti [Analysis of financial statements] : a textbook.* – M: *Yevraziyskiy otkrytyy institut*, 2009. – 179 p.
2. Pozhidayeva T. A. *Analiz finansovoy otchetnosti [Analysis of financial statements] : a manual.* – M: *KnoRus*, 2016. – 320 p.
3. Sapozhnikova N. G. *Mezhdunarodnyye standarty finansovoy otchetnosti [International Financial Reporting Standards] : A Training Manual.* – M: *KnoRus*, 2016. – 367 p.
4. Chubukova I. A. *Data Mining : a tutorial.* – M: *Intuit NOU*, 2016. – 471 p.
5. Yeh C. C., Chi D. J., Hsu M. F. A hybrid approach of DEA, rough set and support vector machines for business failure prediction, *Expert Syst Appl*, 2011, 37:1535–1541.
6. Zhou W., Kapoor G., Detecting evolutionary financial statement fraud, *Decis Support Syst*, 2011, 50:570–575.
7. Salehi M., Fard F. Z., Data mining approach to prediction of going concern using classification and regression tree (CART), *Glob J Manag Bus Res*, 2013, 13(3):24–30.
8. Shin K. S., Lee T. S., Kim H. J. An application of support vector machines in bankruptcy prediction model. *Expert Syst Appl*, 2005, 28:127–135.
9. Humpherys S. L., Moffitt K. C., Burns M. B., Burgoon J. K., Felix W. F. (2011) Identification of fraudulent financial statements using linguistic credibility analysis. *Decis Support Syst* 50: 585–594.
10. Kamarudin KA, Ismail WAW, Mustapha WAHW (2012) Aggressive financial reporting and corporate fraud. *Procedia Soc Behav Sci* 65:638–643. DOI: 10.1016/j.sbspro.2012.11.177.
11. Pai P. F., Hsu M. F., Wang M. C. (2011) A support vector machine-based model for detecting top management fraud. *Knowl Based Syst* 24:314–321
12. Ravisankar P., Ravi V., Rao G. R., Bose I. (2011) Detection of financial statement fraud and feature selection using data mining techniques. *Decis Support Syst* 50:491–500
13. Kirkos S., Spathis C., Manolopoulos Y. (2007) Data mining techniques for the detection of fraudulent financial statements. *Expert Syst Appl* 32(4): 995–1003
14. Kuhn M., Johnson K. (2013) *Applied Predictive Modeling.* Springer, New York, NY. doi: 10.1007/978-1-4614-6849-3
15. Barber D. (2015) *Bayesian Reasoning and Machine Learning.* – URL: <http://web4.cs.ucl.ac.uk/staff/D.Barber/text-book/181115.pdf>
16. Boris Mirkin. *Core Concepts in Data Analysis: Summarization, Correlation and Visualization.* – URL: https://www.hse.ru/data/2010/10/14/1223126254/Mirkin_All.pdf