

МАТЕМАТИЧЕСКИЕ МОДЕЛИ И ЧИСЛЕННЫЕ МЕТОДЫ

УДК 330.45

МЕТОДЫ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ
В ЗАДАЧАХ ФИНАНСОВОГО КОНТРОЛЯ

© 2022 В.М. Сушков*, П.Ю. Леонов**

Национальный исследовательский ядерный университет «МИФИ», Москва, 115409, Россия

*e-mail: vmsushkov@mail.ru

**e-mail: p.u.leonov@gmail.com

Поступила в редакцию: 28.08.2022

После доработки: 26.12.2022

Принята к публикации: 27.12.2022

С каждым годом увеличивается объем корпоративных данных, подлежащих анализу в рамках финансового контроля, вследствие чего актуальным является внедрение методов обработки больших данных в практику субъектов контроля. Целью исследования является разработка и апробация методов обработки больших данных в целях решения задач организаций государственного и частного сектора, проводящих мероприятия в области финансового контроля. В качестве методов исследования выбрано три наиболее перспективных и эффективных средств обработки больших данных, которые в то же время не требуют использования громоздкого математического аппарата или значительных компьютерных мощностей для их реализации, а именно статистический инструмент выявления искажений в финансовых данных закон Бенфорда, кластеризация методом К-средних и средства BI-системы Power BI. Результатом исследования является подтверждение результативности и экономической эффективности рассматриваемых методов обработки больших данных, а также обоснование практической возможности их внедрения в качестве инструментов финансового контроля. Исследование проведено в студенческой Лаборатории финансовой разведки НИЯУ МИФИ.

Ключевые слова: большие данные, финансовый контроль, мошенничество, закон Бенфорда, метод К-средних, Power BI.

DOI: 10.26583/vestnik.2022.5

ВВЕДЕНИЕ

В настоящее время в связи с увеличивающимся объемом корпоративной информации усложняется процесс проведения контрольных мероприятий как на государственном уровне (проводится Счетной палатой, Банком России, Федеральной налоговой службой, пр.), так и на частном (проводится субъектами первичного финансового мониторинга, аудиторскими организациями, службами внутреннего контроля, пр.). Количество заключаемых договоров, проводимых платежных операций, составляемых первичных документов, а также бухгалтерских записей, оформляющих всю совокупность фактов хозяйственной жизни экономического субъекта, порой исчисляется сотнями миллионов за период. В этой связи актуальным представляется внедрение современных методов и средств обработки больших данных (от англ. – Big Data) в практику субъектов контроля в целях оптимизации их деятельности. Технологии больших

данных позволяют значительно сократить временные издержки на проведение проверок посредством отбора наиболее подозрительных элементов из общего массива данных.

МЕТОДЫ

В качестве методов исследования принято решение выбрать наиболее перспективные и эффективные средства обработки больших данных, которые в то же время не требуют использования громоздкого математического аппарата или значительных компьютерных мощностей для их реализации. Таким образом, методологическую основу исследования составили следующие технологии:

- 1) статистический инструмент закон Бенфорда;
- 2) алгоритм кластеризации метод К-средних;
- 3) средства BI (Business Intelligence) системы Power BI.

**Метод 1. Выявление искажений
в финансовых данных с помощью закона
Бенфорда**

В мировой практике для выявления преднамеренного искажения информации широкое распространение получил метод статистического анализа, известный как закон Бенфорда (реже закон первой цифры, закон аномальных чисел) [1]. Данный закон определяет вероятность появления конкретной значащей цифры в распределениях величин, взятых из реальной жизни [2].

Чтобы проверить закон Бенфорда на практике, необходимо рассмотреть множество первых цифр – элементов исследуемого

числового массива и сравнить фактическую частоту их появления с теоретической, определенной Ф. Бенфордом в 1938 г. [3]. Аналогично можно провести анализ второй, третьей и т.д. цифры. Выявляя отклонение теоретического распределения от эмпирического, можно судить о вероятности ошибки в данных или их намеренного искажения.

Ожидаемая вероятность появления цифры d_1 в первом разряде описывается формулой:

$$P(d_1) = \log_{10} \left(1 + \frac{1}{d_1} \right); \quad d_1 = 1, 2, \dots, 9. \quad (1)$$

Распределение Бенфорда для десятичной системы счисления, таким образом, можно представить в виде гистограммы (рис. 1).

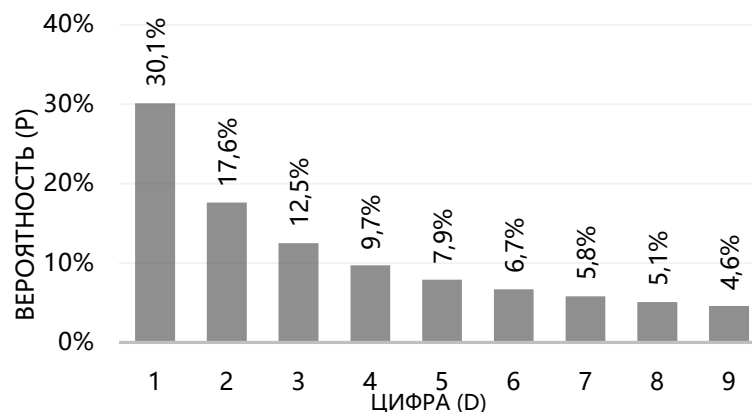


Рис. 1. Распределение Бенфорда для десятичной системы счисления

Теоретически доказано, что распределение Бенфорда для первых двух цифр имеет вид, представленный на рис. 2 [4].

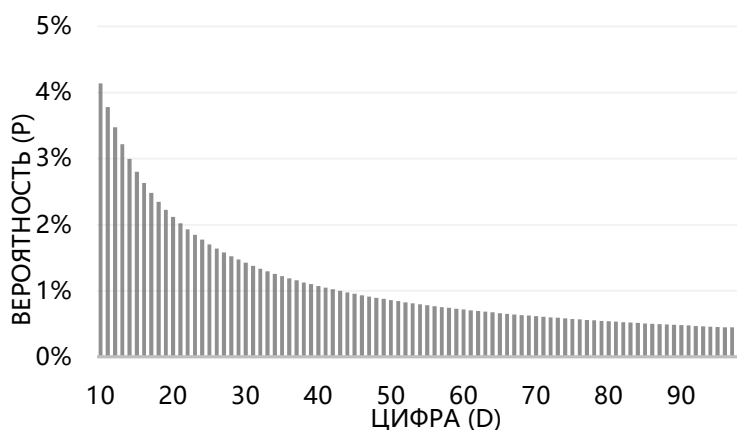


Рис. 2. Распределение Бенфорда для первой пары цифр

В конце XX в. были опубликованы работы американского ученого в области бухгалтерского учета, аудита и математики Марка Нигрини. В своих работах М. Нигрини проанализировал более 200 000 налоговых деклараций и пришел к выводу, что закон Бенфорда применим ко многим наборам финансовых данных, включая данные о дебиторской и кредиторской задолженности, подоходном налоге, фондовых биржах, корпоративных расходах и продажах, показателям выручки и прибыли организации [5].

Метод 2. Обнаружение корпоративного мошенничества методом К-средних

Кластерный анализ – многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о совокупности объектов, с последующим упорядочиванием объектов в сравнительно однородные группы. В зависимости от объема n классифицируемых наблюдений и от априорной информации о числе кластеров методы кластеризации делятся на иерархические и неиерархические (итерационные) [6].

Метод К-средних является одним из наиболее распространенных и часто используемых итерационных методов кластерного анализа. Данный метод был изобретен в 1950-х гг. математиком Гуго Штейнгаузом и почти одновременно Стюартом Ллойдом. Особую популярность приобрел после работы Мак-Куина, поэтому этот метод также называют методом Мак-Куина.

Предлагается применить метод К-средних в целях формирования репрезентативной нестатистической выборки с элементами нестатистического метода «серийного отбора» при проведении мероприятий финансового контроля.

Метод 3. Идентификация подозрительных бухгалтерских записей с помощью средств Power BI

На данный момент контрольные субъекты зачастую испытывают сложности в случае необходимости проверки всего массива бухгалтерских записей экономического субъекта, внесенных в учетную систему даже за небольшой период. Вследствие этого необходима разработка удобного и интуитивно понятного инструмента, позволяющего точно и быстро выявить наиболее высокорискованные записи с минимальными временными затратами.

В качестве среды разработки инструмента было выбрано программное обеспечение Power BI, а в качестве тестируемой системы – бухгалтерская система 1С как самая распространенная на российском рынке.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Результатом исследования является подтверждение результативности и экономической эффективности рассматриваемых методов обработки больших данных, а также обоснование практической возможности их внедрения в качестве инструментов финансового контроля.

Результаты апробации метода 1 (закон Бенфорда)

В проведенном в Лаборатории финансовой разведки НИЯУ МИФИ исследовании были взяты две компании, функционирующие в разных отраслях: «А» (IT-сфера) и «Б» (строительный бизнес). Компания «А» не скрывает свои реальные доходы, совершает законные операции с денежными средствами, не манипулирует финансовыми данными и в полном размере выплачивает налоги в пользу государства. Эта компания уже более 20 лет ведет свою деятельность и на сегодняшний день занимает лидирующие позиции на рынке предлагаемых услуг. В компании «Б» ситуация обратная. Известно, что в прошлых отчетных периодах руководство осуществляло неоднократные действия по незаконному уходу от налогообложения. Вуалируя реальные показатели о результатах финансовой деятельности, руководство заключало фиктивные договорные отношения с третьими лицами, занижало доходы и завышало расходы, а также реализовывало сложные схемы с участием подставных фирм.

Ниже представлены данные компании «А» по платежным операциям за оказанные услуги в 2018 г. Из всего массива анализ проводился только в отношении оплат поставщикам, которых за год было более 3000 шт. В каждой операции были выделены две первые цифры и проанализированы частоты их появления. Результаты представлены на рис. 3.

Полученные результаты близки к теоретическим, однако можно заметить выбросы. Они соответствуют многочисленным фиксированным платежам одному контрагенту. Несмотря на отсутствие идеального совпадения с теоретическими значениями, наблюдаемые результаты хорошо с ними согласуются и наблюдается экспоненциальное убывание в первых двух цифрах старших разрядов финансовых операций.

Аналогичным способом были проанализированы расчеты с контрагентами в компании «Б», которая была замешана в проведении финансовых махинаций и привлечена к ответственности за свои действия. Сгруппированные данные по оплатам поставщикам за 2018 г. представлены в виде гистограммы, при этом определены частоты появления первой пары цифр в старших разрядах операций (рис. 4).

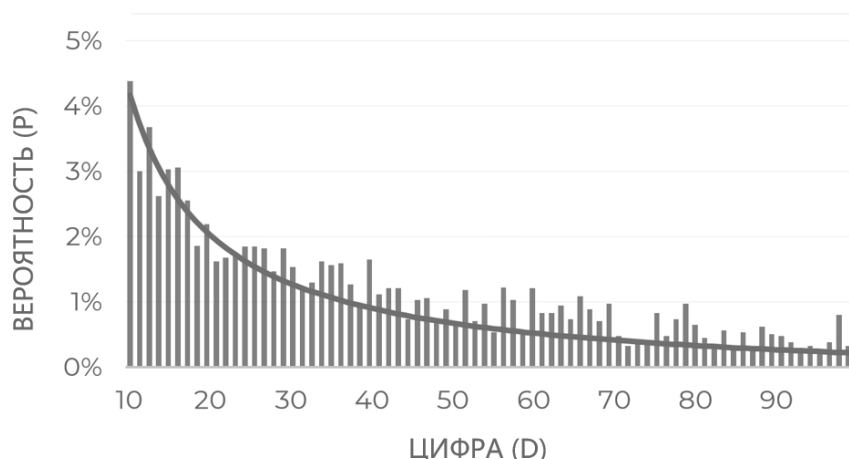


Рис. 3. Частота совпадений первой пары цифр в операциях компании «А»

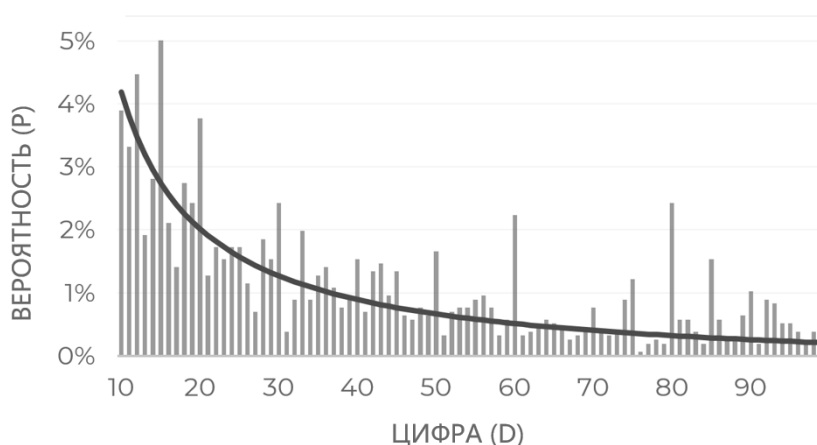


Рис. 4. Частота совпадений первой пары цифр в операциях компании «Б»

Как видно из приведенных данных, результаты существенно отличаются от теоретически рассчитанных. Количество выбросов значительно превышает число однотипных операций, проводимых в отношении одного контрагента. Наблюдается множество отклонений как в большую, так и в меньшую сторону, которые нельзя объяснить особенностями проводимых операций. Оценивая компанию «Б», как неизвестную ранее, можно предположить, что в предоставляемых финансовых отчетах существуют ошибки или намеренные искажения. Подобная ситуация должна послужить сигналом для более детального анализа хозяйственной деятельности компании.

На основании результатов, представленных выше, можно сделать вывод о целесообразности применения закона Бенфорда для диагностики недостоверности финансовой отчетности организации. Установлено, что расчеты по оплате поставщикам за год имеют характер, близкий к теоретическому в случае, если исследуемая организация не маскирует реально существующие финансовые показатели и предоставляет фактические данные в налоговые

органы. В том случае если организация в целях незаконного ухода от налогообложения прибегает к различным способам манипулирования отчетными данными, характер их распределения будет значительно отличаться от теоретического, определенного Фрэнком Бенфордом. Неподчинение учетных данных закону Бенфорда является поводом для их уточнения и проведения более глубокого анализа.

Результаты апробации метода 2 (метод K-средних)

Апробация метода производилась в рамках выявления хищений при продаже нефтепродуктов на автозаправочных станциях.

Проверка проводилась в три этапа:

1) подготовительный этап (определение проверяемой совокупности, выбор свойств объекта и количества кластеров, обоснование выбора кластера (кластеров) после применения метода K-средних);

2) применение метода K-средних (этап проведения расчетов);

3) анализ полученных результатов (сравнение полученных результатов с результатами, полученными при проведении сплошной проверки).

Подготовительный этап

Кластерный анализ проводился относительно дат первого полугодия 2017 г. (181 объект), в которых предполагалось присутствие хищений горюче-смазочных материалов сотрудниками. В качестве свойств дат предложено выбрать:

- X – суточная выручка автозаправочных станций;

- Y – доля суточной выручки автозаправочных станций в максимальной за неделю, рассчитываемая по формуле:

$$y_i = \frac{x_i}{\max_{i-3 \leq p \leq i+3} x_p}; i = 1, \dots, 181, \quad (2)$$

где i – дни первого полугодия 2017 г.; p – дни недели; x_i – суточная выручка; $\max_{i-3 \leq p \leq i+3} x_p$ – максимальная за неделю выручка.

Количество кластеров решено принять равное четырем. Данный выбор обусловлен содержанием кластеров, а именно:

1) дни, в которых получена наибольшая выручка в течение текущей недели и периода в целом;

2) дни, в течение которых получена выручка, имеющая небольшое отклонение от максимальной в течении текущей недели;

3) дни, в течение которых получена средняя выручка относительно текущей недели и периода в целом;

4) дни, в которых получена наименьшая выручка в течение текущей недели и периода в целом.

Для дальнейшего исследования необходимо взять элементы второго кластера, поскольку вероятность того, что при большом количестве продаж за день и располагая временем на совершение манипуляций с суточным отчетом за смену сотрудник решится на проведение хищения горюче-смазочных материалов больше, чем в остальных случаях:

- при очень большом количестве продаж нефтепродуктов легко скрыть их хищение, однако у работников практически не будет времени проводить различные мошенничества по хищению нефтепродуктов, поэтому включение первого кластера в аудиторскую выборку является спорным;

- средние отклонения в выручке при продаже нефтепродуктов могут быть связаны как с небольшим количеством покупателей на протяжении всего дня, так и с влиянием погоды (при низкой температуре объем горюче-смазочных материалов падает, несмотря на то что масса не меняется), следовательно, даты, попавшие в третий кластер, не попадут в аудиторскую выборку;

- большие отклонения от максимальной выручки редко связаны с хищениями, разумный работник понимает – хищение на большую сумму имеет высокий уровень риска для работника, так как такие хищения легко обнаружить, следовательно, исследование объектов четвертого кластера неразумно.

Применение метода К-средних

Учитывая тот факт, что выбранные свойства имеют разный диапазон значений, в первую очередь необходимо провести стандартизацию данных. Для этого были рассчитаны средние значения и среднеквадратические отклонения каждого свойства.

По итогам кластеризации 181 объект был разделен следующим образом: 1 кластер – 36 дней; 2 кластер – 70 дней; 3 кластер – 42 дня; 4 кластер – 33 дня.

В подлежащую анализу выборку было включено 70 дат, принадлежащим второму кластеру. Проверка в целях выявления хищений нефтепродуктов проходила следующим образом:

- проведено сравнение данных по реализации нефтепродуктов через топливно-распределительные колонки по видам оплат, отраженных в итоговых отчетах за смену с аналогичными итоговыми отчетами за смену, распечатанных из программы «ТОПАЗ-АЗС»;

- проведено сравнение данных по расходу нефтепродуктов согласно показаниям программного счетчика учета нефтепродуктов, прошедших через рукава топливно-распределительных колонок, отраженных в итоговых отчетах за смену с аналогичными итоговыми отчетами за смену, распечатанных из программы «ТОПАЗ-АЗС».

Вместо проведения сравнения различных отчетов за 181 день, было проведено сравнение отчетов за 70 дней, попавших в выборку, что на 61.33 % сократило время проверки. По итогам сравнения было выявлено, что в 31-й день из 70 принадлежащих выборке были произведены хищения нефтепродуктов.

Анализ полученных результатов

Для того чтобы проанализировать эффективность применения кластерного анализа методом *K*-средних была проведена проверка сплошным методом, т.е. сравнение отчетов производилось относительно всех 181 дней. По результатам проведенной сплошной проверки было выяв-

лено, что в 50 днях выбранного периода производились хищения горюче-смазочных материалов сотрудниками автозаправочных станций. Выявленные дни мошенничества были рассмотрены на принадлежность к полученным методом *K*-средних кластерам (рис. 5).

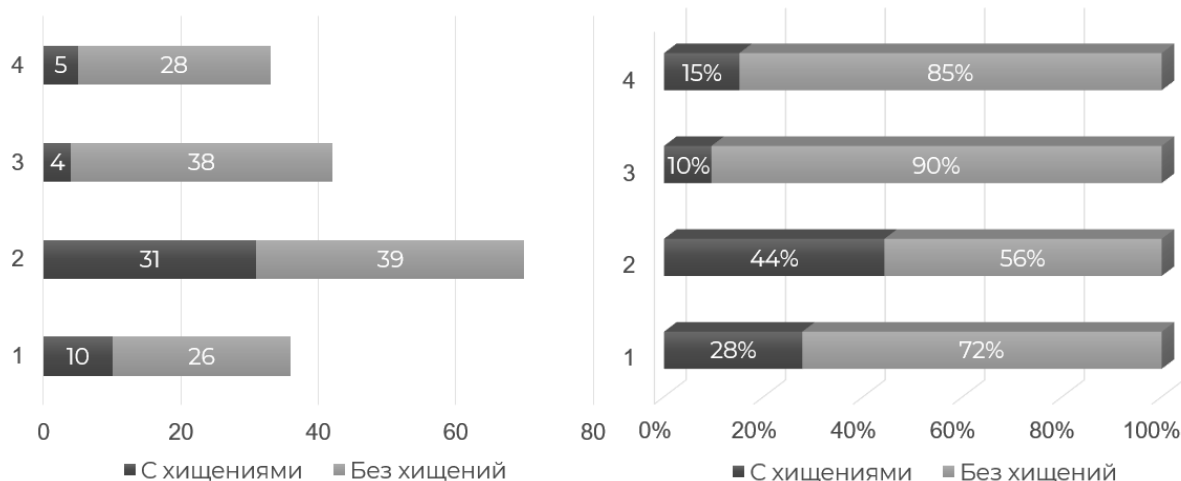


Рис. 5. Принадлежность дней с хищениями к полученным кластерам в числовом и процентном выражении

Анализ полученных результатов позволил сделать вывод о том, что при использовании метода *K*-средних при построении выборки при проведении контрольного мероприятия, направленного на выявление хищений горюче-смазочных материалов на автозаправочных станциях, длительность проверки сократится почти в три раза (на 61,33 %), при этом будет выявлено 62 % дат, в которых были произведены хищения.

Метод *K*-средних служит инструментом, способствующим разделению данных на группы, однако направления его использования определяются специалистами на основе собственного опыта и профессионального суждения.

Результаты апробации метода 3 (средства Power BI)

Первым этапом проведения анализа явилась выгрузка данных. Совокупность бухгалтерских записей может быть выгружена из учетной системы 1С экономическим субъектом самостоятельно или проверяющими лицами. Вместе с бухгалтерскими записями должна быть сформирована и выгружена оборотно-сальдовая ведомость, необходимая на этапе сверки [7].

Вторым этапом является заполнение специального файла (рис. 6) в формате .xlsx, в

котором указывается необходимая для анализа информация, а именно:

- даты начала и окончания финансовых кварталов в организации;
- праздничные дни и официальные выходные;
- режим работы компании в субботу и воскресенье;
- уровень существенности.

Третьим этапом является приведение данных в необходимый вид, который указан в инструкции, и загрузка бухгалтерских записей, оборотно-сальдовой ведомости, а также заполненного файла в готовый отчет.

Инструмент в Power BI состоит из подготовительной и рабочей части, которая представляет собой совокупность различных тестов. В подготовительную часть входит общий обзор массива данных, проведение этапа сверки и расчет порогового значения.

Сверка необходима для подтверждения полноты и целостности выгруженных данных (рис. 7). Для этого сверяются общие суммы по дебету и кредиту счетов из массива бухгалтерских записей с суммами, указанными в оборотно-сальдовой ведомости. В случае расхождений необходимо выявить причину ошибок, при их отсутствии можно продолжать работу.

Данный документ заполняется для получения необходимой информации для проведения тестирования.

Наименование организации ООО "НПП"	
Рабочая материальность 300,000,000	Рабочую материальность стоит указывать, в случае необходимости расчета порога. Если необходим анализ полной выборки, то РМ указывать не нужно.
Дата начала первого квартала 1/1/2019	Дата окончания первого квартала 3/31/2019
Дата начала первого квартала 4/1/2019	Дата окончания первого квартала 6/30/2019
Дата начала первого квартала 7/1/2019	Дата окончания первого квартала 9/30/2019
Дата начала первого квартала 10/1/2019	Дата окончания первого квартала 12/31/2019

Интересующие даты	
1/1/2019	Новогодние праздники
1/2/2019	Новогодние праздники
1/7/2019	Рождество Христово
2/23/2019	День защитника Отечества
3/8/2019	Международный женский день
5/1/2019	Праздник Весны и Труда
5/9/2019	День Победы
6/12/2019	День России

Выходные	
Суббота	Да
Воскресенье	Нет

Рис. 6. Пример заполненного файла, предваряющего анализ

Счет	Кол-во проводок Дт	Кол-во проводок Кт	Проводки Дт	Проводки Кт	ОСВ Дт	ОСВ Кт	Разница Дт	Разница Кт
01	42	6	0.00	0.00	0.00	0.00	0.00	0.00
01.01			0.00	0.00	0.00	0.00	0.00	0.00
02			0.00	0.00	0.00	772,091.10	0.00	772,091.10
02.01		36	0.00	772,091.10	0.00	772,091.10	0.00	0.00
10			0.00	0.00	181,510.75	181,510.75	181,510.75	181,510.75
10.09	25	26	181,510.75	181,510.75	181,510.75	181,510.75	0.00	0.00
19			0.00	0.00	136,469,238.05	182,913,107.54	136,469,238.05	182,913,107.54
19.01		1	0.00	527,980.34	0.00	527,980.34	0.00	0.00
19.03	518	552	134,871,039.98	180,777,130.20	134,871,039.98	180,777,130.20	0.00	0.00
19.04	165	158	1,598,198.07	1,607,997.00	1,598,198.07	1,607,997.00	0.00	0.00
41			0.00	0.00	951,243,105.01	1,091,195,837.79	951,243,105.01	1,091,195,837.79
41.01	713	1766	951,243,105.01	1,091,195,837.79	951,243,105.01	1,091,195,837.79	0.00	0.00
44			0.00	0.00	18,864,555.53	18,746,874.08	18,864,555.53	18,746,874.08
44.01	592	167	18,864,555.53	18,746,874.08	18,864,555.53	18,746,874.08	0.00	0.00
50			0.00	0.00	45,001.00	38,963.75	45,001.00	38,963.75
50.01	3	11	45,001.00	38,963.75	45,001.00	38,963.75	0.00	0.00
51	1003	1205	1,671,732,108.59	1,683,165,818.28	1,671,732,108.59	1,683,165,818.28	0.00	0.00
Всего	10530	10488	8,587,405,385.15	8,587,405,385.15	16,633,701,067.06	16,839,291,688.88	8046295681.91	8251886303.73

Рис. 7. Пример результатов сверки

Этап расчета порогового значения необходим для отсека записей, сумма которых в совокупности не превышает уровень существенности.

Рабочая часть инструмента состоит из тестов, в которых субъект контроля самостоятельно определяет подозрительность записей на основе понимания организации и итоговых выборок, формирующихся на основе заранее подготовленных критериев подозрительности.

В инструмент внедрены следующие тесты:

1) корреспонденция – тест позволяет более подробно проанализировать все корреспонденции счетов и отобрать те, которые являются нетипичными;

2) сторнирование выручки – выявление неосновательного признания выручки, которое в последующем было сторнировано, особенно в периоде, близком к отчетной дате;

3) счета – тест схож с тестом «корреспонденция» и позволяет выявлять редко используемые и нетипичные счета, используемых в оформлении операций;

4) интересующие даты – тест выявляет записи, внесенные в систему в нерабочие дни (праздники и выходные);

5) записи, сделанные в даты закрытия кварталов;

6) записи с подозрительными суммами;

7) записи на круглые суммы;

8) записи на дублированные числа – записи, заканчивающиеся на одно и тоже повторяющееся число, например 111;

9) дублированные записи – записи по одному документу, с одним и тем же содержанием на одинаковую сумму;

10) пользователи – статистика пользователей для определения тех, кто не уполномочен вносить записи или внес небольшое число записей, но на большую сумму.

Тесты с анализом подозрительных записей находятся на нескольких слайдах среды Power BI. Пример результатов одного из тестов (подозрительные счета) представлен на рис. 8.

Созданный инструмент по выявлению подозрительных бухгалтерских операций с помощью средств Power BI успешно справляется с поставленными задачами по анализу больших данных, позволяя значительно экономить время субъектов контроля на проведение анализа.

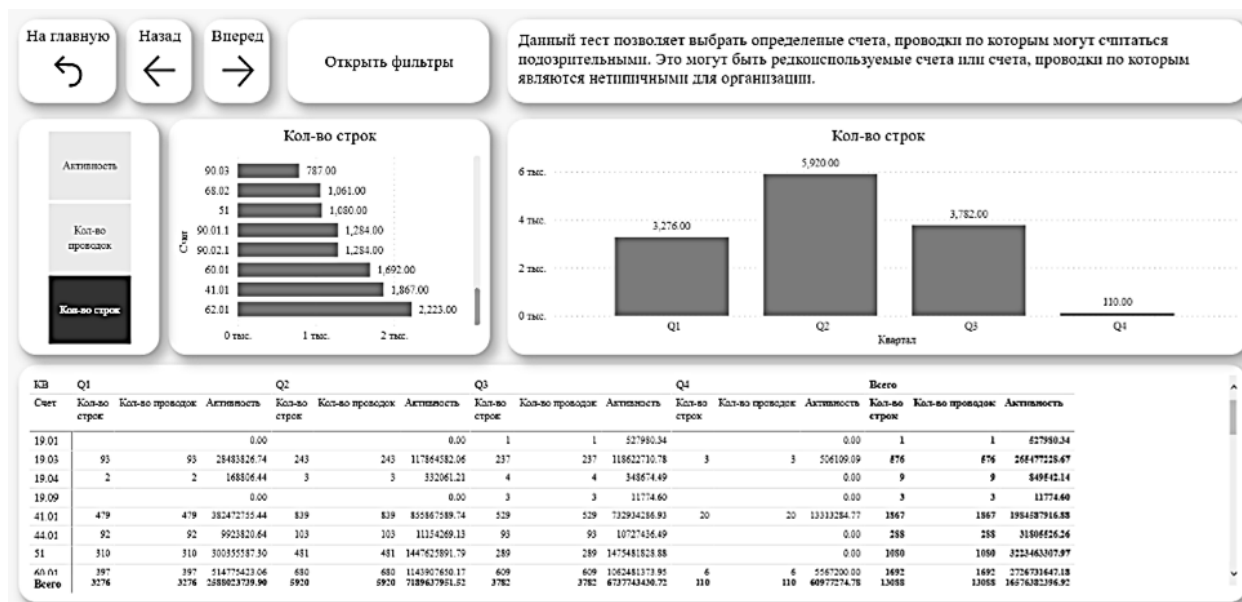


Рис. 8. Слайд с тестом на подозрительные счета

ЗАКЛЮЧЕНИЕ

Таким образом, современные методы и средства обработки больших данных обеспечивают всесторонний анализ массивов корпоративных данных любого объема, позволяя выявлять в них наиболее уязвимые элементы. В

этой связи внедрение описанных в работе инструментов в практику организаций государственного и частного сектора, проводящих мероприятия в области финансового контроля, а также разработка новых на сегодняшний день являются крайне актуальными задачами.

СПИСОК ЛИТЕРАТУРЫ

1. Leonov P.Y., Suyts V.P., Rychkov V.A., Ezhova A.A., Sushkov V.M., Kuznetsova N.V. Possibility of Benford's Law Application for Diagnosing Inaccuracy of Financial Statements / Klimov, V.V., Kelley, D.J. (eds) *Biologically Inspired Cognitive Architectures* 2021. BICA 2021. // *Studies in Computational Intelligence*, 2021. Vol. 1032. Springer, Cham. P. 243–248.
2. Алексеев М.А. Применимость закона Бенфорда для определения достоверности финансовой отчетности // *Вестник НГУЭУ*. 2016. № 4. С. 114–128.
3. Суйц В.П., Хорин А.Н., Жакипбеков Д.С. Диагностика недостоверности отчетности организации: статистические методы в оценке достоверности бухгалтерской отчетности // *Аудит и финансовый анализ*. М.: ООО Издательство «ДСМ Пресс». 2015. № 1. С. 179–188.
4. Зверев Е., Никифоров А. Распределение Бенфорда: Выявление нестандартных элементов в больших совокупностях финансовой информации // *Внутренний контроль в кредитной организации*. 2018. № 4 (40). С. 4–18.
5. Nigrini M.J. Benford's law: applications for forensic accounting, auditing and fraud detection. – Hoboken, New Jersey: John Wiley & Sons, 2012. С. 320.
6. Leonov P.Y., Suyts V.P., Kotelyanets O.S., Ivanov N.V. K-Means Method as a Tool of Big Data Analysis in Risk-Oriented Audit // *Communications in Computer and Information Science*. 2019. Т. 1054. Кв. 3. С. 206–216.
7. Лушин Л.Э. Алгоритм применения программного обеспечения Power BI субъектами первичного финансового мониторинга в целях выявления подозрительных бухгалтерских записей: дипл. работа. М.: НИЯУ МИФИ, 2022.

Vestnik Natsional'nogo issledovatel'skogo yadernogo universiteta «MIFI», 2022, vol. 11, no. 5, pp. 348–357

BIG DATA PROCESSING TECHNIQUES IN FINANCIAL CONTROL TASKS

V.M. Sushkov*, P.Y. Leonov**

National Nuclear Research University MEPhI, Moscow, 115409, Russia

**e-mail: vmsushkov@mail.ru*

***e-mail: p.u.leonov@gmail.com*

Received August 28, 2022; revised December 26, 2022; accepted December 27, 2022

Every year the volume of corporate data to be analysed in financial controls increases, which makes it relevant to introduce Big Data processing techniques into the practice of control subjects. The purpose of the study is to develop and test three techniques of Big Data processing in order to solve the problems of public and private sector organisations carrying out activities in the field of financial control. As research methods, we have chosen three most promising and effective means of processing Big Data, which at the same time do not require the use of complicated mathematical apparatus or significant computer power for their implementation, namely, a statistical tool for detecting errors in financial data Benford's law, clustering method of K-means and the means BI-system Power BI. The result of the study is the confirmation of the effectiveness and cost efficiency of the considered techniques of processing Big Data and the justification of the practical possibility of their implementation as financial control tools. The research was conducted in the student Financial Intelligence Laboratory of NRNU MEPhI.

Keywords: Big Data, financial control, fraud, Benford's Law, K-means, Power BI.

REFERENCES

1. Leonov P.Y., Suyts V.P., Rychkov V.A., Ezhova A.A., Sushkov V.M., Kuznetsova N.V. Possibility of Benford's Law Application for Diagnosing Inaccuracy of Financial Statements. Klimov, V.V., Kelley, D.J. (eds) *Biologically Inspired Cognitive Architectures* 2021. BICA 2021. *Studies in Computational Intelligence*, 2021. Vol. 1032. Springer, Cham. P. 243–248.
2. Alekseev M.A. *Primenimost' zakona Benforda dlja opredeleniya dostovernosti finansovoj otchetnosti* [Applicability of Benford's Law for determining the reliability of financial statements]. *Vestnik NGUJeU*, 2016. № 4. P. 114–128 (in Russian).
3. Sujc V.P., Horin A.N., Zhakipbekov D.S. *Diagnostics nedostovernosti otchetnosti organizatsii: statisticheskie metody v ocenke dostovernosti buhgalterskoj otchetnosti* [Diagnosing the unreliability of an organisation's accounts: statistical methods in assessing the reliability of accounting records]. *Audit i finansovyy analiz*. M.: ООО Izdatel'stvo «DSM Press» Publ. 2015. № 1. P. 179–188 (in Russian).

4. *Zverev E., Nikiforov A.* Raspredelenie Benforda: Vyjavlenie nestandartnyh jelementov v bol'shih sovokupnostjah finansovoj informacii [Benford's distribution: Identifying irregularities in large sets of financial information]. *Vnutrennij kontrol' v kreditnoj organizacii*. 2018. № 4 (40). P. 4–18 (in Russian).

5. *Nigrini M.J.* Benford's law: applications for forensic accounting, auditing and fraud detection. Hoboken, New Jersey, John Wiley & Sons. 2012. P. 320.

6. *Leonov P.Y., Suyts V.P., Kotelyanets O.S., Ivanov N.V.* K-Means Method as a Tool of Big Data Analysis in

Risk-Oriented Audit. *Communications in Computer and Information Science*, 2019. Vol. 1054. Q. 3. P. 206–216.

7. *Lushin L.Je.* Algoritm primenenija programmnogo obespechenija Power BI sub#ektami pervichnogo finansovogo monitoringa v celjah vyjavlenija podozritel'nyh buhgalterskih zapisej: dipl. rabota [Algorithm of application of Power BI software by reporting entities to identify suspicious accounting entries: graduate thesis]. M.: NIJaU MIFI [NRNU MEPhI], 2022 (in Russian).