УДК 519.873

А.С. Колтайс, А.А. Шатрова, А.А. Шелупанов

Математическая модель выбора контрагента при оценке рисков информационной безопасности

Рассмотрены основные методы, используемые при построении математической модели оценки благонадёжности контрагента. Описаns форматы преобразования данных для возможности их использования в ходе моделирования, а также предъявляемые требования к модели. Была проведена проверка зависимости между входными данными для исключения из модели одинаковых атрибутов. Выявлены основные сложности в оценке точности и полноты модели. Описано использование кроссвалидации для устранения сложностей при построении модели. Разработанная модель дает точный результат при малом количестве сравниваемых контрагентов, что соответствует порядку проверки контрагента в реальной системе.

Ключевые слова: модель, благонадёжность, риски, информационно-аналитические системы, машинное обучение.
doi: 10.21293/1818-0442-2020-23-2-36-41

В последние годы внимание к проблеме наличия недекларированных возможностей в программном обеспечении усилилось – выпущены несколько регламентирующих документов.

ФСТЭК России приказом № 131 от 30 июля 2018 г. утвердил «Требования по безопасности информации, устанавливающие уровни доверия к средствам технической защиты информации и средствам обеспечения безопасности информационных технологий» [1], а 11 февраля 2019 г. утвердил «Методику выявления уязвимостей и недекларированных возможностей в программном обеспечении» [2]. Банк России утвердил стандарт СТО БР ИББС-1.4-2018 «Обеспечение информационной безопасности организаций банковской системы Российской Федерации. Управление риском информационной безопасности при аутсорсинге» [3], в котором обязывает подведомственные организации управлять и контролировать риски нарушения информационной безопасности, в том числе при аутсорсинге разработки программного обеспечения. Выявление вероятностных нарушителей, создание модели нарушителя безопасности конфиденциальной информации – один из основных этапов проведения предпроектного обследования и формирования требований по защите информации, обрабатываемой и хранимой в информационных системах [4]. Таким образом, оценка вероятности внедрения разработчиком недекларированных возможностей в программное обеспечение, в частности, при аутсорсинге разработки, является актуальной задачей.

Специфической особенностью разработки АСУ ТП является распределение программного обеспечения на несколько уровней – уровень исполнительных устройств, уровень автоматического управления и уровень операторского управления. Для качественной оценки вероятности внедрения недекларированных возможностей необходимо учитывать структуру системы, а также интерфейсы взаимодействия между подсистемами АСУ ТП и между различными программными комплексами.

Вероятностная модель для оценки потенциального внедрения недекларированных возможностей в программные комплексы является актуальной для двух этапов жизненного цикла разработки ПО – выбор исполнителя (компании-разработчика) и согласования технического задания. Для этапа приемки работ актуальным является применение методов выявления уязвимостей и недекларированных возможностей.

К примеру, при разработке вероятностной модели для этапа выбора компании-разработчика – основными направлениями построения планируются:

– анализ потенциальных компаний разработчиков на основе параметров, характеризующих благонадёжность компаний и их потенциальные связи с конкурентами (в том числе иностранными);

– учет имеющихся в открытом доступе источников информации о деятельности контрагента в сфере информационной безопасности («Государственный реестр сертифицированных средств защиты информации» и т.п.).

При этом под благонадёжностью будем понимать свойство контрагента, отражающее отсутствие предрасположенности совершать в процессе совместной договорной деятельности действия или бездействие, наносящие ущерб компании в форме прямых материальных потерь и (или) в качестве снижения или потери его деловой репутации.

Проверку корпоративной информации компании в своей работе рассмотрели О. На, В.П. Ли, Ю.Я. Ким и Г. Чанг, а также предложили рейтинговую модель проверки информации в рамках обеспечения экономической безопасности организации, основанной на 14 факторах, которые включают общую информацию об организации, внутреннюю информацию организации и дополнительные данные. Но эта модель дает только общую оценку потенциальному контрагенту, не затрагивая особенности законодательства в различных сферах и оценку рисков со стороны информационной безопасности [5].

Описание модели

В настоящее время частичную проверку благонадёжности контрагента, как правило, осуществляют отдельные специализированные отделы компании на основе информационных систем, например,

систем СПАРК, Интегрум и др. Так как данные проверки затратны по времени и могут содержать ошибки вследствие влияния человеческого фактора, а также не учитываются специфические для отрасли параметры, существует необходимость автоматизировать процесс оценки рисков управленческого решения по сотрудничеству с контрагентом (компанией-разработчиком) [6–8].

Для этого была построена модель по оценке риска сотрудничества с контрагентом.

В процессе моделирования в качестве входных данных для модели использовалась выборка из информационно-аналитической системы (ИАС) СПАРК [9], которая представляет собой список контрагентов и их характеристики: возраст компании, среднесписочная численность, важная информация, чистая прибыль (убыток), период погашения дебиторской задолженности (в днях), период погашения кредиторской задолженности (в днях), период оборота основных средств, период оборота активов, коэффициент текущей ликвидности, коэффициент быстрой ликвидности, коэффициент абсолютной ликвидности.

В качестве выходных данных выступает список числовых состояний контрагентов: (−1) – контрагент не благонадёжный, (+1) – контрагент благонадёжный.

При этом модель должна соответствовать следующим требованиям:

– адекватность;
– адаптивность;
– эффективность;
– точность модели должна превышать 80%;
– полнота модели должна превышать 80%.

При этом под точностью модели [10] понимается отношение верных срабатываний к сумме верных и ложных срабатываний.

Под полнотой модели [10] понимается отношение верных срабатываний к сумме верных срабатываний и ложных пропусков.

В качестве метода построения модели будет применяться метод стохастического градиента, который может использоваться при машинном обучении [11]. При этом сначала будет производиться обучение на выборке $x^l = (x_i, y_i)_{i=1}^l$, где $x_i \in \mathbb{R}^n$ – объекты, $y_i \in \{-1; +1\}$ – ответы «учителя».

Самой моделью будет являться линейная модель классификации

$$a(\mathbf{x};\mathbf{w}) = \operatorname{sign}\langle \mathbf{x},\mathbf{w}\rangle, \qquad (1)$$

где $\operatorname{sign}\langle \mathbf{x},\mathbf{w}\rangle$ – знак скалярного произведения вектора **x** на вектор **w**.

Чтобы найти неизвестный вектор **w**, поставим задачу минимизации функционала доли неправильных ответов

$$Q(a, X) = \frac{1}{l}\sum_{i=1}^{l}\left[a(x_i;\mathbf{w}) \neq y_i\right] = \\ = \frac{1}{l}\sum_{i=1}^{l}\left[\operatorname{sign}\langle \mathbf{x},\mathbf{w}\rangle \neq y_i\right] \to \min_w. \qquad (2)$$

Этот функционал является дискретным относительно весов, и поэтому, чтобы можно было воспользоваться градиентом методом, необходимо свести задачу к минимизации гладкого функционала:

$$Q(\mathbf{w}, X) = \frac{1}{l}\sum_{i=1}^{l}\left[y_i \langle x_i, \mathbf{w}\rangle < 0\right] \to \min_w. \qquad (3)$$

Введем величину $M_i(\mathbf{w}) = y_i(x_i\mathbf{w})$ – отступ (margin) объекта x_i:

$$Q(\mathbf{w}, X) = \frac{1}{l}\sum_{i=1}^{l}[M_i(\mathbf{w}) < 0] \to \min_w. \qquad (4)$$

Знак отступа показывает корректность ответа классификатора (положительный отступ соответствует правильному ответу, отрицательный – неправильному), а его абсолютная величина характеризует степень уверенности классификатора в своём ответе. Скалярное произведение $\langle \mathbf{x},\mathbf{w}\rangle$ пропорционально расстоянию от разделяющей гиперплоскости до объекта соответственно, чем ближе отступ к нулю, тем ближе объект к границе классов, тем ниже уверенность в его принадлежности.

Функционал согласно формуле (2) оценивает ошибку алгоритма на объекте с помощью пороговой функции потерь $L(M) = [M < 0]$, где аргументом функции является отступ $M(\mathbf{w}) = y\langle \mathbf{x},\mathbf{w}\rangle$. Чтобы ввести гладкий функционал, оценим эту функцию во всех точках M: $L(M) \le \tilde{L}(M)$.

После этого можно получить верхнюю оценку на функционал

$$Q(\mathbf{w}, X) = \frac{1}{l}\sum_{i=1}^{l}\tilde{L}(y_i \langle \mathbf{x},\mathbf{w}\rangle) \to \min_w. \qquad (5)$$

При этом если верхняя оценка $\tilde{L}(M)$ будет гладкой, то и $L(M)$ будет гладкой. Если верхнюю оценку удастся приблизить к нулю, то и доля неправильных ответов тоже будет близка к нулю.

В качестве верхней оценки будет использоваться логистическая функция потерь

$$\tilde{L}(M) = \log(1 + e^{-M}). \qquad (6)$$

Таким образом, целевая функция будет выражаться в минимизации верхней оценки $\tilde{L}(M)$.

Построение модели

Для моделирования были рассмотрены компании, основной вид деятельности которых, согласно ОКВЭД, «62 – разработка компьютерного программного обеспечения, консультационные услуги в данной области и другие сопутствующие услуги».

Общее количество компаний согласно ИАС СПАРК [9] составило 34 324, из которых 46,53% компаний относятся к «благонадёжным», 53,47% – к «неблагонадёжным». Для моделирования была взята выборка с расширением .csv из 3 432 компаний, которая соответствует распределению благонадёжности в исходной выборке (табл. 1). При этом благонадёжность определялась по свободному индикатору риска (табл. 2): если риск «низкий», то компания благонадёжна, если «средний» или «высокий», то компания неблагонадёжна [9].

Таблица 1

Распределение благонадежности в исходной выборке и в выборке для моделирования

Вид компании	Кол-во	%	Выборка для моделирования	%
Благонадёжная	15 970	46,53	1 597	46,53
Неблагонадёжная	18 354	53,47	1 835	53,47
Общее кол-во	34 324	100	3 432	100

Таблица 2

Риски в ИАС СПАРК

Риски в ИАС СПАРК		
ИДО	ИФР	ИПД
Индекс должной осмотрительноти представляет собой значение от 1 до 99, где более высокое значение отражает большую вероятность того, что компания создана не для уставных целей, а в качестве «транзакционной единицы», не имеющей существенных собственных активов и операций, или является «брошенным» активом	Индекс финансового риска представляет собой значение от 1 до 99, где более высокое значение указывает на наличие признаков неудовлетворительного финансового состояния, которые могут привести к тому, что компания утратит платежеспособность	Индекс платежной дисциплины (Paydex) представляет собой значение от 0 до 100, где более низкое значение указывает на высокий риск просрочки платежей. Индекс платежной дисциплины рассчитывается автоматически на основании данных по платежам компании

Была проведена проверка наличия зависимости между входными данными путем подсчета коэффициента корреляции и его оценки по шкале Чеддока. В результате было решено исключить «коэффициент текущей ликвидности» и «коэффициент быстрой ликвидности» из входных данных, так как между ними и «коэффициентом абсолютной ликвидности» присутствует сильная и очень сильная связь (значения коэффициента корреляции от 0,7 до 0,9 и больше 0,9 соответственно).

Для успешного применения метода машинного обучения над входными данными были осуществлены следующие преобразования:

– разделитель целой и дробной части в Excel был определен как «.»;

– убрано разделение групп разрядов в Excel;

– преобразованы данные в текстовом формате к числовому формату согласно табл. 3;

– преобразованы аномальные значения: все отрицательные значения периода погашения кредиторской задолженности, периода погашения дебиторской задолженности, периода оборота основных средств, периода оборота активов были определены как «–1»;

– убраны пропуски в данных: пропуски в «Размер компании» были определены как «–1», остальные пропуски как «0»;

– приведены данные к одному масштабу от 0 до 1.

Таблица 3

Преобразование данных в текстовом формате к числовому формату

Наименование	Преобразование
Размер компании	0 – микропредприятия, 1 – малые предприятия, 2 – средние предприятия, 3 – крупные предприятия
Важная информация	0 – важной информации нет, 1 – важная информация есть

При моделировании использовался метод linear_model.SGDClassifier() библиотеки sklearn Python. Значения атрибутов принимались по умолчанию [12], кроме следующих:

– loss – функция потерь, значение – '*log*';

– random_state используется, чтобы была возможность повторить эксперименты, значение – 70;

– alpha – константа, которая умножается на penalty (по умолчанию – 0,0001), проведены эксперименты для 0,01, 0,001, 0,0001;

– max_iter – максимальное количество проходов по тренировочным данным (так называемые эпохи), по умолчанию – 1 000, было выявлено оптимальное максимальное количество проходов, чтобы избежать проблему переобучения;

– tol – критерий остановки: обучение остановится, когда loss > best_loss – tol (по умолчанию – 0,001), проведены эксперименты для None, 0,01, 0,001, 0,0001.

При проведении экспериментов оказалось, что точность и полноту в силу несбалансированности ответов модели и реальной системы сложно оценивать, поэтому было решено заменить оценку точности и полноты на оценку ROC-кривой (AUC-ROC), которая показывает зависимость количества верно классифицированных положительных ответов от количества неверно классифицированных отрицательных ответов. При этом критерий AUC-ROC должен быть более 80% [13–15].

Эксперименты проводились отдельно для каждого атрибута (табл. 4).

Таблица 4

Результаты экспериментов над атрибутами модели

Атрибут	Точность	Полнота	Доля верных ответов	AUC-ROC
alpha = 0,01	0,69	0,84	0,76	0,75
alpha = 0,001	0,75	0,73	0,76	0,76
alpha = 0,0001	0,76	0,69	0,76	0,75
tol = None	0,76	0,70	0,76	0,75
tol = 0,01	0,70	0,87	0,77	0,76
tol = 0,001	0,76	0,69	0,76	0,76
tol = 0,0001	0,82	0,62	0,71	0,75

По итогам экспериментов было решено установить max_iter = 30, чтобы уменьшить время обучения, alpha = 0,001, tol оставить по умолчанию. Итоговое значение критерия AUC-ROC составляет порядка 75%.

При этом модель не соответствует требованиям (критерий AUC-ROC меньше 80%). Для повышения точности результатов моделирования на этапе обучения модели можно использовать кроссвалидацию [15]: при оценке модели имеющиеся в наличии данные разбиваются на k частей, затем на $k-1$ частях данных производится обучение модели, а оставшаяся часть данных используется для тестирования. Процедура повторяется k раз; в итоге каждая из k частей данных используется для тестирования.

В результате модель на более равномерных данных обучается эффективнее, вследствие чего повышается точность её результатов.

Было использовано четыре разновидности кроссвалидации [16]:

1) KFold – классический метод кросс-валидации, выборка разделяется на k групп, при этом один объект может появиться в группах только один раз, соотношение классов не сохраняется.

2) StratifiedKFold – метод кросс-валидации, при котором выборка разделяется на k групп с сохранением соотношений классов.

3) ShuffleSplit – метод кросс-валидации, при котором выборка разделяется на k групп со случайными перестановками, т.е. один объект может появиться в группах несколько раз.

4) StratifiedShuffleSplit – метод, который представляет собой слияние StratifiedKFold и ShuffleSplit, т.е. выборка разделяется на k групп с сохранением соотношений классов и со случайными перестановками.

В результате по методу

– KFold качество модели – 0,83;

– StratifiedKFold – 0,83;

– ShuffleSplit – 0,83;

– StratifiedShuffleSplit – 0,84.

Все значения критерия AUC-ROC больше 80%, а значит, требования качества выполняются при использовании всех методов. Так как при StratifiedShuffleSplit достигается максимальная точность модели, итоговая модель будет содержать этот метод.

Далее были произведены анализ результатов и проверка выполнимости требований к модели:

– адекватность;

– адаптивность;

– эффективность;

– критерий AUC-ROC должен быть более 80%.

На тестовых данных было проведено прогнозирование благонадёжности контрагента при помощи построенной модели, при этом количество объектов (компаний) изменялось от 2 до 100. Выяснилось, что при малом количестве сравниваемых контрагентов (меньше 7) модель правильно классифицирует благонадёжность контрагента. При 17 и более сравниваемых контрагентов качество модели уже не соответствует требованиям качества по критерию AUC-ROC. Но в силу специфики проверки контрагента перед заключением договорных отношений (как правило, сравнивается до 10 компаний), можно считать, что модель соответствует данному требованию.

Также была проверена адекватность модели по средним значениям отклика модели и системы при помощи t-критерия Стьюдента [17] с уровнем значимости в 5% (табл. 5). На всех тестовых данных pvalue больше, чем 0,05, значит, гипотеза отвергается, средние значения отклика модели и системы равны, т.е. модель адекватна.

Таблица 5

Пример качества прогнозирования благонадёжности контрагента при помощи построенной модели

Количество объектов	2	3	4	5	6	7	8	9	10	11
AUC-ROC	1	1	1	1	1	1	0,83	0,83	0,87	0,87
Statistic	none	none	none	none	none	none	none	none	–1	–1
Pvalue	none	none	none	none	none	none	none	none	0,3434	0,34
Количество объектов	12	13	14	15	16	17	18	19	20	22
AUC-ROC	0,9	0,9	0,84	0,85	0,85	0,79	0,74	0,74	0,75	0,75
Statistic	–1	–1	0	0	0	–0,57	–1	–1	–1	–1
Pvalue	0,34	0,34	1	1	1	0,58	0,33	0,33	0,33	0,33

Адаптивность показывает возможность модели приспосабливаться к изменениям внешней среды и оригинала. Так как модель можно переобучать на новых данных, данное требование выполняется.

Эффективность отражает приемлемое время решения проблемы, объем занимаемой памяти и ресурсов. При моделировании были подобраны атрибуты модели так, чтобы время выполнения и качество модели были оптимальными.

Заключение

В результате экспериментов была разработана модель по оценке рисков управленческого решения по сотрудничеству с контрагентом, которая классифицирует контрагента как благонадёжный («+1») или неблагонадёжный («–1»). При этом модель дает точный результат при малом количестве сравниваемых контрагентов, что соответствует порядку проверки контрагента в реальной системе. Данная модель может быть использована при разработке политики информационной безопасности, в частности в вопросах, связанных с оценкой рисков надёжности контрагента.

Литература

1. Выписка из требований по безопасности информации, утвержденных приказом ФСТЭК России от 30 июля

2018 г. № 131 [Электронный ресурс]. – Режим доступа: https://fstec.ru /2050-vypiska-iz-trebovanij-po-bezopasnosti-informatsii-utverzhdennykh-prikazom-fstek-rossii-ot-30-iyulya-2018-g-n-131 (дата обращения: 03.04.2020).

2. Методика выявления уязвимостей и недекларированных возможностей в программном обеспечении [Электронный ресурс]. – Режим доступа: http://new.groteck.ru/images/catalog/70840/e75c72a254fcc880fa65657fdb144063.pdf (дата обращения: 15.04.2020).

3. СТО БР ИББС-1.4-2018. Обеспечение информационной безопасности организаций банковской системы Российской Федерации. Управление риском нарушения информационной безопасности при аутсорсинге [Электронный ресурс]. – Режим доступа: http://docs.cntd.ru/document/556983635 (дата обращения: 23.04.2020).

4. Миронова В.Г. Модель нарушителя безопасности конфиденциальной информации / В.Г. Миронова, А.А. Шелупанов // Информатика и системы управления. – 2012. – № 1(31). – С. 28–35.

5. Park C. Improvement of personal information protection level in the military using the measurement of disclosure risk / C. Park, C. Kim, Y. Kim, H. An, J. Bae // Journal of Security Engineering. – 2015. – Vol. 12, № 6. – P. 581–596.

6. Информационно-аналитическая система «СПАРК» при обеспечении экономической безопасности предприятия / А.С. Колтайс, А.А. Конев // Экономическая безопасность: финансовые, правовые и IT-аспекты: сб. матер. первой Всерос. науч.-практ. онлайн-конф.: Иркутск – Томск: Изд-во ТГУ, 2017. – С. 178–182.

7. Мероприятия по проверке контрагентов с помощью информационно-аналитических систем на основе существующих методик / А.С. Колтайс, А.А. Колтайс, А.А. Шатрова, Ю.А. Рубцова // Информационные технологии в науке, управлении, социальной сфере и медицине: сб. науч. трудов V Междунар. науч. конф. в 2 ч. – Томск: Изд-во ТПУ, 2018. – Ч. 2. – С. 418–422.

8. Automation of tax control mechanism with the use of specialized information and analytical systems within the framework of ensuring security / A. Koltays, A. Shatrova, A. Konev, P. Shelupanova // International Journal of Emerging Trends in Engineering Research – 2020. – Vol. 8, No. 4. – P. 1405–1409. – https://doi.org/10.30534/ijeter/2019/027122019

9. ИАС «СПАРК» [Электронный ресурс]. – Режим доступа: http://www.spark-interfax.ru/ (дата обращения: 30.04.2020).

10. Звонарев С.В. Основы математического моделирования: учеб. пособие. – Екатеринбург: Изд-во Урал. ун-та, 2019. –112 с

11. Гасников А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска: учеб. пособие. – 2-е изд., доп. – М.: МФТИ. – 2018. – 288 с.

12. sklearn.linear_model.SGDClassifier [Электронный ресурс]. – Режим доступа: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html/ (дата обращения: 07.05.2020).

13. Петри А. Наглядная медицинская статистика: учеб. пособие / А. Петри, К. Сэбин; пер. с англ. под ред. В.П. Леонова. – 3-е изд., перераб. и доп. – М.: ГЭОТАР-Медиа, 2015. – 216 с.

14. Файнзильберг Л.С. Гарантированная оценка эффективности диагностических тестов на основе усиленного ROC анализа // Управляющие системы и машины. – 2009. – № 5. – С. 3–13.

15. Fawcett T. An introduction to ROC analysis //Pattern Recognition Letters. – 2006. – Vol. 27. – P. 861–874.

16. Refaeilzadeh P. Cross Validation. Encyclopedia of Database Systems/ Refaeilzadeh P., Tang L., Liu H.// New York: Springer, 2009. – P. 532–538.

17. Scipy.stats.ttest_relvalidation [Электронный ресурс]. – Режим доступа: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_rel.html (дата обращения: 13.05.2020).

Колтайс Андрей Станиславович
Аспирант каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Ленина пр-т, д. 40, г. Томск, Россия, 634050
Тел: +7-923-403-40-93
Эл. почта: kas2@keva.tusur.ru

Шатрова Александра Александровна
Студентка каф. безопасности информационных систем (БИС) ТУСУРа
Ленина пр-т, 40, г. Томск, Россия, 634050
Тел.: +7 (382-2) 70-15-29
Эл. почта: shatrovaaleks@gmail.com

Шелупанов Александр Александрович
Д-р техн. наук, профессор, президент ТУСУРа
Ленина пр-т, 40, г. Томск, Россия, 634050
Тел.: +7 (382-2) 90-71-55
Эл. почта: saa@tusur.ru

Koltays A.S., Shatrova A.A., Shelupanov A.A.

Mathematical model for selecting a contractor to assess information security risks

The article deals with the main methods used to construct a mathematical model for assessing the welfare of the counterparty. The formats of data transformation are described in order to use them when modeling, as well as the requirements to the model. The dependence between input data was checked to exclude equal attributes from the model. The main difficulties in assessing the accuracy and completeness of the model were identified. The use of cross-validation to eliminate the difficulties in model building was described. The developed model gives an accurate result with a small number of compared counterparties, that corresponds to the checking order of the counterparty in the real system.

Keywords: model, dependability, risks, information-analytical systems, machine training.
doi: 10.21293/1818-0442-2020-23-2-36-41

References

1. *Vypiska iz Trebovanii po bezopasnosti infor-matsii, utverzhdennykh prikazom FSTEK Rossii ot 30 iyulya 2018 g. No* 131. Available at: https://fstec.ru /2050-vypiska-iz-trebovanij-po-bezopasnosti-informatsii-utverzhdennykh-prikazom-fstek-rossii-ot-30-iyulya-2018-g-n-131 (Accessed: April 3, 2020) (in Russ.).

2. *Metodika vyyavleniya uyazvimostei i nedeklariro-vannykh vozmozhnostei v programmnom obespechenii*. Available at: http://new.groteck.ru/images/catalog/70840/e75c72a254fcc880fa65657fdb144063.pdf (Accessed: April 15, 2020) (in Russ.).

3. STO BR IBBS-1.4-2018 *Obespechenie informatsionnoi bezopasnosti organizatsii bankovskoi sistemy Rossiiskoi Federatsii. Upravlenie riskom narusheniya informatsionnoi bezopasnosti pri autsorsinge*. Available at: http://docs.cntd.ru/document/556983635 (Accessed: April 23, 2020) (in Russ.).

4. Mironova V.G. Model' narushitelya bezopasnosti konfidentsial'noi informatsii / Mironova V.G., Shelupa-nov A.A. // Informatika i sistemy upravleniya. 2012, № 1(31), pp. 28–35.

5. Park C., Kim C., Kim Y., An H., Bae J. Improvement of personal information protection level in the military using the measurement of disclosure risk. Journal of Security Engineering, 2015, vol. 12, no. 6, pp. 581–596.

6. Koltais A.S., Konev A.A. *Informatsionno-analiticheskaya sistema «SPARK» pri obespechenii ekonomicheskoi bezopasnosti predpriyatiya*. [Ekonomicheskaya bezopasnost': finansovye, pravovye i IT-aspekty: sbornik materialov pervoi Vserossiiskoi nauchno-prakticheskoi onlain-konferentsii], 2017, Baikal'skii gosudarstvennyi universitet, pp. 178–182 (in Russ.).

7 Koltais A.S., Koltais A.A., Shatrova A.A., Rubtsova Yu.A. *Meropriyatiya po proverke kontragentov s pomoshch'yu informatsionno-analiticheskikh sistem na osnove sushchestvuyushchikh metodik.* [Informatsionnye tekhnologii v nauke, upravlenii, sotsial'noi sfere i meditsine: sbornik nauchnykh trudov V Mezhdunarodnoi nauchnoi konferentsii, Tomsk]: TPU, 2018, vol. 2, pp. 418–422 (in Russ.).

8. Koltays A. Shatrova A. Shelupanova P. Automation of tax control mechanism with the use of specialized information and analytical systems within the framework of ensuring security. International Journal of Emerging Trends in Engineering Research, 2020, vol. 8, no. 4, pp. 1405–1409. https://doi.org/10.30534/ijeter/2019/027122019

9. IAS «SPARK» Available at: http://www.spark-interfax.ru/ (Accessed: April 30, 2020) (in Russ.).

10. Zvonarev, S.V. *Osnovy matematicheskogo modelirovaniya: uchebnoe posobie*. [Basics of mathematical modeling] Ekaterinburg: Ural. un-t, 2019. 112 p. (in Russ.)

11. Gasnikov A.V. Sovremennye chislennye metody optimizatsii. Metod universal'nogo gradientnogo spuska: uchebnoe posobie [Modern numerical optimization methods. The universal gradient descent method: a training manual], 2 ed. M.: MFTI, 2018, 288 p. (in Russ.)

12. sklearn.linear_model.SGDClassifier. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html/ (Accessed: May 7, 2020).

13. Petri A., Sebin K. Naglyadnaya meditsinskaya statistika [Visual mathematical statistics]. Tutorial. Ed. V.P. Leonov. M., GEOTAR-Media Pub.l, 2015, 216 p. (in Russ.).

14. Fainzil'berg L.S. *Garantirovannaya otsenka effektivnosti diagnosticheskikh testov na osnove usilennogo ROC analiza*. [Upravlyayushchie sistemy i mashiny], 2009, no. 5, pp. 3–13 (in Russ.).

15. Fawcett T. An introduction to ROC analysis. [Pattern Recognition Letters], 2006, vol. 27, pp. 861–874.

16 Refaeilzadeh P., Tang L., Liu H Cross Validation. Encyclopedia of Database Systems. New York: Springer, 2009, pp. 532–538.

17. Scipy.stats.ttest_relvalidation. Available at: https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_rel.html (Accessed: May 13, 2020).

Andrey S. Koltays
Postgraduate student, Department of Complex Information Security of Electronic Computer Systems (KIBEVS),
Tomsk State University of Control Systems and Radioelectronics (TUSUR)
40, Lenin pr., Tomsk, Russia, 634050
Phone: +7 (382-2) 70-15-29
Email: kas2@keva.tusur.ru

Alexandra A. Shatrova
Student, Department of Information Systems Security (BIS)
40, Lenin pr., Tomsk, Russia, 634050
Phone: +7 (382-2) 70-15-29
Email: shatrovaaleks@gmail.com

Alexander A. Shelupanov
Doctor of Engineering Sciences, Professor,
President of TUSUR
40, Lenin pr., Tomsk, Russia, 634050
ORCID: 0000-0003-2393-6701
Phone: +7 (382-2) 90-71-55
Email: saa@tusur.ru