

УДК 659.1

<sup>1</sup>Н.И. Ломакин, <sup>2</sup>Е.М. Арсеньева, <sup>3</sup>В.В. Покидова, <sup>1</sup>Н.О. Могхарбел, <sup>4</sup>Н.В. Иевлева,  
<sup>1</sup>А.А. Сложеникина

<sup>1</sup>Волгоградский государственный технический университет, г. Волгоград, email: tel9033176642@yahoo.com

<sup>2</sup>Волгоградский филиал РЭУ им. Г.В. Плеханова, г. Волгоград, email: bn700@mail.ru

<sup>3</sup>Волгоградский филиал ФГБОУ ИВО МГГЭУ, Волгоград, email: pakidova@mail.ru

<sup>4</sup>Волжский политехнический институт (филиал) Волгоградский государственный технический университет, Волгоградская обл., г. Волжский, email: nievleva@gmail.com

## АРХИТЕКТУРА ИННОВАЦИОННОЙ AI-МОДЕЛИ ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННОГО РЯДА ФЬЮЧЕРСА SIU0 С ИСПОЛЬЗОВАНИЕМ ВЕКТОРИЗОВАННЫХ НОВОСТЕЙ С WEB-САЙТОВ В УСЛОВИЯХ ЦИФРОВИЗАЦИИ ЭКОНОМИКИ

**Ключевые слова:** AI-система, фьючерсный контракт, прогноз, большие данные, нейросеть.

Представлена разработанная инновационная AI-модель «Bot-эдвайзер», предназначенная для прогнозирования цены закрытия фьючерсного контракта SIU0 на московской бирже, включающая в себя программы Scrapy, нейросеть Word2vec, нейросеть Perseptron на платформе Deductor, а также торговый терминал QUIK с интегрированным Lua-сокетом. Исследованы теоретические основы прогнозирования временных рядов финансовых инструментов. Рассмотрен опыт применения систем искусственного интеллекта для сбора и обработки Big Data в целях прогнозирования временных рядов, в том числе фьючерсного контракта SIU0. Выдвинута и доказана гипотеза, что с помощью инновационной AI-модели, можно получить прогноз цены закрытия фьючерсного контракта SIU0 на 15-ти минутном таймфрейме. Предложена AI-система для повышения точности и снижения финансового риска при прогнозировании временного ряда фьючерсного контракта SIU0 для биржевой торговли. Авторами предложена инновационная AI-система, в которой для предсказания цены фьючерсного контракта SiU0. Инновационность предложенного подхода состоит в том, что для предсказания временного ряда использовались не только исторические данные – параметры японских свечей и объема, но и оцифрованные «новостные колебания» с web-сайтов. Кроме того, в целях минимизации ошибки предсказания нейросеть Perseptron, сформированная на платформе Deductor, была обучена на двух типах данных 1) стоимостном – (Pclose) и 2) логарифмированном – (ln). Предложенная инновационная AI-модель имеет большое практическое значение, поскольку она обеспечивает высокую точность прогноза. Так, средний размер ошибки нейросети в первом случае (Pclose) составил 0,000927425, тогда как при работе второй нейросети (Pln) средний размер ошибки составил -0,051026481. Дисперсия значений ошибки в процентном отношении к цене закрытия в первом случае была 0,304107913, тогда как во втором 0,343654316, или на 4 сотых лучше.

<sup>1</sup>N.I. Lomakin, <sup>2</sup>E.M. Arsenyeva, <sup>3</sup>V.V. Pokidova, <sup>1</sup>N.O. Mogharbel, <sup>4</sup>N.V. Ievleva,  
<sup>1</sup>A.A. Slozhenkina

<sup>1</sup>Volgograd State Technical University, Volgograd, email: tel9033176642@yahoo.com

<sup>2</sup>Volgograd branch of the PRUE G.V. Plekhanov, Volgograd, email: bn700@mail.ru

<sup>3</sup>Volgograd branch of FSBEI IVO MGGEU, Volgograd, email: pakidova@mail.ru

<sup>4</sup>Volzhsky Polytechnic Institute (branch) Volgograd State Technical University, Volgograd region, Volzhsky, email: nievleva@gmail.com

## ARCHITECTURE OF THE INNOVATIVE AI-MODEL FOR FORECASTING THE TIME SERIES OF SIU0 FUTURES USING VECTORIZED NEWS FROM WEBSITES IN THE CONDITIONS OF DIGITAL ECONOMY

**Keywords:** AI system, futures contract, forecast, big data, neural network.

The developed innovative AI-model designed to predict the closing price of a SIU0 futures contract on the Moscow stock exchange, which includes Scrapy programs, Word2vec neural network, Perseptron neural network on the Deductor platform, as well as the QUIK trading terminal with an integrated Lua socket, is presented. The theoretical foundations of forecasting the time series of financial instruments have been investigated. The experience of using artificial intelligence systems for collecting and processing Big Data in order to forecast time series, including the SIU0 futures contract, is considered. The hypothesis was put forward and proved that with the help of an innovative AI-model, it is possible to obtain a forecast of the closing price of the SIU0 futures contract on a 15-minute timeframe. An AI system is proposed to improve accuracy and reduce financial risk when forecasting the time series of a SIU0 futures contract for exchange trading. The authors have proposed an innovative AI system in which to predict the price of a futures contract SiU0. The innovation of the proposed approach lies in the fact that not only historical data – parameters of Japanese candlesticks and volume, but also digitized “news fluctuations” from websites were used to predict the time series. In addition, in order to minimize the prediction error, the Perseptron neural network, formed on the Deductor platform, was trained on two data types 1) cost – (Pclose) and 2) logarithmic – (ln). The proposed innovative AI-model is of great practical importance, since it provides a high forecast accuracy. So, the average size of the neural network error in the first case (Pclose) was 0.000927425, while when the second neural network (Pln) was operating, the average error size was -0.051026481. The variance of error values as a percentage of the closing price in the first case was 0.304107913, while in the second 0.343654316, or 4 hundredths better.

Актуальность исследования состоит в том, что выдвинута и доказана гипотеза, что с помощью разработанной AI-системы, с использованием «новостных колебаний» и параметров японских свечей, можно получить прогноз цены закрытия SiU0, причем точность прогноза зависит от варианта используемого датасета.

По оценке экспертов на сегодняшний день существует около ста методов и моделей прогнозирования. Прогнозирование временных рядов в условиях рыночной неопределенности имеет важное значение. Однако, несмотря на множество проведенных научных разработок, отдельные аспекты проблемы остаются недостаточно изученными и требуют дополнительных исследований.

В настоящей статье поднимается проблема, использования систем искусственного интеллекта, в целях прогнозирования параметров временного ряда биржевого инструмента. Решение поставленной проблемы имеет большую практическую значимость, Улучшение точности прогнозирования позволяет повысить эффективность биржевой торговли благодаря минимизации финансового риска.

Авторами предложена AI-система, в которой для предсказания цены фьючерсного контракта SiU0 использовались не только параметры японских свечей и объема, но и «новостные колебания» с web-сайтов. Нейросеть Perseptron, сформированная на платформе Deductor была обучена на двух типах данных 1) стоимостном (Pclose) и 2) логарифмированном– (ln).

### **Цель исследования**

Выдвинуть и доказать гипотезу, что с помощью нейросети, можно получить более точный прогноз цены закрытия фьючерсного контракта SIU0 на 15-ти минутном таймфрейме, поскольку предсказание цены временного ряда будет происходить не только на исторических данных, но с использованием оцифрованных в формате 300 – мерного вектора с «новостных колебаний»

### **Материал и методы исследования**

В работе применялись такие методы исследования, как: монографический,

аналитический, расчетно-конструктивный и нелинейные математические модели Word2Vec и Perseptron, а также анализ, моделирование, изучение и обобщение.

## **Результаты исследования и их обсуждение**

### **1.1 Система искусственного интеллекта Bot-эдвайзер на основе обработки Big Data**

Создание и продвижение области науки о данных и аналитики с точки зрения публикации теоретических основ, алгоритмов и моделей, оценки и экспериментов, приложений и систем по конкретным вопросам представлены в книге Data Analytics. Заслуживает внимания SpringerBriefs в области расширенной обработки информации и знаний, поскольку представляет собой краткое исследование в рассматриваемой области.

Платформа Nadoop представляет собой open-source фреймворк, который позволяет сделать так, чтобы приложение было разделено на несколько фрагментов. В результате каждый из полученных фрагментов обрабатывается на любом узле (компьютере, «ноде») в кластере сформированной вычислительной системы.

Как показывает практика, важное значение имеет повышение точности прогноза цены фьючерсного контракта SIU0, который торгуется на московской бирже с использованием Bot-эдвайзера, блок-схема которого представлена на рис. 1.

Bot-эдвайзер взаимодействует с ботом биржевым торговым терминалом QUIK посредством интегрируемого Lua-сокета. В процессе работы исторические данные временного ряда фьючерсного контракта SIU0 на 15-ти минутном таймфрейме импортируются в режиме онлайн и передаются для соединения с 300 – мерным вектором от Word2vec для формирования обучающей выборки для нейросети Perseptron, при этом, используются параметры: цена открытия – (Po), цена закрытия – (Pc), цена максимальная – (Ph), цена минимальная – (Pl), а также объем торгов (V) в двух форматах: 1) стоимостном (Pclose) и 2) логарифмированном– (ln).

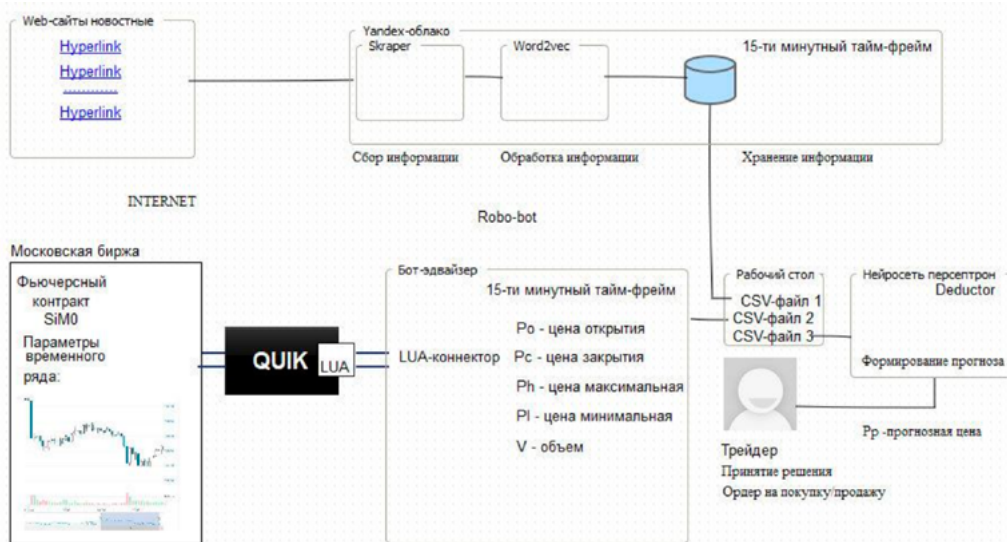


Рис. 1. Архитектура инновационной AI-модели «Bot-адвайзер»

### 1.2 Программа Skraпер для сбора и подсчёта слов с новостных Web-сайтов для подготовки VI

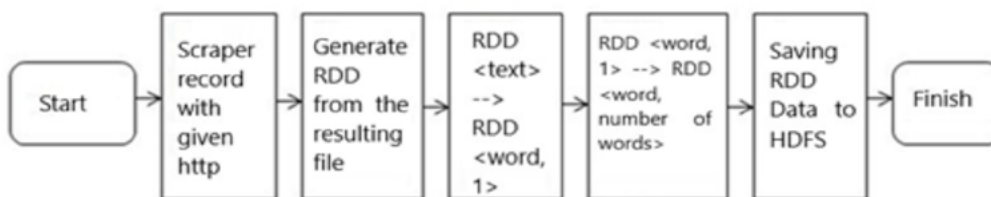


Рис. 2. Блок-схема программы Skraпер

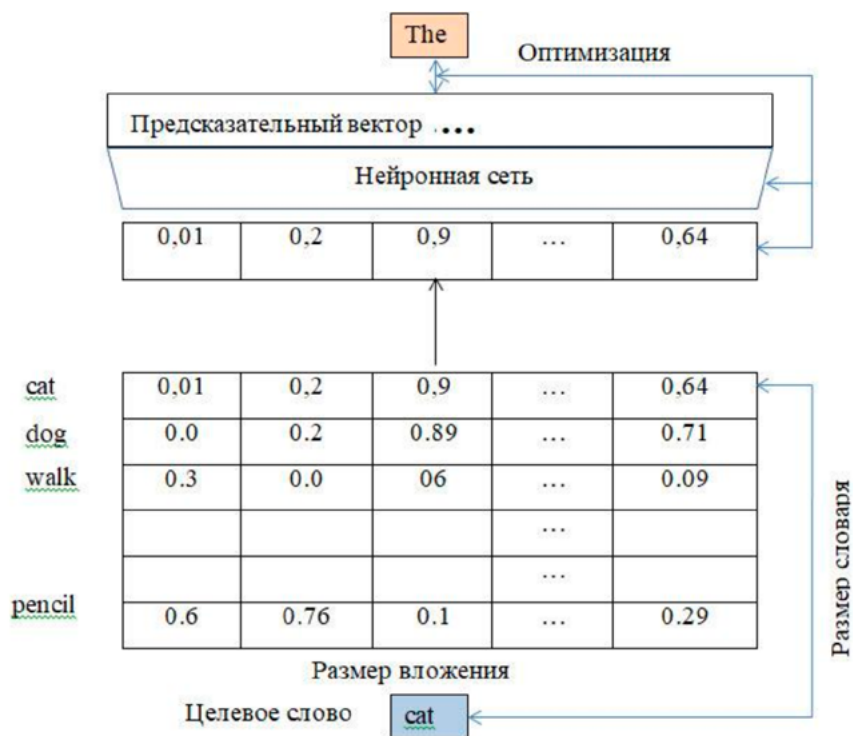


Рис. 3. Нейросеть Word2vec

Слова с новостных Web-сайтов были собраны слова с помощью программы Skraper, архитектура которого представлена ниже (рис. 2).

VI-данные, собранные программой Skraper были переданы и для дальнейшей обработки программой Word2vec для их последующей векторизации (рис. 3).

Word2vec – общее название для совокупности моделей на основе искусственных нейронных сетей, предназначенных для получения векторных представлений слов на естественном языке. Используется для анализа семантики естественных языков. Векторное представление – общее название для различных подходов к моделированию языка и обучению представлений в обработке естественного языка, направленных на сопоставление словам (и, возможно, фразам) из некоторого словаря векторов из для значительно меньшего количества слов в словаре.

Теоретической базой для векторных представлений является дистрибутивная семантика [1].

Из одного контекстного окна получают два элемента данных (на одно целевое слово приходится два соседних). Размер окна обычно определяется пользователем. Чем больше размер контекстного окна, тем лучше наша модель, но это влияет на время выполнения алгоритма. Эта модель известна как «skip-gram» алгоритм, это один из алгоритмов word2vec. Другой алгоритм известен как «непрерывный мешок со словами» (continuous bag-of-words model, CBOW).

### **1.3 Нейросеть Perseptron для прогноза цены закрытия цены фьючерсного контракта SIU0 на следующие 15 минут на основе VI с новостных Web-сайтов и данных временного ряда**

Программа Perseptron для прогноза цены закрытия цены фьючерсного контракта SIU0 сформирована на платформе Deductor. Deductor – аналитическая платформа, разработанная компанией BaseGroup Labs. В Deductor встроены самые востребованные алгоритмы анализа (деревья решений, нейронные сети,

самоорганизующиеся карты и т. д.), есть десятки способов визуализации и предусмотрена интеграция с множеством источников/приемников данных.

Нейросеть Perseptron, сформированная на платформе Deductor была обучена на двух типах данных 1) стоимостном (Pclose) и 2) логарифмированном – (ln).

Первый тип «Стоимостной (Pclose)» представлял собой 300 – мерный вектор от Word2vec и параметры: цена открытия – (Po), цена закрытия – (Pc), цена максимальная – (Ph), цена минимальная – (Pl), а также объем торгов (V). Прогнозное значение затем сравнивалось с фактической ценой закрытия, если прогнозное значение оказывалось больше, то открывалась длинная позиция и наоборот.

Второй тип – «Логарифмированный – (ln)» представлял собой 300 – мерный вектор от Word2vec и параметры: натуральный логарифм отношения цен P/Pt-1: открытия – (Po), цена закрытия – (Pc), цена максимальная – (Ph), цена минимальная – (Pl) ln отношения объемов торгов (V). Над прогнозным значением проводилось потенцирование, то есть  $P_{progn} = e^{P_{progn}}$ . Затем прогнозное значение сравнивалось с фактической ценой закрытия, если прогнозное значение оказывалось больше, то открывалась длинная позиция и наоборот.

Первый вариант Perseptron обучался на датасете, который включал 305 – мерные векторы и прогнозные значения цены закрытия Pt+1, датасет содержал данные с 14 апреля 2020 г. по 04 июля 2020 г. и насчитывал 1930 строк (таблица 1).

Разработанный персеptron содержит 305 параметров на входном слое, два скрытых слоя по 100 и 10 узлов, соответственно, и выходной слой с одним параметром – прогнозной ценой. Персеptron был сформирован и использовался на платформе Deductor. Граф нейросети представлен на рис. 4.

Точность прогноза и экономическая результативность работы нейросети были исследованы на данных, не вошедших ни в обучающую, ни в тестовую выборки от 03 июля 2020 г. с 10.00 до 15.30.

Графики, отражающие точность прогноза по обоим вариантам представлены ниже (рис. 5).

Таблица 1

Фрагмент обучающей выборки первого варианта Perseptron

Date, time	X1	X2	...	X300
14.04.2020 22.00.00	-0,062179142	0,020744415	...	0,033938743
14.04.2020 21.45.00	0,059611453	-0,062379819	...	0,034278696
...	...	...	...	...
04.06.2020 16.15.00	0,058385131	-0,059321518	...	0,037604395

...	Po	Pc	Ph	Pl	Volume	Prognosis(Pt+1)
...	71065	71107	71114	71062	6721	71140
...	71065	71067	71080	71035	12248	71107
...	...	...	...	...	...	...
...	74849	74717	74861	74715	28285	74878

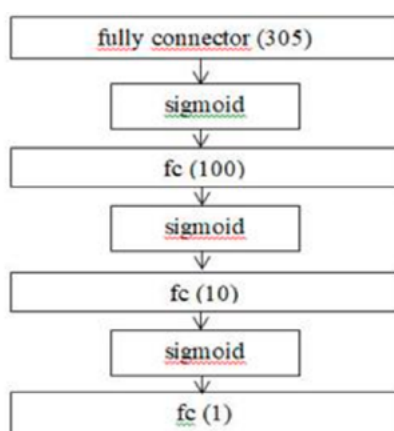


Рис. 4. Граф нейросети персептрон

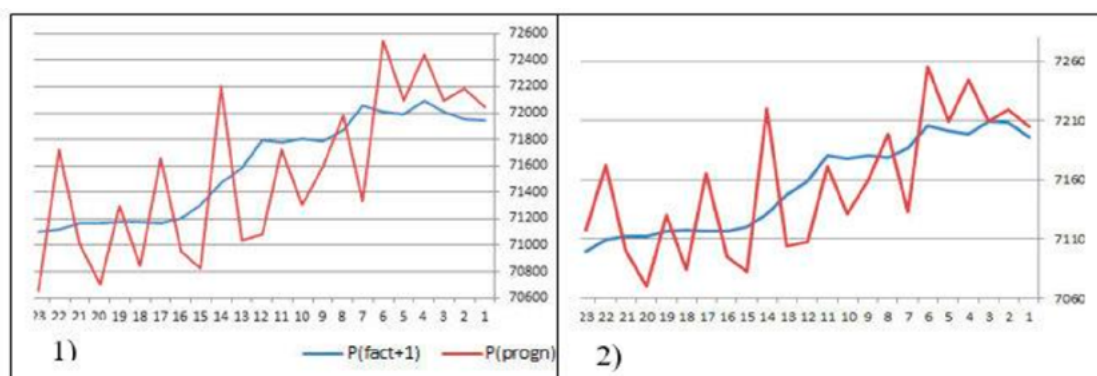


Рис. 5. Кривые «факт-прогноз» по двум вариантам

Анализ результатов эксперимента показал, что все рассматриваемые параметры в первом варианте оказались лучше в сравнении со вторым вариантом. Так, средний размер ошибки нейросети в первом случае ( $P_{close}$ ) составил 0,000927425, тогда как при работе второй нейросети ( $P_{ln}$ ) средний размер

ошибки составил -0,051026481. Дисперсия значений ошибки в процентном отношении к цене закрытия в первом случае была 0,304107913, тогда как во втором 0,343654316, или на 4 сотых лучше.

Сигма значений ошибки нейросети в первом случае ( $P_{close}$ ) составила 0,551459802, тогда как при работе вто-

рой нейросети (PIn) 0,586220365. Что говорит о том, что в первом случае разброс значений относительно среднего значения меньше.

Как известно, временной ряд – это ряд последовательных значений, характеризующих изменение некоторого показателя во времени. Анализ временных рядов представляет собой совокупность математико-статистических методов анализа, направленных на выявление структуры временных рядов и для их прогнозирования.

Математические модели временных рядов могут иметь различные формы и представлять различные стохастические процессы. Можно выделить три широких класса моделей, в которых последующие данные линейно зависят от предшествующих: авторегрессионные модели, интегральные модели, модели скользящего среднего.

На их основе строятся модели авторегрессионного скользящего среднего (Autoregressive Moving Average, ARMA) и модель авторегрессии и проинтегрированного скользящего среднего (Autoregressive Integrated Moving Average, ARIMA).

Среди нелинейных моделей временных рядов выделяют такие, как: GARCH, TAR, EGARCH, FIGARCH, CGARCH и др.

Отечественный опыт свидетельствует о том, что системы искусственного интеллекта находят все более широкое применение в решении практических задач на крупных предприятиях, в банках и IT-компаниях. Следует констатировать, что в России сложился довольно низкий уровень применения цифровых инноваций, по сравнению с зарубежными компаниями развитых стран и продолжает оставаться на низком уровне. В целях кардинального изменения ситуации в мае 2017 в России были изданы соответствующие нормативные документы, целью которых является стимулирование организаций к активной цифровизации всех сфер и функциональных зон своей деятельности. Важной проблемой остается вопрос оценки и снижения финансового риска.

Научный интерес представляют исследование Рупперта по статистике и анализу данных для финансового инжиниринга [2] и модель ценообразования капитальных активов Дженсена, Фишера

и Майрона [3]. Фама и Макбет сочли необходимым рассматривать риск как категорию, в которой наблюдается доходность и равновесие [4]. Фразинни и Педерсен приняли во внимание роль бета-портфеля финансовых инструментов [5].

Практика показывает, что применение систем искусственного интеллекта позволяет решать широкий круг проблем. Например, важное значение имеет поиск оптимальных решений в области создания организационных механизмов выделения субсидий на общественный транспорт в Санкт-Петербурге [6, с. 4706-4711], для оптимизации управления инновационным процессом на производственном предприятии [7].

Как показывает практика, искусственный интеллект находит все более широкое применение в условиях цифровизации экономики, для которых характерно внедрение технологий «Индустрия 4.0». Внимание многих ученых сфокусировано на исследовании технологических процессов, обусловленных внедрением цифровой экономики.

### Выводы

Предложенное решение имеет большое практическое значение, поскольку разработанная AI-система обеспечивает высокую точность прогноза. Так, средний размер ошибки нейросети в первом случае ( $P_{close}$ ) составил 0,000927425, тогда как при работе второй нейросети (PIn) средний размер ошибки составил -0,051026481. Дисперсия значений ошибки в процентном отношении к цене закрытия в первом случае была 0,304107913, тогда как во втором 0,343654316, или на 4 сотых лучше.

Сигма значений ошибки нейросети в первом случае ( $P_{close}$ ) составила 0,551459802, тогда как при работе второй нейросети (PIn) 0,586220365. Что говорит о том, что в первом случае разброс значений относительно среднего значения меньше.

Первый вариант оказался более результативным по размеру полученной прибыли (вариационной маржи). В первом случае прибыль составила 748,0 руб., тогда как во втором величина прибыли была меньше, лишь 548,0 руб. Максимальная просадка, в обоих вариантах оказалась одинаковой -169 руб.

*Библиографический список*

- 1 Word2Vec: как работать с векторными представлениями слов [Электронный ресурс] – URL: <https://neurohive.io/ru/osnovy-data-science/word2vec-vektornye-predstavlenija-slov-dlja-mashinnogo-obucheniya/> (дата обращения 2 мая 2020 г.)
- 2 Ruppert D. Statistics and data analysis for financial engineering // Springer. 2019.
- 3 Jensen M., Fischer V. and Myron V. The Capital Asset Pricing Model: some empirical tests // Praeger Publishers Inc. 1972.
- 4 Eugene F. Fama and James D. MacBeth Risk, Return and Equilibrium: Empirical Tests // Journal of Political Economy, 1973. Vol. 81(3) 5 Frazzini A. and Pedersen L.H. Betting Against Beta. // NBER Working Paper. 2010.
6. Kulachinskaya A., Kravchenko V., Bezdenezhnykh T. Organizational mechanisms of allocation of subsidies for public transport in St. Petersburg // Proceedings of the 31st International Business Information Management Association Conference. – 2018. – С. 4706-4711
7. Demidenko D.S. Malevskaia-Malevich E.D., Dubolazova Y.A., Victorova N.G. Optimization of the innovation process management at a manufacturing enterprise // Proceedings of the 31st International Business Information Management Association Conference. – 2018. – С. 996-1003.