

УДК 303.733.4

**Р.Н. Яковлев**, мл. научный сотрудник, Санкт-Петербургский институт информатики и автоматизации Российской Академии наук (Россия, 199178, Санкт-Петербург, 14 линия В.О., 39) (e-mail: iakovlev.r@mail.ru)

## **ПРИМЕНЕНИЕ СРЕДСТВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДЛЯ РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ ДЕЯТЕЛЬНОСТИ СКЛАДА**

*Данная работа посвящена актуальной на сегодняшний день проблематике – оптимизации складского хозяйства. Непосредственно в работе рассматривался процесс приемки и отправки грузов на некотором складском помещении среднего масштаба. В рамках исследования, по результатам проведения предварительного анализа, были выбраны показатели, необходимые для применения средств интеллектуального анализа с целью прогнозирования количества сотрудников на складских подъездах. В соответствии с поставленной целью были сформированы модели по прогнозированию необходимого числа сотрудников на подъездах для обеспечения работы последних в режиме, соответствующем оптимальному значению показателя «загруженность подъезда». Для решения данной задачи рассмотрено применение различных методов машинного обучения, таких как дерево решений, регрессор на основе метода ближайших соседей, случайный лес, нейронная сеть прямого распространения. Каждая из задействованных моделей проходила обучение с различными значениями гиперпараметров модели, которые выбирались как в ручном режиме на основе эвристик, так и с использованием специализированных программных средств перебора по сетке GridSearchCV из библиотеки scikitlearn, предназначенных для поиска оптимальных значений гиперпараметров. С использованием автоматизации поиска гиперпараметров при обучении моделей было достигнуто меньшее значение среднеквадратической ошибки по сравнению с ручным выбором гиперпараметров. По результатам проведенного анализа качества прогнозов моделей было установлено, что спрогнозированные значения числа сотрудников в большей степени соответствуют реальной ситуации нежели плановые значения, используемые компанией на данный момент. На основе анализа полученных результатов были составлены рекомендации для оценки роста экономической эффективности предприятия.*

**Ключевые слова:** средства интеллектуального анализа; складское хозяйство; экономическая эффективность; управление персоналом.

**DOI:** 10.21869/2223-1560-2018-22-6-127-135

**Ссылка для цитирования:** Яковлев Р.Н. Применение средств интеллектуального анализа для решения задач оптимизации деятельности склада // Известия Юго-Западного государственного университета. 2018. Т. 22, № 6(81). С. 127-135.

\*\*\*

### **Введение**

В рамках данной работы рассматривается проблема управления и оптимизации складского хозяйства. На протяжении многих лет оптимизация складского хозяйства не теряет своей актуальности [1, 2]. Множеством исследователей для решения задач в рамках данной проблематики применялись различные группы подходов и методов, в числе которых: игровые методы, применение которых рассмотрено в работе [3], логистический подход к оптимизации складской грузопереработки, представленный, в частности, в [4], а также подходы, связанные с

применением имитационных моделей, задействованные в [5].

Целью данной работы является формирование решений по оптимизации процессов приемки и отправки грузов с применением средств интеллектуального анализа данных. В качестве исходных данных были взяты сведения о складе, который относится к складским помещениям класса В+ (по классификации Knight Frank и Swiss Realty Group) [6]. Анализируемый склад имеет шесть подъездов для загрузки и разгрузки грузов. При этом подъезды отличаются уровнем технологического оснащения, что позво-

ляет принимать к погрузке/разгрузке только определенные виды продукции.

### **Постановка задачи**

При первичном анализе складского хозяйства было выявлено, что одним из ключевых факторов, влияющих на скорость погрузки/разгрузки, является число сотрудников на подъезде, также скорость разгрузки зависит и от физических свойств продукции. На данный момент в соответствии с внутренними инструкциями организации все сотрудники работают по сменам (длительность смены 4 часа), а их число на конкретном подъезде определяется в соответствии с планом. Данный план составляется на основе величины среднемесячной загруженности. На данный момент в компании применяется подход по ликвидации очередей, который имеет определенные недостатки, в частности, из-за отсроченной реакции на возникновение очередей происходит рост издержек:

- связанных с временем ожидания и временем проведения погрузок/разгрузок;
- связанных с необходимостью в содержании потенциально избыточного персонала;
- связанных с необходимостью по внеплановому вызову дополнительных сотрудников (по более высокой ставке на оплату труда работника).

Исходя из результатов первичного анализа цель работы была детализована следующим образом:

1. На основе исторических данных, сформировать модели по прогнозированию необходимого числа сотрудников на подъездах для обеспечения работы последних в режиме, соответствующем оптимальному значению показателя «загруженность подъезда».

2. Обозначить возможности по оценке потенциального экономического эффекта от применения полученных моделей.

### **Исследование загруженности подъезда. Построение и обучение моделей**

Для достижения поставленных в работе целей был проведен предварительный анализ статистики по подъезду с целью выбора показателей, которые будут использованы для прогнозирования оптимального числа сотрудников на подъезде. На основании предварительного анализа был сформирован итоговый список параметров для дальнейшего построения моделей по прогнозированию оптимального числа сотрудников на подъезде. В него вошли следующие показатели: число поступивших за смену машин; вес поступившего за смену груза; загруженность подъезда (внутренний показатель, используемый компанией для оценки интенсивности рабочего процесса, а также при принятии решений относительно изменения числа работников на подъезде, определяется по окончанию смены начальником смены); календарный месяц; порядковый номер смены; плановое число сотрудников. Выбор параметров обосновывается тем, что их либо можно определить заранее с высокой точностью, либо у них имеется целевое значение, в частности, для показателя «загруженность подъезда» целевым значением является «оптимальная нагрузка». В рамках рассматриваемых задач, отдельные подъезды являются независимыми единицами, а потому дальнейшее исследование и прогнозирование проводилось только для одного из подъездов.

Прогнозирование осуществлялось с использованием 4 различных типов моделей:

- дерево решений;
- регрессор на основе метода ближайших соседей;
- случайный лес;
- нейронная сеть прямого распространения.

Каждая из представленных моделей базируется на различных принципах и широко используется в области интеллектуального анализа данных в задачах прогнозирования. Несмотря на то, что первые две модели являются более простыми, они, тем не менее, не теряют своей актуальности и часто демонстрируют весьма высокие показатели качества работы [7–9]. Две другие используемые модели (случайный лес [10] и нейронная сеть прямого распространения) являются более сложными и трудно интерпретируемыми для пользователя, и преимущественно используются как «черный ящик» при формировании прогнозов, однако данные модели могут достигать более высоких показателей качества работы [11–14], чем и обусловлено их применение в данном исследовании.

В данной работе для каждой из первых трех моделей была проведена следующая последовательность действий:

- Исходная выборка делилась на обучающую и тестовую в соотношении 75 к 25.

- Модель обучалась на обучающей выборке с произвольно выбранными параметрами, после чего проводилась ее проверка на тестовой выборке, качество модели оценивалось посредством определения значения коэффициента детерминации между спрогнозированными и реальными значениями.

- Далее с использованием возможностей библиотеки `scikit-learn` [15] проводилась подборка наилучших параметров для модели (с использованием кросс-валидации) и обучение модели с нуля с уже оптимальными параметрами. Качество такой модели также оценивалось с помощью коэффициента детерминации.

Поскольку целевой параметр является целочисленным, вычисление коэффициента детерминации также проводилось и после округления прогнозных значений «фактического числа сотрудников на подъезде».

В рамках четвертого подхода с использованием программного средства `matlab` была создана и обучена нейронная сеть прямого распространения со следующей архитектурой (рис. 1).

Таблица 1

Результаты обучения и проверки качества модели №1 (дерево решений)

Формат результатов	Максимальная глубина дерева	Число используемых признаков	Коэффициент детерминации ( $R^2$ )
Дерево решений с параметрами, выбранными вручную			
Исходные результаты (вещественные числа)	5	-	0.72
Результаты после округления (натуральные числа)	5	-	0.65
Дерево решений с оптимальными параметрами			
Исходные результаты (вещественные числа)	7	4	0.73
Результаты после округления (натуральные числа)	7	4	0.66

Таблица 2

Результаты обучения и проверки качества модели №2 (метод ближайших соседей)

Формат результатов	Число рассматриваемых соседей	Коэффициент детерминации ( $R^2$ )
Метод ближайших соседей с параметрами, выбранными вручную		
Исходные результаты (вещественные числа)	10	0.61
Результаты после округления (натуральные числа)	10	0.55
Метод ближайших соседей с оптимальными параметрами		
Исходные результаты (вещественные числа)	9	0.73
Результаты после округления (натуральные числа)	9	0.66

Таблица 3

Результаты обучения и проверки качества модели №3 (случайный лес)

Формат результатов	Максимальная глубина деревьев	Число используемых признаков	Коэффициент детерминации ( $R^2$ )
Случайный лес с параметрами, выбранными вручную			
Исходные результаты (вещественные числа)	7	-	0.73
Результаты после округления (натуральные числа)	7	-	0.67
Случайный лес решений с оптимальными параметрами			
Исходные результаты (вещественные числа)	10	3	0.76
Результаты после округления (натуральные числа)	10	3	0.69

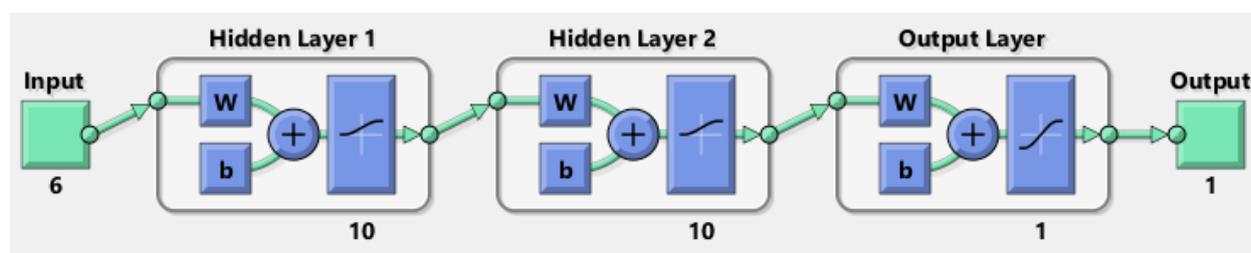


Рис. 1. Архитектура нейронной сети прямого распространения

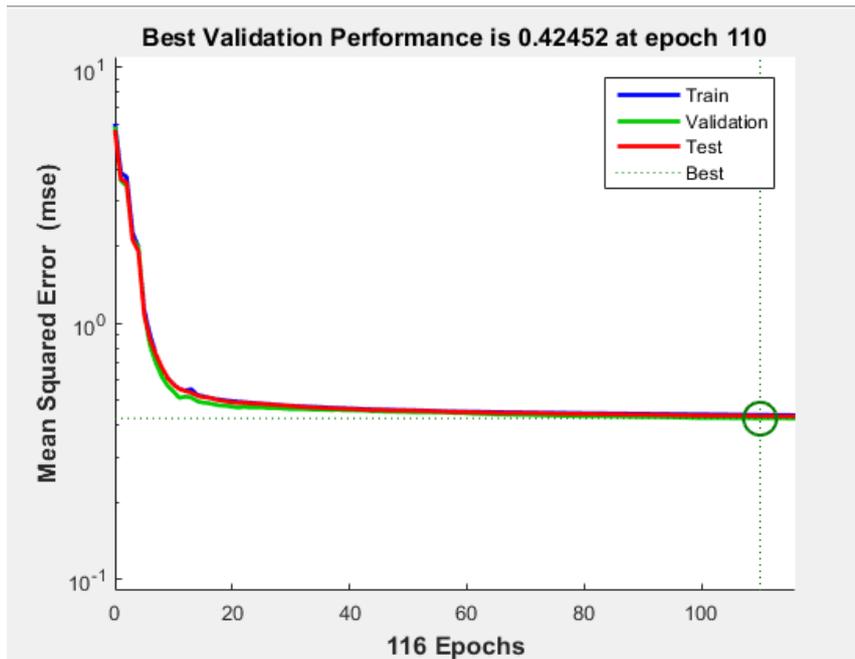


Рис. 2. Величина среднеквадратической ошибки в процессе обучения модели

Вычисленные значения показателей качества моделей ( $R^2$  и среднеквадратической ошибки) позволяют нам сделать вывод о состоятельности построенных моделей, а соответственно и об их пригодности к дальнейшему прогнозированию. Стоит отметить, что при проверке на тестовой выборке показатели качества работы дерева решений, регрессора на основе метода ближайших соседей и случайного леса существенно возросли после обучения данных моделей с оптимальными параметрами (рис. 2).

**Прогнозирование оптимального числа сотрудников**

После завершения процессов построения и обучения моделей было осуществлено непосредственное прогнозирование значений числа сотрудников на подъезде. Прогнозы формировались на период с 01.10.2018 по 31.10.2018 включительно, соответственно значения за данный период не входили в выборку, на основе которой проводилось обучение моделей. Данный период был выбран поскольку в та-

ком случае появляется возможность сравнить спрогнозированные значения величин с их реальными значениями.

Прогнозное число сотрудников может быть принято в качестве оптимального с высокой долей вероятности, поскольку на весь прогнозируемый период показателю «загруженность подъезда» было присвоено оптимальное значение.

Полученные результаты прогнозирования представлены на рисунке 3 ниже.

На основе спрогнозированных и реальных значений показателя «число сотрудников на подъезде» по каждой модели были определены величины среднеквадратической ошибки MSE полученных прогнозов.

Кроме того, в целях формирования более обоснованного заключения относительно целесообразности использования разработанных моделей, дополнительно был построен график загруженности подъезда на основе реальных данных и данных, использованных при прогнозировании числа сотрудников на подъезде, представленный на рисунке 4.

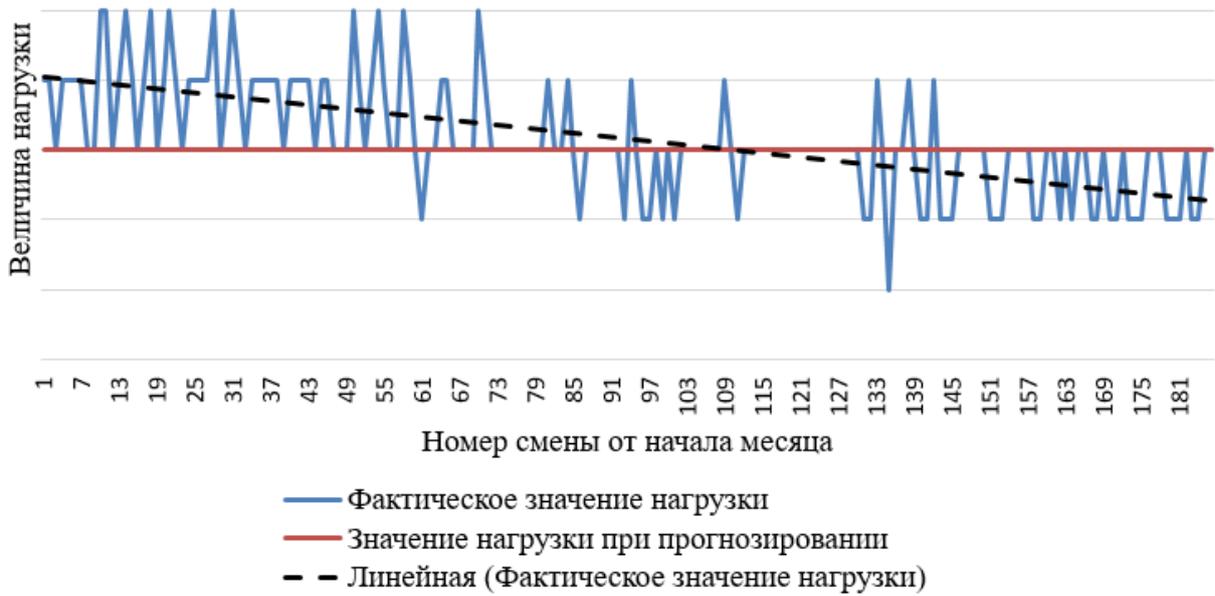


Рис. 3. Результаты прогнозирования фактического числа сотрудников

Таблица 4

Величины среднеквадратических ошибок прогнозирования и планирования

	Дерево решений	Метод ближайших соседей	Случайный лес	Нейронная сеть	Плановые значения
MSE	0,28	0,32	0,31	0,49	0,5

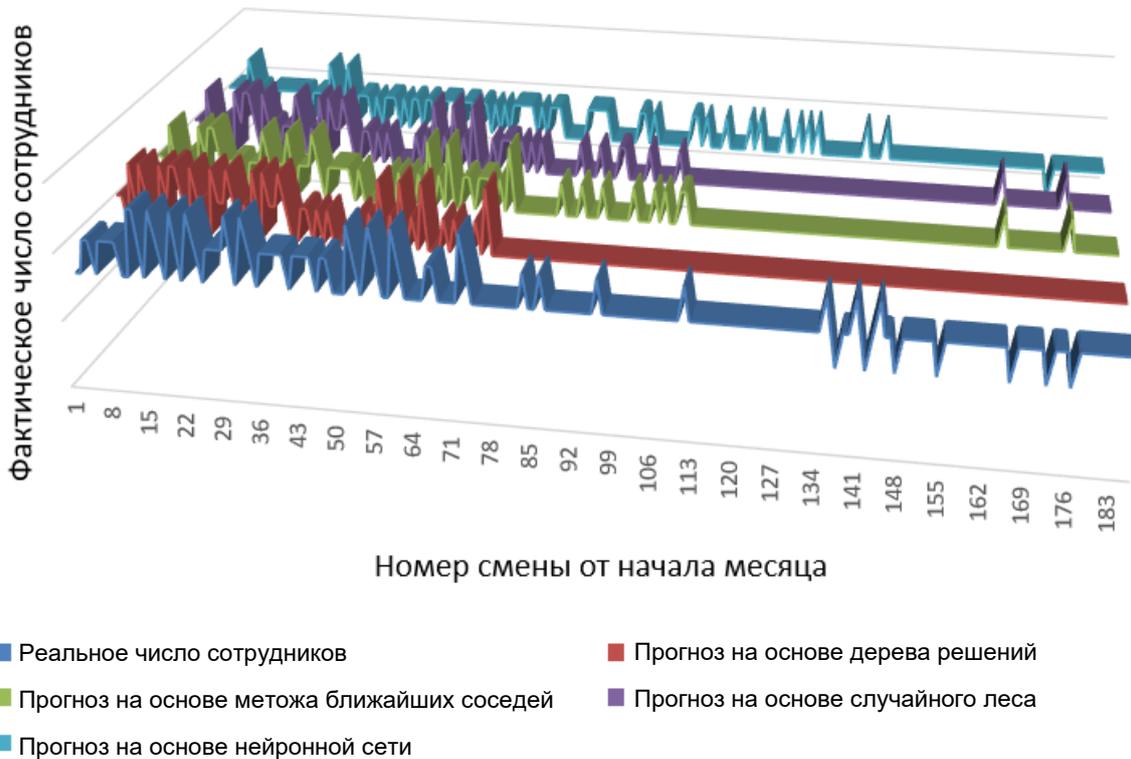


Рис. 4. Величина загруженности за рассматриваемый период

Анализируя графики, представленные на рисунках 3 и 4, можно заметить, что в первой половине рассматриваемого периода фактическое число сотрудников на подъезде зачастую было существенно больше планового. Тем не менее, исходя из значений фактической загруженности подъезда на данном временном промежутке можно сделать вывод о том, что число сотрудников в определенные смены оказывалось недостаточным. Следовательно, можно предположить, что разумное превышение прогнозного числа сотрудников над фактическим числом сотрудников на подъезде на данном временном отрезке является свидетельством более высокого качества модели. Для второй половины рассматриваемого периода ситуация иная, фактическое число сотрудников преимущественно равнялось плановому. Стоит отметить, что исходя из графика загруженности подъезда можно сделать вывод о том, что фактическое число сотрудников зачастую было даже несколько избыточным. Таким образом, превышение прогнозного значения фактическим на данном временном промежутке можно трактовать как свидетельство более низкого качества модели.

Необходимо также отметить, что в течение всего прогнозного периода наблюдаются тренды к уменьшению загруженности подъезда и к уменьшению числа сотрудников, работающих на подъезде. При этом тренд к уменьшению загруженности подъезда сохраняется, даже несмотря на постепенное снижение числа сотрудников.

Исходя из графиков, представленных на рис. 3, можно утверждать, что результаты, спрогнозированные для каждой из разработанных моделей, значительно лучше соответствуют описанным выше трендам и зависимостям в сравнении со стабильными плановыми значениями. Принимая во внимание рассчитанные ранее значения среднеквадратической

ошибки, можно заключить, что использование разработанных моделей по прогнозированию числа сотрудников на подъезде является целесообразным.

### **Выводы**

Основываясь на результатах, полученных в ходе интеллектуального анализа, представляется возможным в дальнейшем оценить изменения в фонде рабочего времени сотрудников. В том случае, если фонд рабочего времени снизился или же не изменился, но эффективность работы при этом увеличилась, то предложенное решение по определению оптимального числа сотрудников на подъезде улучшило работу предприятия. В качестве параметра, определяющего экономическую эффективность работы подъезда, предприятию предлагается использовать параметр выработки, характеризующий производительность труда.

Отметим также, что использование разработанных прогнозных моделей потенциально позволит снизить очереди на погрузку и разгрузку машин, что может стать конкурентным преимуществом, т.к. логистические компании смогут снизить свои издержки, связанные с доставкой грузов: зарплата водителя-экспедитора; бензина; простой машины, как следствие задержка исполнения следующего заказа.

### **Благодарности**

*Исследование выполнено в рамках бюджетной темы (№АААА-А16-116033110095-0).*

### **Список литературы**

1. Ложечник Е.А. Оптимизация складского комплекса предприятия на основе рационализации и автоматизации основных процессов // Транспортное дело России. 2010. № 3.
2. Karásek J. An overview of warehouse optimization // International journal of advances in telecommunications, electro-

technics, signals and systems. 2013. Vol. 2.No. 3. P. 111–117.

3. Оптимизация складского хозяйства предприятий автомобильного транспорта использованием игровых методов / Е.В. Агеева, С.В. Пикалов, И.П. Емельянов, Е.В. Агеев // Известия Юго-Западного государственного университета. Серия: Техника и технологии. 2015. №. 1. С. 21–28.

4. Андрианов А.Ю., Киселева О.М. Использование логистических подходов в управлении складом временного хранения // Вестник университета. 2016. №. 1.

5. Kofjač D., Kljajić M., Rejec V. The anticipative concept in warehouse optimization using simulation in an uncertain environment // European Journal of Operational Research. 2009. Vol. 193. No. 3. P. 660–669.

6. Классификация складских помещений от агентства Swiss Realty Group. URL: <http://www.galaxylogistics.ru/slovar-logistiki/k-ru/klassifikatsiya-skladskih-pomescheniy-ot-agentstva-swiss-realty-group.html> (дата обращения 15.10.2018).

7. Построение деревьев решений и извлечение правил из обученных нейронных сетей / В.Н. Гридин, В.И. Солодовников, И.А. Евдокимов, С.В. Филиппков // Искусственный интеллект и принятие решений. 2013. №. 4. С. 26.

8. Воронцов К.В. Математические методы обучения по прецедентам (теория обучения машин). М., 2011. С. 119–121.

9. An empirical comparison of machine learning models for time series forecasting / N.K. Ahmed, A.F. Atiya, N.E. Gayar, H. El-Shishiny // Econometric Reviews. 2010. Vol. 29. No. 5-6. P. 594–621.

10. Чистяков С.П. Случайные леса: обзор // Труды Карельского научного центра Российской академии наук. 2013. №. 1.

11. Liaw A., Wiener M. Classification and regression by randomForest // R news. 2002. Vol. 2. No. 3. P. 18–22.

12. Arulampalam G., Bouzerdoum A. A generalized feedforward neural network architecture for classification and regression // Neural networks. 2003. Vol. 16. №. 5-6. P. 561–568.

13. Бобырь М.В. Метод нелинейного обучения нейро-нечеткой системы вывода // Искусственный интеллект и принятие решений. 2018. № 1. С. 67–75.

14. Бобырь М.В. Проектирование нейронных и нечетких моделей в области вычислительной техники и систем управления. М.: Аргамак-Медиа, 2018. 110 с.

15. Scikit-learn. Machine Learning in Python. URL: <https://scikit-learn.org/stable/documentation.html> (дата обращения 15.10.2018).

*Поступила в редакцию 16.11.2018*

UDC 303.733.4

**R.N. Yakovlev**, Junior Researcher, St. Petersburg Institute for Informatics and Automation of the Russian Academy Of Sciences (Russia, 199178, St. Petersburg, 14 line V. O., 39) (e-mail: iakovlev.r@mail.ru)

#### **APPLICATION OF INTELLECTUAL ANALYSIS TOOLS FOR SOLVING WAREHOUSE OPERATION OPTIMIZATION PROBLEMS**

*This work is devoted to the currently relevant issues - optimization of warehousing. Cargo receiving and dispatch processes held in a medium-sized warehouse are being considered in this work in detail. Based on the results of preliminary analysis, several indicators were selected, necessary for the usage of intellectual analysis tools in order to predict the required number of employees at the warehouse entrances. In accordance with the purpose set, models for predicting the required number of employees at the entrances were created to ensure the working process in a way, which corresponds to the optimal value of the indicator "workload of the entrance". Various methods of machine learning, such as decision tree, k-nearest neighbors regression, random forest, and feedforward neural network are considered in the context of the problem, mentioned above. Each of the models was trained with*

different values of the model's hyperparameters, which were selected in both manual heuristic-based mode and using specialized software tools for grid search (GridSearchCV) from the scikit-learn library, designed to find the optimal values of the hyperparameters. Using the automated search for hyperparameters when training models yields to a smaller mean-square error in comparison with manual selection of hyperparameters. According to the analysis results of the model prediction quality, it was found that the predicted number of employees closely corresponds to the real situation in comparison with the planned values being used by the company. Based on the obtained results, several recommendations were made to assess the growth of the economic efficiency of the enterprise.

**Key words:** data mining tools; warehousing; economic efficiency; Human Resource management.

**DOI:** 10.21869/2223-1560-2018-22-6-127-135

**For citation:** Yakovlev R.N. Application of Intellectual Analysis Tools for Solving Warehouse Operation Optimization Problems. Proceedings of the Southwest State University, 2018, vol. 22, no. 6(81), pp. 127-135 (in Russ.).

\*\*\*

## References

1. Lozhechnik E.A. Optimizacija sklad-skogo kompleksa predpriyatija na osnove racionalizacii i avtomatizacii osnovnyh processov. *Transportnoe delo Rossii*, 2010, no. 3.
2. Karásek J. An overview of warehouse optimization. *International journal of advances in telecommunications, electro-technics, signals and systems*, 2013, vol. 2, no. 3, pp. 111–117.
3. Ageeva E.V., Pikalov S.V., Emel'janov I.P., Ageev E.V. Optimizacija sklad-skogo hozjajstva predpriyatij avtomobil'nogo transporta ispol'zovaniem igrovyyh metodov. *Izvestija Jugo-Zapadnogo gosudarstvennogo universiteta. Serija: Tehnika i tehnologii*, 2015, no. 1, pp. 21–28.
4. Andrianov A.Ju., Kiseleva O.M. Ispol'zovanie logisticheskikh podhodov v upravlenii skladom vremennogo hranenija. *Vestnik universiteta*, 2016, no. 1.
5. Kofjač D., Kljajić M., Rejec V. The anticipative concept in warehouse optimization using simulation in an uncertain environment. *European Journal of Operational Research*, 2009, vol. 193, no. 3, pp. 660–669.
6. Klassifikacija sklad-skikh pomeschenij ot agentstva Swiss Realty Group. URL: <http://www.galaxylogistics.ru/slovar-logistiki/k-ru/klassifikatsiya-sklad-skikh-pomeschenij-ot-agentstva-swiss-realty-group.html> (data obrashhenija 15.10.2018).
7. Gridin V.N., Solodovnikov V.I., Evdokimov I.A., Filippkov S.V. Postroenie derev'ev reshenij i izvlechenie pravil iz obuchennyh nejronnyh setej. *Iskusstvennyj intellekt i prinjatje reshenij*, 2013, no. 4, pp. 26.
8. Voroncov K.V. Matematicheskie metody obuchenija po precedentam (teorija obuchenija mashin). Moscow, 2011, pp. 119–121.
9. Ahmed N.K., Atiya A.F., Gayar N.E., El-Shishiny H. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 2010, vol. 29, no. 5-6, pp. 594–621.
10. Chistjakov S.P. Sluchajnye lesa: obzor. *Trudy Karel'skogo nauchnogo centra Rossijskoj akademii nauk*, 2013, no. 1.
11. Liaw A., Wiener M. Classification and regression by randomForest. *R news*. 2002, vol. 2, no. 3, pp. 18–22.
12. Arulampalam G., Bouzerdoum A. A generalized feedforward neural network architecture for classification and regression. *Neural networks*, 2003, vol. 16, no. 5-6, pp. 561–568.
13. Bobyr' M.V. Metod nelinejnogo obuchenija nejro-nechetkoj sistemy vyvoda. *Iskusstvennyj intellekt i prinjatje reshenij*, 2018, no. 1, pp. 67–75.
14. Bobyr' M.V. Proektirovanie nejronnyh i nechetkih modelej v oblasti vychislitel'noj tehniki i sistem upravlenija. Moscow, Argamak-Media Publ., 2018, 110 p.
15. Scikitlearn. Machine Learning in Python. URL: <https://scikit-learn.org/stable/documentation.html> (data obrashhenija 15.10.2018).