

УДК 677:382  
DOI 10.47367/0021-3497\_2021\_4\_37

**ПРОБЛЕМЫ ПОСТРОЕНИЯ  
ПРОГНОСТИЧЕСКИХ МОДЕЛЕЙ ПОКАЗАТЕЛЕЙ  
ВНЕШНЕЭКОНОМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ  
ПРЕДПРИЯТИЙ ТЕКСТИЛЬНОЙ ПРОМЫШЛЕННОСТИ**

**PROBLEMS OF PREDICTIVE MODEL CONSTRUCTING  
OF INDICATORS OF FOREIGN ECONOMIC ACTIVITY  
OF TEXTILE INDUSTRY ENTERPRISES**

*А.Н. КИСЛЯКОВ*

*A.N. KISLYAKOV*

**(Российская академия народного хозяйства и государственной службы  
при Президенте Российской Федерации,  
Российская академия народного хозяйства и государственной службы  
при Президенте Российской Федерации, Владимирский филиал)**

**(Russian Presidential Academy of National Economy and Public Administration,  
Russian Presidential Academy of National Economy and Public Administration, Vladimir branch)**

E-mail: ankislyakov@mail.ru

*Работа посвящена актуальной проблеме построения прогностических моделей показателей внешнеэкономической деятельности предприятий текстильной промышленности, основанных на алгоритмах машинного обучения. Цель работы заключается в описании подхода к построению моделей*

*на основе кластерного анализа, деревьев принятия решений и нейронных сетей. Показаны примеры построения прогностических моделей на основе ансамблевых методов и искусственных нейронных сетей для набора данных, описывающего факты внешнеэкономической деятельности по товарно-номенклатурным группам и регионам Центрального федерального округа Российской Федерации. Описанный подход позволяет повысить эффективность построения и описания торгового профиля региона в целях управления развитием внешнеэкономической деятельностью региона в сегменте производства текстильной продукции.*

*The paper is devoted to the actual problem of predictive model constructing of indicators of foreign economic activity of textile industry enterprises based on machine learning algorithms. The purpose of this paper is to describe an approach to creating models based on cluster analysis, decision trees, and neural networks. Examples of predictive model constructing based on ensemble methods and artificial neural networks for a data set describing the facts of foreign economic activity by commodity-nomenclature groups and regions of the central Federal District of the Russian Federation are shown. The described approach makes it possible to increase the efficiency of constructing and describing the trade profile of the region in order to manage the development of the region's foreign economic activity in the textile production segment.*

**Ключевые слова:** текстильная промышленность, кластерный анализ, деревья решений, нейронные сети.

**Keywords:** textile industry, cluster analysis, decision trees, neural networks.

#### *Введение*

На сегодняшний день управление внешнеэкономической деятельностью предприятий текстильной промышленности в условиях товарно-санкционных ограничений и последствий пандемии COVID-19 как никогда нуждается в разработке качественных прогностических моделей показателей эффективности функционирования предприятий текстильной промышленности и исследовании новых возможностей и форматов производства, а также реализации текстильной продукции Российской Федерации в контексте управления развитием территорий. Особую актуальность приобретает задача исследования результатов деятельности предприятий как в разрезе отдельных регионов, так и в разрезе товарных групп в целях повышения эффективности экономического планирования и прогнозирования условий выпуска продукции. Решение подобных задач основано на грамотном использовании методов экономико-математического моделирования, алгоритмов анализа данных и машинного обучения.

Одним из перспективных подходов к исследованию временных рядов применительно к задаче построения прогностических моделей является использование моделей машинного обучения [1], [2], которые опираются как на методы кластерного анализа, позволяющие обобщить признаки и характеристики взаимодействия отдельных групп (кластеров) объектов социально-экономической системы, так и методы построения прогностических моделей на основе временных рядов.

Цель работы заключается в описании подхода к построению прогностических моделей показателей внешнеэкономической деятельности предприятий текстильной промышленности на основе моделей машинного обучения.

#### *Методы и инструменты*

Начальным этапом построения любой прогностической модели является изучение набора данных, который описывает предметную область. Обычно исходный набор данных представляет собой несколько показателей, представленных в виде времен-

ного ряда. Однако чаще всего временной ряд является коротким, нестационарным, с резкими изменениями значений переменных, разрывами и скачкообразными изменениями, что в значительной мере снижает качество прогностических моделей [3]. Методы виртуального расширения статистической выборки и аргументации [4] позволяют частично решить данную проблему за счет использования инструментария имитационного моделирования [5], однако более важным является необходимость учета взаимных связей отдельных предикторов модели. Указанная задача требует сравнительного анализа эффективности различных алгоритмов машинного обучения. Ряд исследований показал, что высокую эффективность при описании пространства признаков социально-экономических систем имеют алгоритмы на основе деревьев решений и искусственных нейронных сетей [6], [7].

Классические методы регрессионного анализа и описательной статистики [8] не позволяют эффективно обучить модель на основе комплекса признаков и требуют

тщательного их отбора, в то время как методы на основе глубокого обучения и теории графов менее требовательны к набору предикторов и позволяют автоматически их отбирать в процессе обучения. Основной задачей является лишь определение тех признаков, которые могут быть включены в модель.

Для описания указанного подхода из открытых источников были взяты показатели объемов импорта и экспорта по одной из групп товарной номенклатуры внешнеэкономической деятельности (ТНВЭД): 63 – Прочие готовые текстильные изделия; наборы; одежда и текстильные изделия и т.д. На рис. 1-а и б показаны результаты анализа объемов импорта/экспорта по регионам, в дол. (а), и результаты кластеризации регионов Центрального федерального округа (ЦФО) Российской Федерации (РФ) по объемам импорта и экспорта (б): объемы товарооборота в денежном выражении по осям приведены в логарифмическом масштабе.

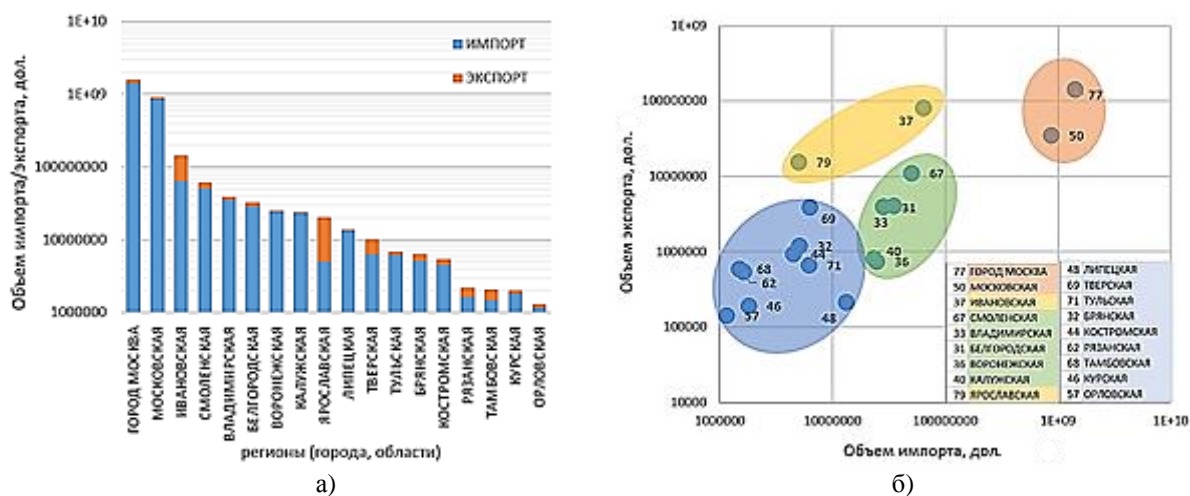


Рис. 1

Анализ результатов, представленных на рис. 1, показывает, что объемы импорта и экспорта продукции данной товарной группы значительно отличаются в зависимости от региона, что обусловлено как исторически сложившимися обстоятельствами – регионы с наиболее развитым текстильным производством: Ивановская обл., Ярославская обл., так и высокой плотностью населения – г. Москва и Московская

обл. Простая визуализация в виде временных диаграмм (рис. 2 – изменение объемов импорта и экспорта текстильной продукции в Ярославской, Ивановской и Московской областях) соотношения импорта и экспорта показывает, что в регионах с большей плотностью населения наблюдается резкое возрастание объемов импорта в начале 2020 г. в разгар "первой волны" пандемии COVID-19. Это связано с закупкой сырья

для пошива защитных медицинских масок и костюмов различных видов. Однако регионы с традиционно высоким уровнем экспорта, по всей видимости, в меньшей степени задействованы в процессе производства указанных видов продукции. Так, Ярославская и Ивановская области показали незначительную отрицательную динамику в объемах экспорта. При этом Ивановская обл. показала значительное снижение объемов импорта. Данные выводы говорят о необходимости учета указанных дополнительных признаков при построении модели.

Кроме регионального аспекта следует также обратить внимание и на товарно-номенклатурные группы [9]. На рис. 3 показано соотношение объемов импорта и экспорта текстильной продукции по товарным

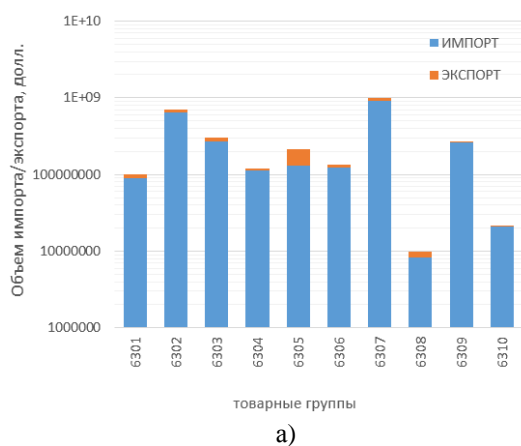


Рис. 3

группам и результаты кластерного анализа по товарным группам. Описание уточняющих кодов ТНВЭД группы 63 приведено в табл. 1 по объемам импорта и экспорта.

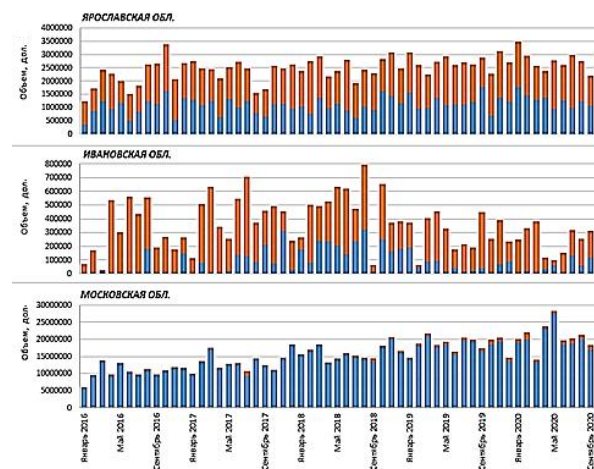


Рис. 2

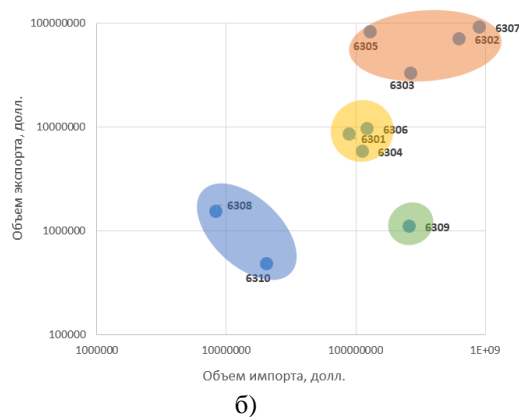


Таблица 1	
6301	Одеяла и пледы дорожные
6302	Белье постельное, столовое, туалетное и кухонное
6303	Занавеси (включая портьеры) и внутренние шторы; ламбрекены или подзоры для кроватей
6304	Изделия декоративные прочие, кроме изделий товарной позиции 9404
6305	Мешки и пакеты упаковочные
6306	Брезенты, навесы, тенты; палатки; паруса для лодок, досок для виндсерфинга или сухопутных транспортных средств; снаряжение для кемпинга
6307	Готовые изделия прочие, включая выкройки одежды
6308	Наборы, сост. из тканей и пряжи или нитей с принадлеж. или без них, для изгот. ковров, гобеленов, вышитых скатертей или салфеток или аналог. текст. изделий, упакованные или для розн. продажи
6309	Одежда и прочие изделия, бывшие в употреблении
6310	Тряпье, использованное или новое, куски бечевки, веревок, канатов и тросов и изделия из бечевки, веревок, канатов или тросов, из текстильных материалов, бывшие в употреблении

Результаты, показанные на рис. 3, позволяют сделать вывод о том, что указанные характеристики оказывают влияние на целевые показатели и также могут быть ис-

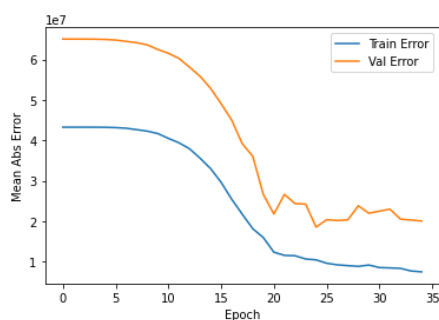
пользованы для конструирования признаков модели. При этом следует отметить, что перечень признаков может быть расширен аналогично после проведения дополни-

тельных исследований по возможности включения их в модель, например, страна импортер/экспортер.

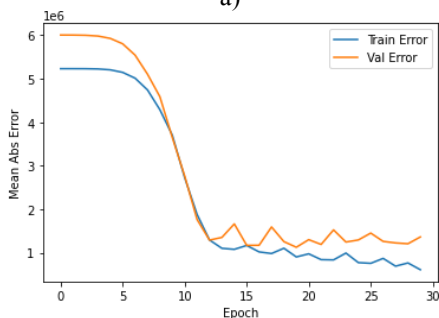
#### Результаты и обсуждение

На основе описанных признаков был сформирован набор данных, включающий описание объемов импорта/экспорта по ЦФО в дол. США в разрезе регионов и товарных групп. Показатели отражены по периодам с шагом в 1 месяц, начиная с января 2016 г. по октябрь 2020 г. Указанный набор данных состоит из 57 точек данных и 28 признаков: 10 товарно-номенклатурных групп и 18 регионов, которые являются предикторами модели. Требуется выполнить прогноз показателей объемов импорта и экспорта по ЦФО.

Результаты обучения и настройки алгоритма на основе искусственных нейронных сетей [6] для показателей объемов импорта (а) и экспорта (б) по ЦФО с использованием библиотеки tensorflow для глубокого обучения на языке Python приведены на рис. 4.



а)



б)

Рис. 4

Конфигурация многослойной сверточной нейронной сети с оптимальными характеристиками позволяет создать прогностическую модель на основе регрессии при средней абсолютной ошибке (mean absolute error – MAE) 10066350 дол. для показателя

импорта и 2158024 дол. для показателя экспорта, что составляет порядка 10 и 5% от среднего значения показателей объемов импорта и экспорта соответственно.

Дальнейшие эксперименты с конфигурацией показали, что в данном случае потенциал ИНС исчерпан и дополнение новых слоев и количества нейронов в скрытых слоях не приводит к улучшению точностных характеристик модели. Обучение выполняется за 30 эпох (epoch). Причина столь посредственных характеристик точности – недостаток обучающего набора данных, при этом даже методы расширения статистической выборки не способны кардинально исправить ситуацию.

В свою очередь, использование методов на основе деревьев решений позволяет более глубоко изучить небольшой массив данных. Однако обучение дерева решений может создать сверхсложные деревья, которые не обобщаются хорошо из обучающих данных (эффект "переобучения"). В связи с этим разработаны более сложные ансамблевые методы, такие как, например, random forest (случайный лес) с последующей выборкой данных для обучения модели (bagging) в целях повышения адекватности разрабатываемых моделей с расширением не исходной выборки данных, а результатов работы моделей [6], [10]. Результат использования алгоритма случайный лес для 10 тыс. деревьев решений и параметров для управления генератором случайных чисел, равным 50, на указанном выше наборе показал лучшие точностные характеристики при средней абсолютной ошибке 6334538,76 дол. для объемов импорта и 331616,62 для объемов экспорта, что составляет порядка 6,3 и 0,4% от среднего значения показателей объемов импорта и экспорта соответственно. Разумеется, указанные результаты могут быть улучшены при грамотной аргументации исходного набора данных.

## ВЫВОДЫ

При реализации прогностической модели на основе деревьев решений были получены лучшие характеристики точности,

чем при использовании ИНС, в связи с особенностями исходного набора данных. Описанный математический и программный инструментарий позволяет более эффективно выполнять прогноз показателей рынка текстильной продукции в ЦФО с учетом изменения внешних факторов, оказывающих влияние на хозяйственную деятельность. Указанный подход позволит не только раскрыть общее состояние показателей внешнеэкономической деятельности по сегменту текстильной продукции в ЦФО, но и дать понимание, какие алгоритмы и методы более эффективно использовать для обработки имеющихся данных.

#### ЛИТЕРАТУРА

1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning with Applications in R / Publisher: Springer. – 2013.
2. Shmueli Galit. Data Mining for Business Analytics: Concepts, Techniques, and Applications in R. – 2018.
3. Kislyakov A., Tikhonuyk N. Principles for Development of Predictive Stability Models of Social and Economic Systems on the basis of DTW // First Conference on Sustainable Development: Industrial Future of Territories (IFT 2020). – Vol. 208. № 08001. 2020. DOI: 10.1051/e3sconf/202020808001
4. Q. Wen, L. Sun, X. Song, J. Gao, X. Wang, H. Xu. Time Series Data Augmentation for Deep Learning. 2020: A Survey arXiv:2002.12478v1 [cs.LG] 27 Feb 2020
5. Кисляков А.Н. Асимметрия информации в задачах анализа социально-экономических процессов. Вестник НГУЭУ. – 2020, №1. С. 64...75. <https://doi.org/10.34020/2073-6495-2020-1-064-075>
6. Ioannou Y.A., Robertson D., Zikic D., Kotschieder P., Shotton J., Brown M., Criminisi A. Decision Forests, Convolutional Networks and the Models in-Between. – 2016. ArXiv, abs/1603.01250.
7. De'ath G. Multivariate Regression Trees: A New Technique for Modeling Species-Environment Relationships. Ecology. – № 83, 2002. P. 1105...1117.
8. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition // Publisher: Springer. – 2017.
9. Wooldridge J. M. "Introductory Econometrics - A Modern Approach" South-Western Pub. – 2004. Ch.8. P.248...271.
10. Liu X., Liu X., Lai Y., Yang F., Zeng Y. Random Decision DAG: An Entropy Based Compression Approach for Random Forest. In: Li G., Yang J., Gama J., Natwchai J., Tong Y. (eds) Database Systems for Advanced Applications. DASFAA 2019. Lecture Notes in Computer Science. – Vol. 11448, 2019. Springer, Cham. [https://doi.org/10.1007/978-3-030-18590-9\\_37](https://doi.org/10.1007/978-3-030-18590-9_37)

Natwchai J., Tong Y. (eds) Database Systems for Advanced Applications. DASFAA 2019. Lecture Notes in Computer Science. – Vol. 11448, 2019. Springer, Cham. [https://doi.org/10.1007/978-3-030-18590-9\\_37](https://doi.org/10.1007/978-3-030-18590-9_37)

#### REFERENCES

1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning with Applications in R / Publisher: Springer. – 2013.
2. Shmueli Galit. Data Mining for Business Analytics: Concepts, Techniques, and Applications in R. – 2018.
3. Kislyakov A., Tikhonuyk N. Principles for Development of Predictive Stability Models of Social and Economic Systems on the basis of DTW // First Conference on Sustainable Development: Industrial Future of Territories (IFT 2020). – Vol. 208. № 08001. 2020. DOI: 10.1051/e3sconf/202020808001
4. Q. Wen, L. Sun, X. Song, J. Gao, X. Wang, H. Xu. Time Series Data Augmentation for Deep Learning. 2020: A Survey arXiv:2002.12478v1 [cs.LG] 27 Feb 2020
5. Kislyakov A.N. Asimetriya informatsii v zadachakh analiza sotsial'no-ekonomicheskikh protsessov. Vestnik NGUEU. – 2020, №1. S. 64...75. <https://doi.org/10.34020/2073-6495-2020-1-064-075>
6. Ioannou Y.A., Robertson D., Zikic D., Kotschieder P., Shotton J., Brown M., Criminisi A. Decision Forests, Convolutional Networks and the Models in-Between. – 2016. ArXiv, abs/1603.01250.
7. De'ath G. Multivariate Regression Trees: A New Technique for Modeling Species-Environment Relationships. Ecology. – № 83, 2002. P. 1105...1117.
8. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition // Publisher: Springer. – 2017.
9. Wooldridge J. M. "Introductory Econometrics - A Modern Approach" South-Western Pub. – 2004. Ch.8. P.248...271.
10. Liu X., Liu X., Lai Y., Yang F., Zeng Y. Random Decision DAG: An Entropy Based Compression Approach for Random Forest. In: Li G., Yang J., Gama J., Natwchai J., Tong Y. (eds) Database Systems for Advanced Applications. DASFAA 2019. Lecture Notes in Computer Science. – Vol. 11448, 2019. Springer, Cham. [https://doi.org/10.1007/978-3-030-18590-9\\_37](https://doi.org/10.1007/978-3-030-18590-9_37)

Рекомендована кафедрой экономики РАНХ и ГС, Владимирский филиал. Поступила 28.07.21.