

УДК 519.254

https://doi.org/10.33619/2414-2948/97/07

ПРЕДЛОЖЕНИЯ ПО ИСПОЛЬЗОВАНИЮ БОЛЬШИХ ДАННЫХ ДЛЯ ПРОИЗВОДСТВЕННОЙ КОМПАНИИ

- ©**Павлов Е. М.**, ORCID: 0009-0002-0028-8270, SPIN-код: 3158-0223,
Национальный исследовательский университет «МЭИ»,
г. Москва, Россия, PavlovyEgM@mpei.ru
- ©**Рыжов А. В.**, ORCID: 0009-0003-4729-9505, Национальный исследовательский
университет «МЭИ», г. Москва, Россия, RyzhovAVI@mpei.ru
- ©**Баланев К. С.**, ORCID: 0009-0002-9722-7262, SPIN-код: 8192-1861, Национальный
исследовательский университет «МЭИ», г. Москва, Россия, BalanевKS@mpei.ru
- ©**Крепков И. М.**, канд. техн. наук, Национальный исследовательский университет «МЭИ»,
г. Москва, Россия, KrepkovIM@mpei.ru

SUGGESTIONS FOR THE USE OF BIG DATA FOR A MANUFACTURING COMPANY

- ©**Pavlov E.**, ORCID: 0009-0002-0028-8270, SPIN-code: 3158-0223, National Research University
Moscow Power Engineering Institute, Moscow, Russia, PavlovyEgM@mpei.ru
- ©**Ryzhov A.**, ORCID: 0009-0003-4729-9505, National Research University Moscow Power
Engineering Institute, Moscow, Russia, RyzhovAVI@mpei.ru
- ©**Balanев K.**, ORCID: 0009-0002-9722-7262, SPIN-code: 8192-1861, National Research
University Moscow Power Engineering Institute, Moscow, Russia, BalanевKS@mpei.ru
- ©**Krepkov I.**, Ph.D., National Research University Moscow Power Engineering Institute,
Moscow, Russia, KrepkovIM@mpei.ru

Аннотация. Рассматриваются этапы и методы работы с большими данными. Описываются задачи и проблемы в области лесозаготовок и производства продукции из древесины, решаемые с помощью анализа обработанных данных. В статье представлен алгоритм работы с большими данными и его реализация в виде программного комплекса, разработанного на языке Python. Перечисляются основные выгоды, получаемые на основе визуализации обработанной информации.

Abstract. This paper discusses the stages and methods of working with big data. It describes the tasks and problems in the field of logging and wood products production, solved by analyzing processed data. The article presents an algorithm for working with big data and its realization in the form of a program complex developed in the Python language. The main benefits derived from the visualization of processed information are listed.

Ключевые слова: большие данные, анализ данных, лесозаготовки, язык Python.

Keywords: big data, data analytics, timber harvesting, Python language.

В качестве организации выступает производственная компания, занимающаяся производством деревянных плит, фанеры, ОСБ и др. Для компании такого рода одной из основных задач является снижение издержек, что, в свою очередь, возможно благодаря анализу больших данных, связанных с лесным хозяйством [3].

В качестве первого этапа работы необходимо осуществить *сбор данных*. Они могут включать в себя включая информацию о лесном хозяйстве, природных явлениях (экологии,

засуха, ливни), миграции деревьев и других релевантных параметрах. Этот этап поможет выявить закономерности и связи между данными. Поскольку большие данные могут быть неструктурированными и содержать ошибки, после этого важно провести процедуры очистки данных, чтобы убрать дубликаты, исправить ошибки и привести данные в пригодный для анализа вид. На следующем этапе создаются предикаты, которые будут использоваться для прогнозирования и выявления возможных проблем или оптимизации процессов. Например, можно создать предикаты, связанные с ростом и зрелостью деревьев, погодными условиями, и т. д.

Завершающим этапом является построение аналитической модели на основе собранных и очищенных данных. Методы данного этапа могут включать в себя машинное обучение, статистический анализ, временные ряды и т. д. [1]. По завершению обработки больших данных возможно их дальнейшее использование для решения различных проблем и задач в области лесозаготовок и производства продукции из древесины. Среди таких задач можно выделить:

- *Оптимизацию заготовки сырья.* Анализ данных о лесных ресурсах позволит определить оптимальные временные интервалы и локации для заготовки древесины, учитывая природные условия и засухи.

- *Управление рисками.* Моделирование данных о природных явлениях позволит компании более эффективно управлять рисками, связанными с экологическими и погодными факторами.

- *Оптимизацию производства.* Анализ данных о производственных процессах и данных о рынке поможет оптимизировать производство и управлять запасами более эффективно.

- *Снижение потерь.* Предсказания и рекомендации, полученные из аналитической модели, помогут уменьшить потери материалов и снизить издержки.

Все этапы работы с большими данными представляют собой определенный алгоритм [4]. В данном случае он имеет следующий вид (Рисунок 1).

Для реализации данного алгоритма был разработан прототип программного комплекса, собирающего большие данные. Он представляет собой программу на языке Python [2].

В качестве собираемых данных были выбраны текущие (эксплуатационные) затраты на охрану окружающей среды по регионам РФ (<https://showdata.gks.ru/report/279336/>).

В процессе сбора происходит запрос к вышеуказанной витрине статистических данных на сайте и последующее извлечение всех данных.

Затем из сырых данных извлекаются лишь необходимые элементы с соответствующими классами и происходит их структуризация. В ходе этого процесса также происходит видоизменение полученных элементов. Например, удаляются лишние цифры в названии регионов (Рисунок 2), и лишние пробелы в числах (Рисунок 3).

После очистки данных на их основе создается база данных SQLite с одной таблицей (Рисунок 4), имеющая один атрибут для региона и ряд атрибутов на каждый из годов. Из таблицы видно, что данные действительно прошли очистку и были приведены к более структурированному виду.

Для визуализации хранимых в БД данных генерируются гистограммы с данными по каждому региону (Рисунок 5). Гистограммы формируются на основе различных запросов к БД. Однако перед этим также происходит преобразование данных из строкового типа к целочисленному.

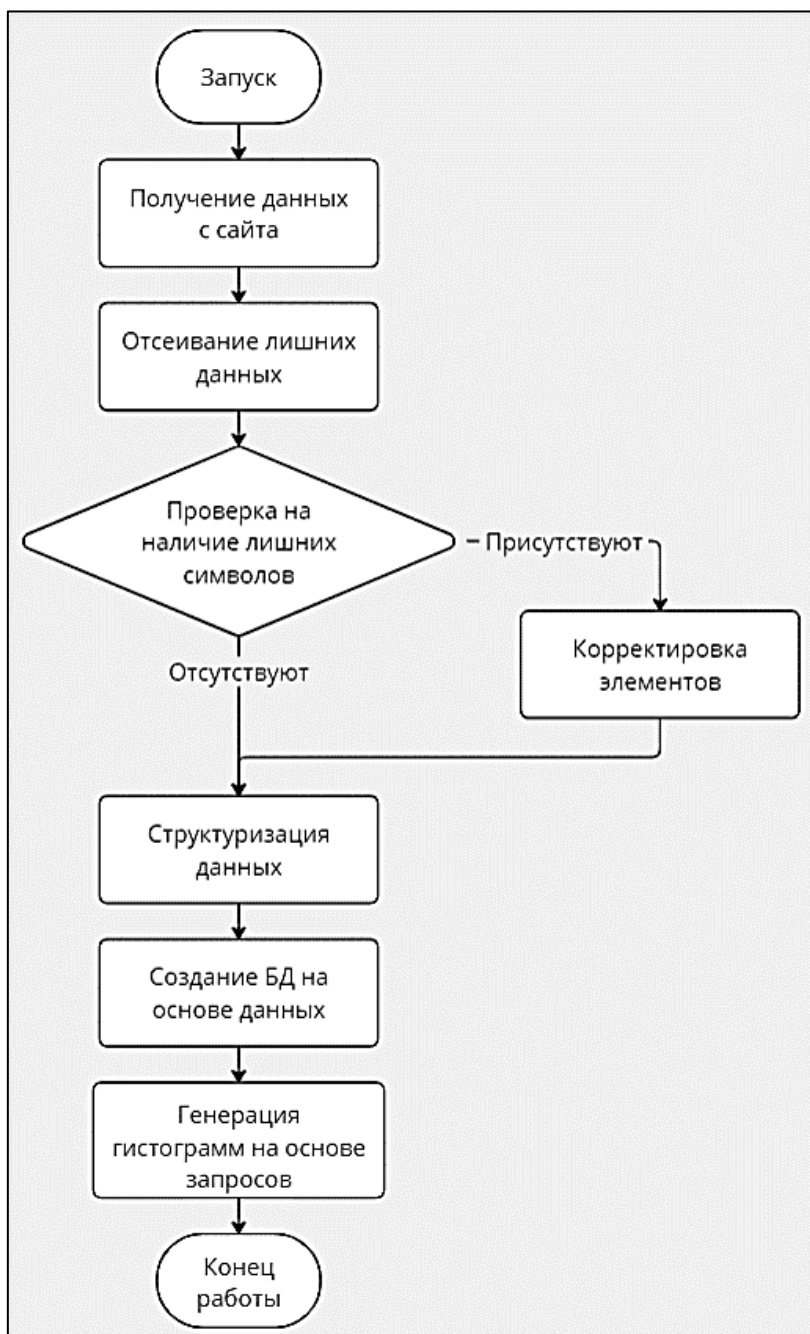


Рисунок 1. Общий алгоритм

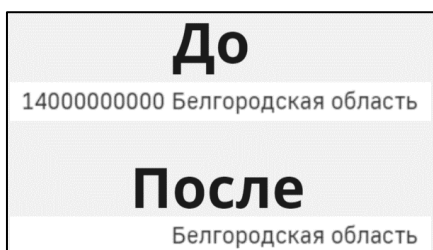


Рисунок 2. Удаление лишних символов

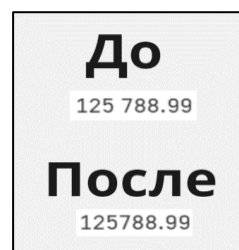


Рисунок 3. Удаление пробелов

Region	2013	2014
.АГ Сухопутные территории Арктической зоны Рос...	NaN	NaN
Российская Федерация	498407.00	525144.0
Центральный федеральный округ	87557.00	119895.0
Белгородская область	8556.00	10429.0
Брянская область	166.00	2462.0
Владимирская область	283.99	544.0
Воронежская область	2095.00	2106.0
Ивановская область	1461.00	1250.0
Калужская область	2156.00	2276.0
Костромская область	1042.00	548.0
Курская область	976.99	2102.0
Липецкая область	595.00	2530.0

2015	2016	2017	2018	2019
NaN	NaN	NaN	NaN	10288.00
485118.00	870629.00	475456.99	684653.00	707598.00
139091.00	125788.99	148923.00	134728.00	170869.00
10150.00	10020.00	13556.00	13033.00	13372.00
332.00	118.99	94.00	978.99	603.00
245.99	270.00	519.99	836.99	893.99
1959.00	2426.00	2348.00	3717.00	8554.00
1100.00	925.00	2470.00	554.99	775.00
2906.00	3510.00	12261.00	4049.00	11459.00
1033.00	890.99	1869.00	2275.00	1988.00
1894.00	1614.00	6483.00	5151.00	6566.00
679.00	676.00	1779.00	2322.00	2371.00

Рисунок 4. Вывод значений из базы данных

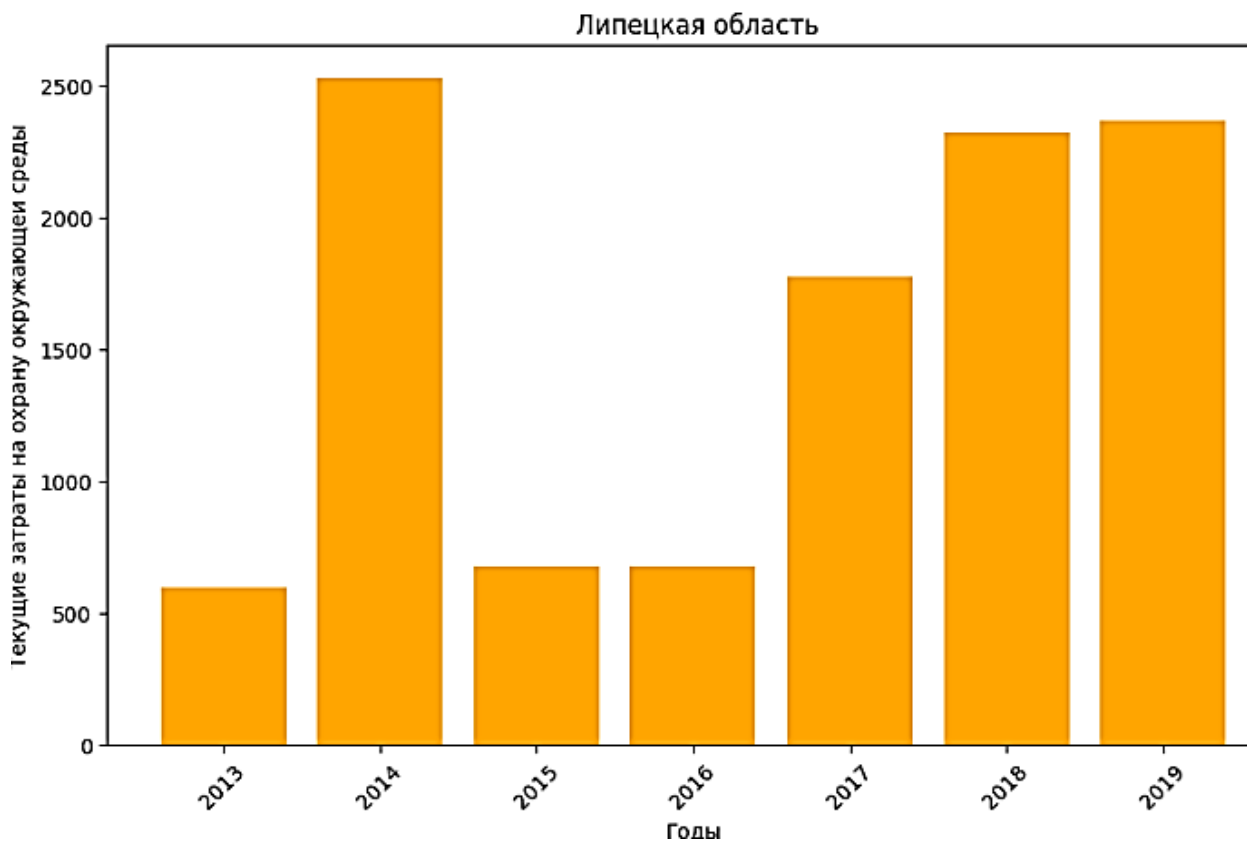


Рисунок 5. Визуализация данных в виде гистограммы

В ходе работы программы задействуется ряд специализированных библиотек. Среди них: matplotlib (построение графиков), pandas (структуризация данных), sqlite3 (создание и взаимодействие с базой данных), BeautifulSoup (взаимодействие с данными через классы), webdriver (взаимодействие с веб-страницами) и ряд других библиотек [5].

Полученная визуализация информации о затратах на охрану окружающей среды по регионам Российской Федерации может быть важным инструментом для производственной компании, занимающейся производством изделий из дерева, чтобы снизить издержки и улучшить свою конкурентоспособность.

Анализ затрат на охрану окружающей среды в различных регионах позволяет компании выбирать местоположение своих производственных объектов так, чтобы минимизировать затраты на соблюдение экологических нормативов и требований. Информация о затратах на охрану окружающей среды также может влиять на стоимость древесного сырья и его доступность в разных регионах. Это позволяет компании оптимизировать цепочку поставок сырья и снизить издержки на его транспортировку.

Знание различий в требованиях к охране окружающей среды в разных регионах позволяет компании эффективно планировать и внедрять меры для соблюдения местных нормативов. Это может включать в себя внедрение технологий с меньшим воздействием на окружающую среду, управление отходами и другие практики. Анализ затрат на охрану окружающей среды может помочь компании лучше взаимодействовать с местными органами власти, предлагать совместные инициативы по улучшению экологической ситуации в регионе и получать поддержку и льготы в обмен на соблюдение экологических стандартов.

Производственные компании, активно заботящиеся об охране окружающей среды, могут использовать эту информацию для маркетинга своих продуктов. Это может привести к увеличению спроса на продукцию компании и увеличению ее репутации. Знание о затратах на охрану окружающей среды также позволяет компании более точно оценивать потенциальные экологические риски в разных регионах и разрабатывать планы для их управления.

Список литературы:

1. Дейтел П., Дейтел Х. Python: Искусственный интеллект, большие данные и облачные вычисления. СПб: Питер, 2020. 864 с.
2. Дэви С., Арно М., Мохамед А. Основы Data Science и Big Data. Python и наука о данных. СПб: Питер, 2017. 336 с.
3. Майер-Шенбергер В., Кукьер К. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. М.: Манн, Иванов и Фербер, 2014. 240 с.
4. Марц Н., Джеймс У. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени. М.: Вильямс, 2018. 368 с.
5. Митчелл Р. Современный скрапинг веб-сайтов с помощью Python. СПб: Питер, 2021. 336 с.

References:

1. Deitel, P., & Deitel, H. (2020). Python: Iskusstvennyi intellekt, bol'shie dannye i oblachnye vychisleniya. St. Petersburg. (in Russian).
2. Devi, S., Arno, M., & Mokhamed, A. (2017). Osnovy Data Science i Big Data. Python i nauka o dannykh. St. Petersburg. (in Russian).
3. Maier-Shenberger, V., & Kuker, K. (2014). Bol'shie dannye. Revolyutsiya, kotoraya

izmenit to, kak my zhivem, rabotaem i myslim. Moscow. (in Russian).

4. Marts, N., & Dzheims, U. (2018). Bol'shie dannye. Printsipy i praktika postroeniya masshtabiruemykh sistem obrabotki dannykh v real'nom vremeni. Moscow. (in Russian).

5. Mitchell, R. (2021). Sovremenniy skraping veb-saitov s pomoshh'yu Python. St. Petersburg. (in Russian).

*Работа поступила
в редакцию 09.11.2023 г.*

*Принята к публикации
19.11.2023 г.*

Ссылка для цитирования:

Павлов Е. М., Рыжов А. В., Баланев К. С., Крепков И. М. Предложения по использованию больших данных для производственной компании // Бюллетень науки и практики. 2023. Т. 9. №12. С. 59-64. <https://doi.org/10.33619/2414-2948/97/07>

Cite as (APA):

Pavlov, E., Ryzhov, A., Balanov, K., & Krepkov, I. (2023). Suggestions for the Use of Big Data for a Manufacturing Company. *Bulletin of Science and Practice*, 9(12), 59-64. (in Russian). <https://doi.org/10.33619/2414-2948/97/07>