



ЭКСПЕРИМЕНТАЛЬНАЯ ЭКОНОМИКА И ФИНАНСЫ

№4 2023

СОДЕРЖАНИЕ

МАТЕМАТИЧЕСКИЕ, СТАТИСТИЧЕСКИЕ И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ В ЭКОНОМИКЕ

Булгаков А.Л., Урядников А.М.
Прогнозирование туристических показателей
с помощью искусственного интеллекта 3

Степанович К.С., Булгаков А.Л.
Использование методов машинного
обучения при обработке данных:
практические аспекты 17

РЕГИОНАЛЬНАЯ И ОТРАСЛЕВАЯ ЭКОНОМИКА

Дмитриева Л.В., Алешина А.В.
Повышение эффективности промышленных
предприятий в условиях санкций 32

ФИНАНСЫ

Воронкова С.Н., Алешина А.В.
Финтех и децентрализованные финансы:
тенденции, проблемы и роль в финансовом
секторе 37

НАУЧНЫЙ ЖУРНАЛ

Основан в июне 2021 года

Выходит 4 раза в год

Зарегистрирован Федеральной службой по надзору в
сфере связи, информационных технологий и массовых
коммуникаций (РОСКОМНАДЗОР).

Рег. ПИ № ФС77-81413 от 30 июня 2021 года

ISSN 2782-3644

Учредители:

ООО «Издательство «КноРус»

ООО «Институт экспериментальной экономики и финансов
МГУ имени М.В. Ломоносова»

Адрес редакции:

Россия, 117218, Москва,
ул. Кедрова, д. 14, корп. 2
Многоканальный телефон/факс: +7 (495) 741-46-28

Сайт: www.eeaf.ru

Почта: welcome@eeaf.ru

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

А.Л. Булгаков

Главный редактор:

А.В. Алешина

Отпечатано в типографии ООО «Русайнс», 117218, Мо-
сква, ул. Кедрова, д. 14, корп. 2

Тираж 300 экз. Формат А4. Подписано в печать: 29.12.2023
Цена свободная

Все материалы, публикуемые в журнале, подлежат вну-
треннему и внешнему рецензированию

Издание не подлежит маркировке согласно п. 2 ст. 1
Федерального закона от 29.12.2010 № 436-ФЗ «О защите
детей от информации, причиняющей вред их здоровью и
развитию»

К СВЕДЕНИЮ АВТОРОВ

Уважаемые коллеги! Обращаем ваше внимание на то, что матери-
алы статей проходят обязательную экспертизу. После экспертизы
статьи поступают в Редакцию журнала, где проходят редактор-
скую и корректорскую правку. Редакция оставляет за собой право
сокращать объем статей и редактировать их в соответствии с тре-
бованиями научного журнала. Рукописи статей не возвращаются;
с авторами в переписку Редакция не вступает; гонорар авторам
не выплачивается.

ПЕРЕЧЕНЬ ДОКУМЕНТОВ И ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ И СОПУТСТВУЮЩИХ МАТЕРИАЛОВ, НЕОБХОДИМЫХ ДЛЯ ПУБЛИКАЦИИ

Текст статьи, при оформлении которого необходимо соблюсти
следующие требования: объем статьи - до 60 тыс. знаков (1,5 авт.
листа); в статье должна быть следующая информация: ФИО авто-
ра(ов) полностью, место работы (учебы), контактная информация
(телефон, E-mail); аннотация и ключевые слова к статье, список
литературы (на русском и английском языках).

TABLE OF CONTENTS

MATHEMATICAL, STATISTICAL AND INSTRUMENTAL METHODS IN ECONOMICS

Bulgakov A.L., Uryadnikov A.M.
Predicting tourism performance using artificial
intelligence. 3

Stepanovich K.S., Bulgakov A.L.
Using Machine Learning Methods in Data
Processing: Practical Aspects. 17

REGIONAL AND INDUSTRY ECONOMICS

Dmitrieva L.V., Aleshina A.V.
Improving the efficiency of industrial
enterprises under sanctions 32

FINANCE

Voronkova S.N., Aleshina A.V.
Fintech and decentralized finance: trends,
problems and role in the financial sector. 37

SCIENTIFIC JOURNAL

Founded in June 2021
Published 4 times a year

*Registered by the Federal Service for Supervision of
Communications, Information Technology and Mass
Media (ROSKOMNADZOR).*

Reg. PI No. FS77-81413 dated June 30, 2021

ISSN 2782-3644

Founders:

*Knorus Publishing House LLC, Institute of
Experimental Economics and Finance of the
Lomonosov Moscow State University LLC*

EDITORIAL OFFICE:

Russia, 117218, Moscow, Kedrova St., 14, bldg. 2
Multi-channel phone/fax: +7 (495) 741-46-28

Website: www.eeaf.ru

Mail: welcome@eeaf.ru

CHIEF EDITOR

Anna Valentinovna Aleshina

EDITORIAL TEAM

Andrey Leonidovich Bulgakov

*Printed at the printing house LLC Rusyns,
117218, Moscow, st. Kedrova, d.14, building 2*

*Circulation 300 copies. A4 format. Signed to print:
29/12/2023*

Free price

*All materials published in the journal are subject to
internal and external review.*

*The publication is not subject to labeling in
accordance with paragraph 2 of Art. 1 of the Federal
Law of December 29, 2010 No. 436-FZ "On the
Protection of Children from Information Harmful to
Their Health and Development"*

МАТЕМАТИЧЕСКИЕ, СТАТИСТИЧЕСКИЕ
И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ В ЭКОНОМИКЕПрогнозирование туристических показателей
с помощью искусственного интеллекта**Булгаков Андрей Леонидович**

к. э. н., доцент МГУ имени М. В. Ломоносова;
Доцент РЭУ им. Г. В. Плеханова
Профессор Московского института
современного академического образования
E-mail: z3900207@mail.ru

Урядников Анатолий Максимович

студент РЭУ им. Г. В. Плеханова
E-mail: u.a-98@yandex.ru

В работе рассматривается создание модели машинного обучения для прогнозирования туристических показателей. Анализируются зависимости между числом прибытий иностранных туристов в страну и различными факторами, включая население, площадь территории, ВНД (валовой национальный доход, метод Атласа), выбросы CO₂, расходы на здравоохранение и индексы потребительских цен. Особое внимание уделяется проблемам отрасли туризма в Российской Федерации, а также факторам ее развития. Исследование опирается на массивы данных из открытых источников, в частности, с сайта ООН и сайта группы Всемирного банка. Эти данные предоставляют обширную базу для анализа и построения прогностической модели. Модель машинного обучения, разрабатываемая в рамках исследования, призвана помочь в принятии стратегических решений в туристической отрасли. Она может быть полезна как для государственных органов, так и для частных компаний, работающих в сфере туризма. Результаты исследования могут способствовать более эффективному планированию и развитию туристической инфраструктуры, а также разработке целенаправленных маркетинговых стратегий для привлечения иностранных туристов.

Ключевые слова: туризм в России, цифровизация туризма, кризисы в туризме, машинное обучение, искусственный интеллект, нейронные сети.

Введение

За счет большого культурного наследия, разнообразия достопримечательностей и обилия природных ландшафтов Россия имеет крупный потенциал для развития туризма. В настоящее время ведется активное инвестирование в модернизацию и развитие данного сегмента рынка, привлекаются новые технологии, строятся достопримечательности и развивается инфраструктура. Развитие инфраструктуры в том числе связано с крупными международными событиями, прошедшими на территории России. Например, Зимние Олимпийские игры 2014 года, Чемпионат мира по футболу 2018 года и Формула 1 в Сочи, которые также являлись драйверами роста интереса иностранных туристов. Согласно Стратегии развития туризма в РФ до 2035 г. ключевым преимуществом развития туризма в РФ для въездных и внутренних туристов является наличие множества точек притяжения. Однако последние несколько лет отрасль сталкивается с новыми вызовами, ограничивающими развитие и заставляющие разрабатывать новые стратегии реализации потенциала.

В связи с кризисом, вызванным распространением COVID-2019 пришлось пересмотреть фактор безопасности точек массового скопления людей, а усиление санкционного давления на РФ вынудило пересмотреть экономическую выгоду некоторых популярных направлений. В тоже время накопительный отложенный спрос спровоцировал повышение популярности внутреннего туризма и про-

лил свет на новые инвестиционные проекты. Появилась резкая необходимость в анализе новых маршрутов и мер по закрытию потребностей участников рынка. Благодаря большому количеству открытых статистических и разнородных данных и удобным инструментам машинного обучения появилась возможность привнести современные технологии в процесс анализа и построения вектора развития государств в различных сферах, а в частности, туризма.

Искусственный интеллект, как инструмент прогноза туристических показателей

Мощным инструментом для развития социальной и экономических сфер государства стала цифровизация. За счет анализа и обработки большого количества данных появилась возможность строить прогнозы развития проектов, получая конкретные численные показатели. Искусственный интеллект (ИИ) способен обнаружить неочевидные закономерности, что позволяет с высокой точностью спрогнозировать тенденции и просчитать экономические выгоды от инвестиций. Алгоритмы машинного обучения позволяют произвести анализ различных показателей, связанных с туризмом во всем мире, чтобы помочь выявить факторы, на которые стоит обратить внимание при составлении стратегии развития в данной сфере.

Имея различные числовые показатели (индикаторы) стран за различные временные промежутки, можно построить прогнозную модель ИИ, для исследования факторов, влияющих на количество прилетов туристов в страну в течении года. Для решения данной задачи был разработан логический формат файла, объединяющий количественные характеристики стран, взятые из открытых источников международных организаций. Был написан скрипт на языке Python, реализующий данный логический формат и использующий методы машинного обучения для обучения модели ИИ и прогнозирование туристических показателей.

Импорт библиотек для обработки данных

Для начала работы с функциями и расширенными возможностями, которые не входят в стандартную библиотеку языка Python, необходимо импортировать дополнительные модули. Далее следует код, подключающий необходимые дополнительные библиотеки для решения задачи.

```
[ ] from google.colab import drive
drive.mount('/content/drive')
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount(«/content/drive», force_remount=True).
```

```
[ ] #Встроенный модуль для работы с регулярными выражениями
#re – https://docs.python.org/3/library/re.html
import re
```

```
#Встроенный модуль для чтения и записи файлов формата «CSV»
#csv – https://docs.python.org/3/library/csv.html
import csv
```

```
#Библиотека для обработки и анализа данных. Высокоуровневые структуры данных и операции для эффективного анализа
#pandas – https://pandas.pydata.org
import pandas as pd
```

```
#библиотека для научных вычислений: обработка многомерных массивов и выполнения математических операций
#NumPy – https://numpy.org/
import numpy as np
```

```
#Библиотека для визуализации данных
#Matplotlib – https://matplotlib.org/
import matplotlib.pyplot as plt
```

```
#Библиотека визуализации данных для создания статистических графиков
```

```
#seaborn – https://seaborn.pydata.org/
import seaborn as sns

#Библиотека машинного обучения с открытым исходным кодом
#sklearn.preprocessing –
https://scikit-learn.org/
from sklearn.preprocessing import
LabelEncoder
from sklearn.model_selection import
train_test_split
from sklearn.linear_model import
LinearRegression
from sklearn.ensemble import
RandomForestRegressor,
GradientBoostingRegressor
from sklearn.svm import SVR
from sklearn.neural_network import
MLPRegressor
from sklearn.cluster import KMeans,
DBSCAN, AgglomerativeClustering,
MeanShift
from sklearn.metrics import mean_
squared_error
from sklearn import metrics

# расширенный анализ DataFrame, по-
зволяющий экспортировать анализ дан-
ных в различные форматы, такие как
html и json
# ydata_profiling – https://ydata-
profiling.ydata.ai/docs/master/
pages/getting_started/overview.html
!pip install ydata-profiling
from ydata_profiling import
ProfileReport

#библиотека для параллельных вы-
числений и масштабирования работы
с данными, позволяющая эффективно
работать с большими датасетами.
#dask поддерживает структуры данных,
аналогичные тем, что есть в Pandas
(например, DataFrame) и NumPy (на-
пример, массивы).
#dask – https://www.dask.org/
import dask.dataframe as dd
```

Импорт неструктурированных данных с сайта ООН

В качестве источников использовались данные с сайта группы Всемирного банка, а также сайта ООН. К сожалению, данные не структурированы, что делает необходи-мым их дальнейшее преобразование к обще-му виду.

Были загружены следующий данные:

- Международный туризм, количество прибы-тий «International tourism, number of arrivals» <https://data.worldbank.org/indicator/ST.INT.ARVL>
- Население страны «Population, total» <https://data.worldbank.org/indicator/SP.POP.TOTL>
- Площадь территории «Surface area (sq. km)» <https://data.worldbank.org/indicator/AG.SRF.TOTL.K2>
- ВНД* (в текущих долларах США) «GNI, Atlas method (current US\$)» <https://data.worldbank.org/indicator/NY.GNP.ATLS.CD>
- Выбросы CO2 «CO2 emissions (kt)» <https://data.worldbank.org/indicator/EN.ATM.CO2E.KT>
- Расходы на здравоохранение «Health expenditure» http://data.un.org/_Docs/SYB/CSV/SYB65_325_202209_Expenditure%20on%20health.csv
- Индексы потребительских цен «Consumer price indices» http://data.un.org/_Docs/SYB/CSV/SYB65_128_202209_Consumer%20Price%20Index.csv

Данные с сайта группы Всемирного банка скачиваются в виде Zip-архива, в который упа-кованы файлы в формате csv. На самом сайте есть удобный инструмент визуализации, по-зволяющий наглядно оценить данные перед использованием (см. рис. 1)

International tourism, number of arrivals

World Tourism Organization, Yearbook of Tourism Statistics, Compendium of Tourism Statistics and data files.

License : CC BY-4.0 [i](#)

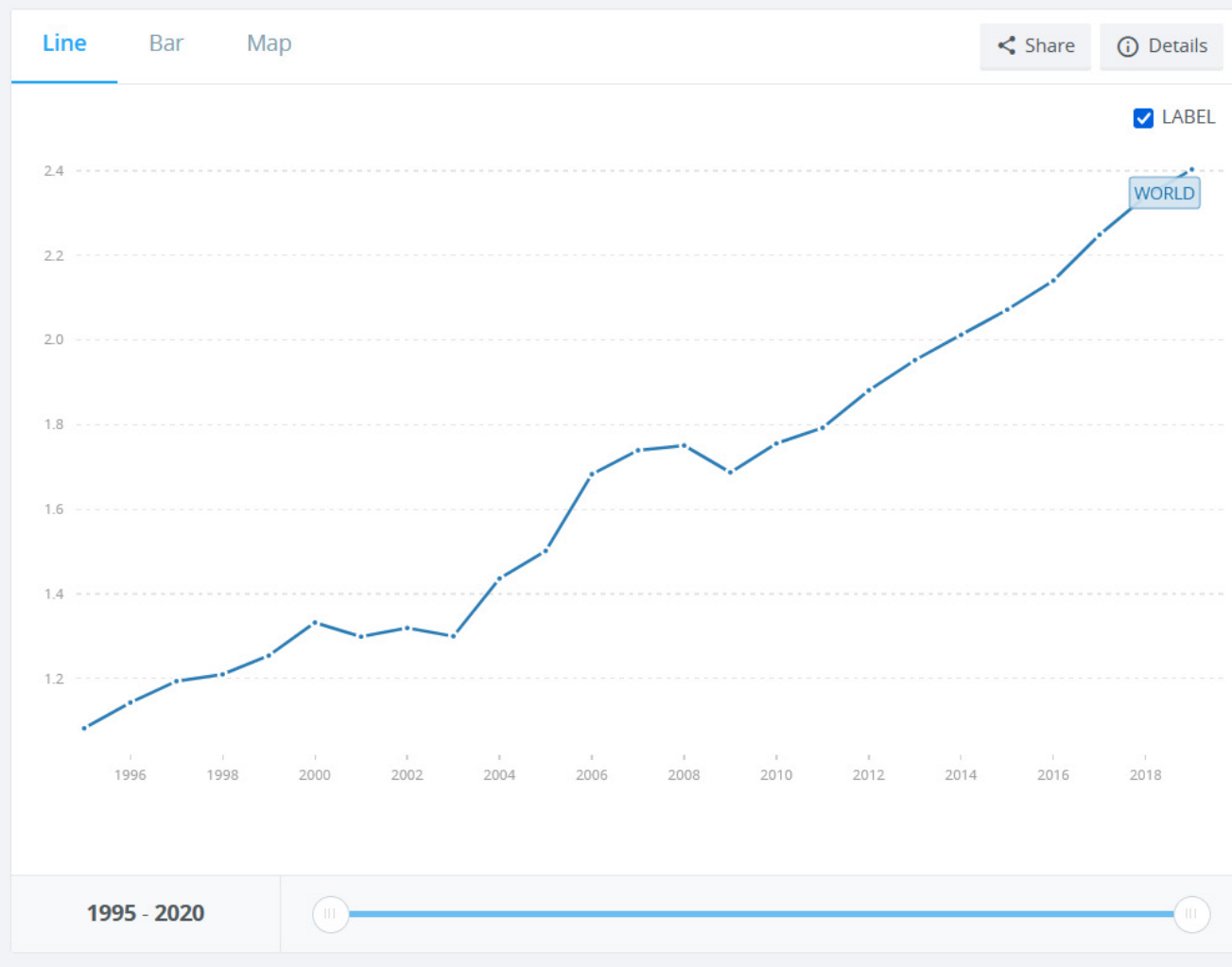


Рис. 1. Данные с сайта Всемирного банка (International tourism, number of arrivals) [26]

В начале файла содержится заголовок с описанием, который необходимо учитывать перед использованием файла. Сами данные перечисляются через запятую.

На сайте ООН <http://data.un.org/> данные доступны для загрузки в двух форматах: pdf и csv. Данные в pdf сведены в таблицу 1 и разбиты по страницам.

,"World Development Indicators",

| Date", "2023-05-10",

,"Country Code", "Indicator Name", "Indicator
 "1961", "1962", "1963", "1964", "1965", "1966", "1967", "1
 ", "1987", "1988", "1989", "1990", "1991", "1992", "1993",
 "12", "2013", "2014", "2015", "2016", "2017", "2018", "2019
 ", "CO2 emissions

Таблица 1. Данные с сайта ООН «Население, площадь и плотность расселения».

2		Population, surface area and density		Population, superficie et densité						
Country or area Pays ou zone	Year Année	Mid-year population estimates and projections (millions) Estimations et projections de population au milieu de l'année (millions)			Sex Ratio (males per 100 females) Rapport des sexes (hommes pour 100 femmes)	Population age distribution (percentage) Répartition par âge de la population (pourcentage)		Population density (per km ²) Densité de population (pour km ²)	Surface area (000 km ²) Superficie (000 km ²)	
		Total	Male Hommes	Females Femmes		Aged 0 to 14 years old âgée de 0 à 14 ans	Aged 60+ years old âgée de 60 ans ou plus			
Total, all countries or areas	2010	6 985.60	3 514.41	3 471.20	101.2	27.1	11.1	53.6	...	...
	2015	7 426.60	3 737.40	3 689.19	101.3	26.4	12.3	57.0	...	...
Total, tous pays ou zones	2020	7 840.95	3 943.61	3 897.34	101.2	25.7	13.5	60.1	130 094	...
	2022 ¹	7 975.11	4 008.58	3 966.53	101.1	25.3	13.9	61.2	...	...
Africa	2010	1 055.23	525.87	529.37	99.3	41.5	5.0	35.7	...	...
Afrique	2015	1 201.11	599.30	601.81	99.6	41.3	5.2	40.6	...	...
	2020	1 360.68	679.36	681.32	99.7	40.5	5.4	46.0	29 648	...
	2022 ¹	1 426.74	712.43	714.31	99.7	40.1	5.5	48.3	...	...
Northern Africa	2010	207.11	104.59	102.52	102.0	32.4	6.6	26.9	...	...
Afrique septentrionale	2015	228.36	115.31	113.05	102.0	32.8	7.3	29.7	...	...
	2020	251.42	126.86	124.56	101.9	32.8	8.2	32.7	7 769	...
	2022 ¹	259.97	131.12	128.85	101.8	32.5	8.5	33.8	...	...
Sub-Saharan Africa	2010	848.12	421.27	426.85	98.7	43.7	4.6	38.8	...	...
Afrique subsaharienne	2015	972.75	483.99	488.76	99.0	43.2	4.7	44.5	...	...
	2020	1 109.26	552.50	556.76	99.2	42.3	4.8	50.7	21 879 ²	...
	2022 ¹	1 166.77	581.31	585.46	99.3	41.8	4.8	53.4	...	...
Eastern Africa	2010	342.74	169.60	173.15	97.9	45.0	4.1	51.5	...	...
Afrique orientale	2015	393.35	194.84	198.51	98.2	43.8	4.3	59.1	...	...
	2020	449.29	222.75	226.54	98.3	42.1	4.5	67.5	6 667	...
	2022 ¹	473.00	234.54	238.47	98.4	41.4	4.5	71.1	...	...
Middle Africa	2010	133.61	66.32	67.29	98.6	45.6	4.5	20.6	...	...
Afrique centrale	2015	157.40	78.22	79.18	98.8	45.7	4.4	24.3	...	...
	2020	184.57	91.79	92.78	98.9	45.5	4.4	28.4	6 497	...
	2022 ¹	196.08	97.51	98.57	98.9	45.4	4.4	30.2	...	...
Southern Africa	2010	59.10	28.50	30.60	93.1	29.7	7.3	22.2	...	...
Afrique australe	2015	63.72	30.99	32.72	94.7	29.2	7.8	24.0	...	...
	2020	67.27	32.74	34.53	94.8	29.5	8.3	25.3	2 651	...
	2022 ¹	68.60	33.44	35.16	95.1	29.3	8.4	25.8	...	...
Western Africa	2010	312.67	156.86 ³	155.80 ³	100.7 ³	44.1 ³	4.7 ³	51.6	...	...
Afrique occidentale	2015	358.28	179.94 ³	178.35 ³	100.9 ³	44.1 ³	4.6 ³	59.1	...	...
	2020	408.12	205.22 ³	202.91 ³	101.1 ³	43.2 ³	4.7 ³	67.3	6 064	...
	2022 ¹	429.08	215.82 ³	213.26 ³	101.2 ³	42.7 ³	4.7 ³	70.8	...	...
Americas ²	2010	935.82	461.93	473.89	97.5	24.7	13.1	24.1	...	...
Amériques ²	2015	983.54	485.81	497.73	97.6	23.2	14.7	25.4	...	...
	2020	1 025.79	508.71	519.09	97.6	21.8	16.5	26.4	38 791	...
	2022 ¹	1 037.14	511.80	525.34	97.4	21.2	17.2	26.7	...	...
Northern America	2010	345.27	170.17 ⁴	175.10 ⁴	97.2 ⁴	19.5 ⁴	18.4 ⁴	18.5	...	...
Amérique septentrionale	2015	360.46	178.29 ⁴	182.18 ⁴	97.6 ⁴	19.0 ⁴	20.5 ⁴	19.3	...	...
	2020	373.96	185.36 ⁴	188.56 ⁴	98.3 ⁴	18.2 ⁴	22.8 ⁴	20.0	18 652	...
	2022 ¹	376.87	186.67 ⁴	190.20 ⁴	98.1 ⁴	17.7 ⁴	23.7 ⁴	20.2	...	...
Latin America & the Caribbean	2010	590.55	291.77	298.78	97.7	27.7	10.0	29.1	...	...
Amérique latine et Caraïbes	2015	623.08	307.53	315.55	97.5	25.6	11.3	30.7	...	...
	2020	651.84	321.31	330.52	97.2	23.9	12.9	32.2	20 139	...
	2022 ¹	660.27	325.13	335.14	97.0	23.2	13.4	32.6	...	...
Caribbean	2010	41.40	20.57 ⁵	20.84 ⁵	98.7 ⁵	26.3 ⁵	11.9 ⁵	186.9	...	...
Caraïbes	2015	42.75	21.21 ⁵	21.55 ⁵	98.4 ⁵	24.8 ⁵	13.1 ⁵	193.0	...	...
	2020	43.96	21.76 ⁵	22.20 ⁵	98.0 ⁵	23.6 ⁵	14.5 ⁵	198.5	226	...
	2022 ¹	44.39	21.94 ⁵	22.45 ⁵	97.8 ⁵	23.2 ⁵	15.1 ⁵	200.4	...	...
Central America	2010	158.06	76.68	79.39	96.6	31.3	8.4	63.3	...	...
Amérique centrale	2015	167.19	82.15	85.04	96.6	28.9	9.5	67.8	...	...
	2020	176.34	86.54	89.81	96.4	26.6	10.9	71.6	2 452	...
	2022 ¹	179.06	87.73	91.33	96.1	25.8	11.3	72.7	...	...

Population and migration - Population et migration

13

Данные в формате csv содержат 1 строку-заголовок, после чего идут через запятую. Стоит отметить, что некоторые данные из файла в формате PDF не соответствуют данным из файла csv.

n, density and surface area",,,,,,
/Area,,Year,Series,Value,Footnotes,Source
countries or areas",2010,Population mid-year estimate
accessed July 2022."
countries or areas",2010,Population mid-year estimate
last accessed July 2022."
countries or areas",2010,Population mid-year estimate
ion, last accessed July 2022."

Следующий код демонстрирует процесс импорта (загрузки) данных в программу:

```
[ ] #Импорт данных с сайта группы
Всемирного банка data.worldbank.org
dataArrivals = pd.read_csv('/
content/drive/MyDrive/colab_data/
arrivals.csv')
dataPopulation = pd.read_csv('/
content/drive/MyDrive/colab_data/
population.csv')
dataSurfaceArea = pd.read_csv('/
content/drive/MyDrive/colab_data/
surfaceArea.csv')
dataGNI = pd.read_csv('/content/
drive/MyDrive/colab_data/gni.csv')
dataCO2 = pd.read_csv('/content/
drive/MyDrive/colab_data/CO2.csv')

#Импорт данных с сайта ООН data.
un.org
dataHealth = pd.read_
csv('http://data.un.org/_Docs/SYB/
CSV/SYB65_325_202209_Expenditure%20
on%20health.csv', header = 1,
thousands=',')
dataPriceIndices = pd.read_
csv('http://data.un.org/_Docs/SYB/
CSV/SYB65_128_202209_Consumer%20
Price%20Index.csv', header = 1,
thousands=',')
```

Предобработка данных

Поскольку данные имеют разнородный характер, их необходимо привести к общему виду. В следующей функции происходит удаление ненужных столбцов из датасетов worldbank, перенос данных из столбцов с годами в строки, удаляются лишние данные, приводятся значения в столбце «год» к числовому типу.

```
[ ] def fix_dataframe(df, df_name):
# Удаление ненужных столбцов
df = df.drop(df.columns[[1, 2, 3]],
axis=1)
#Транспонирование данных из столбцов
в строки
df = df.melt(id_vars=[«Country
Name»], var_name=»Year»)
```

```
#Изменение названий столбцов
df.columns = ['Country Name',
'Year', df_name]
# Удаление лишних строк
df = df.drop(df[df['Year'] ==
'Unnamed: 67'].index)
#Преобразование столбца «Year» в це-
лочисленный формат
df['Year'] = df['Year'].
astype('int64')
return df
```

```
dataArrivalProcessing = fix_
dataframe(dataArrivals, 'Arrivals')
dataPopulationProcessing =
fix_dataframe(dataPopulation,
'Population')
dataSurfaceAreaProcessing =
fix_dataframe(dataSurfaceArea,
'SurfaceArea')
dataGNIProcessing = fix_
dataframe(dataGNI, 'GNI')
dataCO2Processing = fix_
dataframe(dataCO2, 'CO2')
```

Далее необходимо обработать данные, полученный с сайта ООН

```
[ ] #Обработка данных с расхода-
ми на здравоохранение «Health
expenditure»
dataHealthProcessing =
dataHealth[['Region/Country/Area»,
«Unnamed: 1», «Year»,»Series»,
«Value»]]
dataHealthProcessing.columns = ['Id',
'Region', 'Year', 'Series', 'Value']

dfs = [dataHealthProcessing[dataHea
lthProcessing['Series'] == 'Current
health expenditure (% of GDP)'].set_
index(['Id', 'Region', 'Year']),

dataHealthProcessing[dataHealthProce
ssing['Series'] == 'Domestic general
government health expenditure (% of
total government expenditure)'].set_
index(['Id', 'Region', 'Year'])]
```

```
dataHealthProcessing = pd.concat(dfs,
axis=1).reset_index()
```

```
dataHealthProcessing.columns = ['Id',
'Region', 'Year', 'Series', 'Current
health expenditure (% of GDP)',
'Series', 'Domestic general
government health expenditure (% of
total government expenditure)']
dataHealthProcessing =
dataHealthProcessing[['Region',
'Year', 'Current health expenditure
(% of GDP)', 'Domestic general
government health expenditure (% of
total government expenditure)']]
dataHealthProcessing.
columns = ['Country Name',
'Year', 'HealthExpenditure',
'GvmHealthExpenditure']
```

```
[] #Обработка данных индексов по-
требительских цен «Consumer price
indices»
dataPriceIndicesProcessing = dataPri
ceIndices[dataPriceIndices[«Series»]
== «Consumer price index: General»]
dataPriceIndicesProcessing = dataPr
iceIndicesProcessing[['Unnamed: 1»,
«Year», «Value»]]
dataPriceIndicesProcessing.columns =
['Country Name', 'Year', 'Indices']
```

```
[] #Разница между длиной строк в об-
работанном наборе данных с сайта
группы Всемирного банка и сайта ООН
len(dataArrivalProcessing) -
len(dataPriceIndicesProcessing)
15520
```

В связи с тем, что данные от ООН представлены с меньшим отчетным периодом по годам, общее число самих данных сильно уступает данным с сайта группы Всемирного банка, что сильно повлияет на результаты машинного обучения. Далее это будет наглядно продемонстрировано.

Получение нового логического формата данных

Исходные данные были предобработаны для получения нового логического формата. Далее создается два формата: в первом записаны данные с сайта Всемирного банка, второй тип включается в себя также данные с сайта ООН. В процессе объединения названия стран заменяются на числовые значения.

```
[] #Функция слияния обработанных да-
тасетов в единую структуру
def merge_dataframes(*dataframes):
if len(dataframes) < 2:
raise ValueError(«Нужно передать как
минимум два DataFrame»)
```

```
merged_df = dataframes[0]
```

```
for df in dataframes[1:]:
merged_df = pd.merge(merged_df, df,
on=['Country Name', 'Year'])
```

```
return merged_df
```

```
#Функция трансформации названия
страны в датасете
def countryChanger(df):
cc_le = LabelEncoder()
cc_le.fit(df['Country Name'].values)
cc_encodet = cc_
le.transform(df['Country Name'].
values)
global countryDict
countryDict = dict(zip(cc_
le.classes_, cc_le.transform(cc_
le.classes_)))
df['Country Name'] = cc_encodet
return df
```

```
# Словарь для последующей обратной
конвертации названий стран
countryDict = {}
```

```
mergeData = merge_datafra
mes(dataArrivalProcessing
, dataPopulationProcessing,
```

```
dataSurfaceAreaProcessing,
dataGNIProcessing, dataCO2Processing)
logicData = countryChanger(mergeData)
```

```
mergeData2 = merge_dataframes(
dataArrivalProcessing,
dataPopulationProcessing,
dataSurfaceAreaProcessing,
dataGNIProcessing, dataCO2Processing,
dataHealthProcessing,
dataPriceIndicesProcessing)
logicData2 =
countryChanger(mergeData2)
```

Подготовка логического формата к машинному обучению

Полученные логические форматы необходимо подготовить для машинного обучения. Стоит обратить внимание на года, которые остались в форматах после объединения:

```
[] #Оставшиеся года для первого формата:
unique_years = logicData['Year'].
unique()
print(unique_years)
[1960 1961 1962 1963 1964 1965 1966 1967 19
68 1969 1970 1971 1972 1973
1974 1975 1976 1977 1978 1979 1980 1981 198
2 1983 1984 1985 1986 1987
1988 1989 1990 1991 1992 1993 1994 1995 199
6 1997 1998 1999 2000 2001
2002 2003 2004 2005 2006 2007 2008 2009 201
0 2011 2012 2013 2014 2015
2016 2017 2018 2019 2020 2021 2022]
```

```
[] #Оставшиеся года для второго формата:
unique_years = logicData2['Year'].
unique()
print(unique_years)
[2000 2005 2010 2019]
```

Как видно, для формата, где включены данные с сайта ООН, остались только те данные, которые есть в отчетном периоде организации.

Далее необходимо удалить пустые значения, если такие имеются:

```
[] logicData.isnull().sum()
Country Name 0
Year 0
Arrivals 10958
Population 358
SurfaceArea 1150
GNI 5528
CO2 9589
dtype: int64
>[] logicData2.isnull().sum()
Country Name 0
Year 0
Arrivals 58
Population 0
SurfaceArea 0
GNI 13
CO2 1
HealthExpenditure 0
GvmHealthExpenditure 0
Indices 0
dtype: int64
>[] logicData = logicData.
dropna(how='any')
logicData2 = logicData2.
dropna(how='any')
logicData.isnull().sum() +
logicData2.isnull().sum()
Arrivals 0.0
CO2 0.0
Country Name 0.0
GNI 0.0
GvmHealthExpenditure NaN
HealthExpenditure NaN
Indices NaN
Population 0.0
SurfaceArea 0.0
Year 0.0
dtype: float64
```

Полученный формат файла после обработки имеет следующий вид (таблица 2):

```
[] logicData
```

Таблица 2. Обработанные данные с сайта ООН по годам и разным странам.

	Country Name	Year	Arrivals	Population	SurfaceArea	GNI	CO2
9311	1	1995	1.158354e+07	3.534666e+08	1.510674e+07	2.545332e+11	3.274330e+05
9313	2	1995	2.670706e+06	2.358615e+08	9.166270e+06	9.537971e+10	1.161258e+05
9314	7	1995	9.000000e+03	1.391225e+07	1.246700e+06	5.114858e+09	1.271968e+04
9315	3	1995	3.040000e+05	3.187784e+06	2.875000e+04	2.035220e+09	1.928510e+03
9319	10	1995	2.289000e+06	3.494611e+07	2.780400e+06	2.563472e+11	1.128922e+05
...	...	...	...	...	...	...	...
15953	262	2019	2.403074e+09	7.742682e+09	1.346644e+08	8.902440e+13	3.551286e+07
15954	204	2019	1.810000e+05	2.119050e+05	2.840000e+03	8.957249e+08	2.954000e+02
15957	219	2019	1.479700e+07	5.808706e+07	1.219090e+06	3.908618e+11	4.396400e+05
15958	264	2019	1.266000e+06	1.838048e+07	7.526100e+05	2.561246e+10	6.803900e+03
15959	265	2019	2.294000e+06	1.535461e+07	3.907600e+05	2.232742e+10	1.175990e+04

Источник: составлено авторами

5072 rows × 7 columns

Теперь можно оценить плотность данных по годам в полученном формате (рис. 2):

```
[ ] sns.histplot(logicData[«Year»],
binwidth=0.9)
<Axes: xlabel='Year', ylabel='Count'>
```

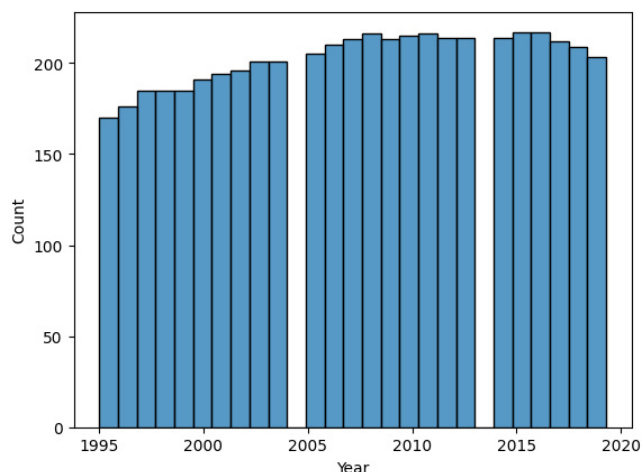


Рис. 2. Распределение данных по годам (составлено авторами)

Для наглядности полученных данных можно построить графики по столбцам для РФ:

```
[ ] mergeData = merge_dataframes(dataArrivalProcessing,
dataPopulationProcessing,
dataSurfaceAreaProcessing,
dataGNIProcessing, dataCO2Processing)
filtered_df =
mergeData[mergeData['Country Name']
== 'Russian Federation']
fig, axs = plt.subplots(nrows=1,
ncols=4, figsize=(16, 4))
```

```
#Построение графика для 'Arrivals'
axs[0].plot(filtered_df['Year'],
filtered_df['Arrivals'], marker='o')
axs[0].set_xlabel('Year')
axs[0].set_ylabel('Arrivals')
axs[0].set_title('Arrivals')
```

```
#Построение графика для 'Population'
axs[1].plot(filtered_df['Year'],
filtered_df['Population'],
marker='o')
axs[1].set_xlabel('Year')
axs[1].set_ylabel('Population')
axs[1].set_title('Population')
```

```
#Построение графика для 'GNI'
axs[2].plot(filtered_df['Year'],
filtered_df['GNI'], marker='o')
```

```

axs[2].set_xlabel('Year')
axs[2].set_ylabel('GNI')
axs[2].set_title('GNI')

#Построение графика для 'C02'
axs[3].plot(filtered_df['Year'],
filtered_df['C02'], marker='o')
axs[3].set_xlabel('Year')

axs[3].set_ylabel('C02')
axs[3].set_title('C02')

#Размещение графиков внутри сетки
plt.tight_layout()

#Отображение графиков (см. рис. 3)
plt.show()

```

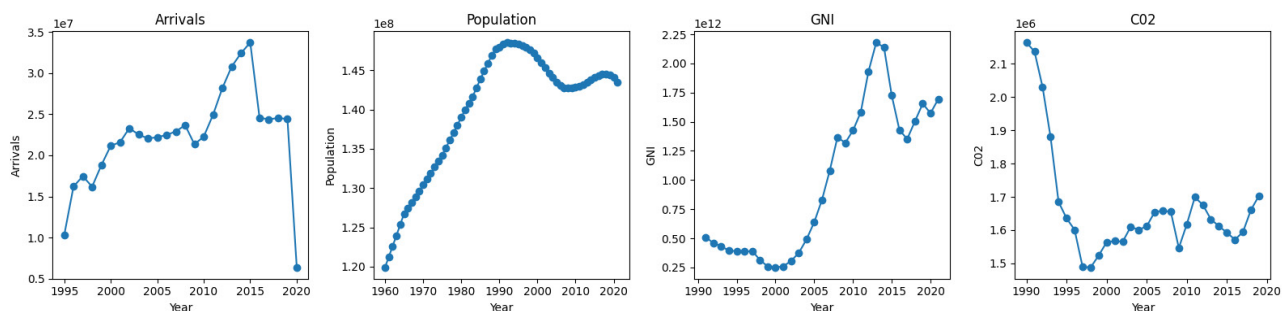


Рис. 3. Функциональные графики, показывающие динамику посещений туристов, количества населения, ВНД, выбросы углекислого газа (составлено авторами)

Создание моделей машинного обучения

Чтобы использовать полученные данные для прогноза показателей, необходимо создать обучающие модели данных. Следующий код создает и обучает две модели линейной регрессии для двух полученных форматов:

```

[] #Модель для первого формата
X = logicData[['Country Name',
'Year', 'Population', 'SurfaceArea',
'GNI', 'C02']].copy()
Y = logicData['Arrivals'].copy()
X_train, X_test, Y_train, Y_test
= train_test_split(X, Y, test_
size=0.3, random_state=0)
model = LinearRegression()
model.fit(X_train, Y_train)

#Модель для второго формата
X2 = logicData2[['Country Name',
'Year', 'Population', 'SurfaceArea',
'GNI', 'C02', 'HealthExpenditure',
'GvmHealthExpenditure', 'Indices']].
copy()
Y2 = logicData2['Arrivals'].copy()

```

```

X_train2, X_test2, Y_train2, Y_test2
= train_test_split(X2, Y2, test_
size=0.3, random_state=0)
model2 = LinearRegression()
model2.fit(X_train2, Y_train2)
LinearRegression
LinearRegression()

```

Оценка точности прогнозирования

Была проведена оценка точности прогнозирования.

```

[] model.score(X_train, Y_train)
0.9134850535676391
[] model2.score(X_train2, Y_train2)
0.3657235780785588

```

В первом случае точность составила ~91%, а во втором ~36,5%. Несмотря на большее число параметров (количество столбцов) ключевую роль при обучении сыграло количество данных (строк). Однозначно стоит использовать первую модель для прогноза будущих показателей. Точность в дан-

ном случае говорит о том, что зная прогнозы страны по определенным выше показателям, можно с высокой точностью определить количество туристов в стране на прогнозируемый год.

Распределенные вычисления

Поскольку для машинного обучения используется большой объем входящих данных. Для оптимизации обработки больших данных можно использовать распределённую обработку. В python существует несколько библиотек, которые помогут реализовать такую задачу, одной из них является dask. Dask эффективно масштабируется от локальных вычислений до кластерных, что позволяет обрабатывать большие объемы данных, размер которых превышает объем доступной в устройстве оперативной памяти. Это особенно важно в областях, где объемы данных могут превышать гигабайты и даже терабайты, например, в генетике, астрофизике и климатологии. Распределенная обработка также будет полезна и для подготовки данных к машинному обучению. Для загрузки и обработки данных, используемых в данном исследовании, код с учетом методов Dask будет выглядеть следующим образом.

```
[ ] def fix_dataframe(df, df_name):
df = df.drop(df.columns[[1, 2, 3]],
axis=1)
df = df.melt(id_vars=[«Country
Name»], var_name=»Year»)
df.columns = ['Country Name',
'Year', df_name]
df = df[df['Year']!= 'Unnamed: 67']
df['Year'] = df['Year'].
astype('int64')
return df
```

```
#Чтение данных с помощью Dask
dataArrivals = dd.read_csv('/
content/drive/MyDrive/colab_data/
arrivals.csv')
dataPopulation = dd.read_csv('/
content/drive/MyDrive/colab_data/
```

```
population.csv')
dataSurfaceArea = dd.read_csv('/
content/drive/MyDrive/colab_data/
surfaceArea.csv')
dataGNI = dd.read_csv('/content/
drive/MyDrive/colab_data/gni.csv')
dataCO2 = dd.read_csv('/content/
drive/MyDrive/colab_data/CO2.csv')
```

```
#Чтение данных из внешних источников
с указанием типов данных
```

```
dataHealth = dd.read_
csv('http://data.un.org/_Docs/SYB/
CSV/SYB65_325_202209_Expenditure%20
on%20health.csv', header=1,
thousands=',', assume_missing=True)
dataPriceIndices = dd.read_
csv('http://data.un.org/_Docs/SYB/
CSV/SYB65_128_202209_Consumer%20
Price%20Index.csv', header=1,
thousands=',', assume_missing=True)
```

```
#Обработка данных
```

```
dataArrivalProcessing = fix_
dataframe(dataArrivals, 'Arrivals')
dataPopulationProcessing =
fix_dataframe(dataPopulation,
'Population')
dataSurfaceAreaProcessing =
fix_dataframe(dataSurfaceArea,
'SurfaceArea')
dataGNIProcessing = fix_
dataframe(dataGNI, 'GNI')
dataCO2Processing = fix_
dataframe(dataCO2, 'CO2')
```

```
#Преобразование данных о здоровье
```

```
dataHealthProcessing =
dataHealth[['Region/Country/Area»,
«Unnamed: 1», «Year», «Series»,
«Value»]]
dataHealthProcessing.columns =
['Id', 'Country Name', 'Year',
'Series', 'Value']
health_expenditure = dataHealthProce
ssing[dataHealthProcessing['Series']
== 'Current health expenditure (% of
GDP)']
```

```
government_health_expenditure = dataHealthProcessing[dataHealthProcessing['Series'] == 'Domestic general government health expenditure (% of total government expenditure)']
```

#Слияние данных

```
dataHealthProcessing = dd.merge(health_expenditure, government_health_expenditure, on=['Id', 'Country Name', 'Year'], suffixes=('', '_y'), how='outer') dataHealthProcessing = dataHealthProcessing.drop(columns=['Series', 'Series_y'])
```

#Переименование столбцов

```
dataHealthProcessing.columns = ['Id', 'Country Name', 'Year', 'HealthExpenditure', 'GvmHealthExpenditure']
```

#Фильтрация и преобразование данных индексов цен

```
dataPriceIndicesProcessing = dataPriceIndices[dataPriceIndices['Series'] == 'Consumer price index: General'] dataPriceIndicesProcessing = dataPriceIndicesProcessing[['Unnamed: 1', 'Year', 'Value']] dataPriceIndicesProcessing.columns = ['Country Name', 'Year', 'Indices']
```

Dask обеспечивает тесную интеграцию с широко используемыми в Python библиотеками, такими как NumPy, Pandas и Scikit-Learn. Это позволяет легко интегрировать Dask в существующие рабочие процессы, минимизируя необходимость в значительных изменениях кода.

Выводы

Полученные модели машинного обучения показали возможность прогнозирования туристических показателей с помощью искусственного интеллекта. Данный код можно

использовать и для других данных с сайтов группы Всемирного банка и ООН, и определить зависимости между различными индикаторами стран. Также возможна реализация прогнозной модели для индикаторов страны, зная ее ближайшие показатели.

Литература

1. F. Alfaro-Almagro, M. Jenkinson, N. K. Bangerter, J. L. Andersson, L. Griffanti, G. Douaud, et al., «Image processing and Quality Control for the first 10000 brain imaging datasets from UK Biobank», *NeuroImage*, vol. 166, pp. 400–424, 2018.
2. V. Hayot-Sasson, S. T. Brown and T. Glatard, «Performance Evaluation of Big Data Processing Strategies for Neuroimaging», 19th IEEE/ACM International Symposium on Cluster Cloud and Grid Computing (CC-Grid), 2019.
3. Matthew Rocklin, «Dask: Parallel Computation with Blocked algorithms and Task Scheduling», *Proceedings of the 14th Python in Science Conference*, pp. 126–132, 2015.
4. Y. Cheng and F. Rusu, «Parallel in-situ data processing with speculative loading», *SIGMOD*, pp. 1287–1298, 2014.
5. Anaconda, [online] // Электронный ресурс // URL: <https://www.continuum.io/>.
6. S. Arumugam, A. Dobra, C. Jermaine, N. Pansare, and L. Perez. The DataPath System: A Data-Centric Analytic Processing Engine for Large Data Warehouses. In *SIGMOD 2010*.
7. J. Dean and S. Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. *Commun. ACM*, 51(1), 2008
8. Announcing the Consortium for Python Data API Standards. // Электронный ресурс // URL: https://data-apis.org/blog/announcing_the_consortium/.
9. Dask. // Электронный ресурс // URL: <https://dask.org/>.
10. D. A. Watt. Programming language design concepts. John Wiley & Sons, 2004.
11. P. Sinthong and M. J. Carey. Aframe: Extending dataframes for large-scale modern data analysis. In 2019 IEEE International Conference

- on Big Data (Big Data), pages 359–371. IEEE, 2019.
12. Stephan Hoyer and Joseph J. Hamman: xarray: N-d labeled arrays and datasets in python. *Journal of Open Research Software*, 5, apr 2017.
13. Ryan Abernathey: Step-by-Step Guide to Building a Big Data Portal.
14. Adnan, K., Akbar, R.: An analytical study of information extraction from unstructured and multidimensional big data. *Journal of Big Data* 6(1), 91 (2019).
15. Angelova, R., Siersdorfer, S.: A neighborhood-based approach for clustering of linked document collections. In: *Proceedings of the 15th ACM International Conference on Information and Knowledge Management*. p. 778–779. CIKM '06, Association for Computing Machinery (2006).
16. Bach, M., Krstic, Z., Seljan, S., Turulja, L.: Text mining for big data analysis in financial sector: A literature review. *Sustainability (Switzerland)* 11(5) (2019).
17. Grishin, V.: Method of analysis and search for borrowings in the text. *Problems of science* 7, 31 (2018)
18. Jalali, S.M.J., Park, H.W., Vanani, I.R., Pho, K.H.: Research trends on big data domain using text mining algorithms. *Digital Scholarship in the Humanities* (04 2020). 19. Logica, B., Magdalena, R.: Using big data in the academic environment. *Procedia Economics and Finance* 33, 277–286 (2015).
19. Rocklin, M.: Dask: Parallel computation with blocked algorithms and task scheduling. In: *Python in Science Conference*. pp. 126–132 (2015).
20. Salvador, S., Chan, P.: Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In: *16th IEEE International Conference on Tools with Artificial Intelligence*. pp. 576–584 (2004).
21. Zambelli, A.: A data-driven approach to estimating the number of clusters in hierarchical clustering. *F1000Research* 5 (2016).
22. Elgendy, N., Elragal, A.: Big data analytics: a literature review paper. *Lect. Notes Comput. Sci.* 8557, 214–227 (2014).
23. Gandomi, A., Haider, M.: Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manage.* 35(2), 137–144 (2015)
24. Shin, D., Shim, J.: A systematic review on data mining for mathematics and science education. *Int. J. Sci. Math. Educ.* 19, 639–659 (2021)
25. Статистические данные по всему миру, предоставляемые Организацией Объединенных Наций // Электронный портал ООН // Электронный ресурс // URL: <http://data.un.org/>
26. Открытые данные Всемирного банка // Электронный портал ООН // Электронный ресурс // URL: <https://data.worldbank.org>

Predicting tourism performance using artificial intelligence

Bulgakov A. L., Uryadnikov A. M.

Russian Economic University. G. V. Plekhanov, MISAO, Lomovosov Moscow State University

This paper examines the creation of a machine learning model for predicting tourism performance. The dependencies between the number of foreign tourist arrivals in the country and various factors, including population, area, GNI (gross national income, Atlas method), CO2 emissions, health care expenditures and consumer price indices are analyzed. Special attention is paid to the problems of the tourism industry in the Russian Federation and the factors of its development. The study relies on massive data sets from open sources, in particular from the UN website and the World Bank Group website. These data provide an extensive base for analyzing and building a predictive model. The machine learning model being developed as part of the study is designed to help in strategic decision making in the tourism industry. It can be useful for both government agencies and private companies operating in the tourism industry. The results of the study can contribute to more effective planning and development of tourism infrastructure, as well as the development of targeted marketing strategies to attract foreign tourists.

Keywords: Tourism in Russia, digitalization of tourism, crises in tourism, machine learning, artificial intelligence, neural networks.

References

1. F. Alfaro-Almagro, M. Jenkinson, N. K. Bangerter, J. L. Andersson, L. Griffanti, G. Douaud, et al., «Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank,» *NeuroImage*, vol. 166, pp. 400–424, 2018.

2. V. Hayot-Sasson, S. T. Brown and T. Glatard, «Performance Evaluation of Big Data Processing Strategies for Neuroimaging,» 19th IEEE/ACM International Symposium on Cluster Cloud and Grid Computing (CC-Grid), 2019.
3. Matthew Rocklin, «Dask: Parallel Computation with Blocked algorithms and Task Scheduling,» Proceedings of the 14th Python in Science Conference, pp. 126–132, 2015.
4. Y. Cheng and F. Rusu, «Parallel in-situ data processing with speculative loading,» SIGMOD, pp. 1287–1298, 2014.
5. Anaconda // Electronic resource // URL: <https://www.continuum.io/>.
6. S. Arumugam, A. Dobra, C. Jermaine, N. Pansare, and L. Perez. The DataPath System: A Data-Centric Analytic Processing Engine for Large Data Warehouses. In SIGMOD 2010.
7. J. Dean and S. Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. Commun. ACM, 51(1), 2008
8. Announcing the Consortium for Python Data API Standards. URL: https://data-apis.org/blog/announcing_the_consortium/.
9. Dask. // Electronic resource // URL: <https://dask.org/>.
10. D. A. Watt. Programming language design concepts. John Wiley & Sons, 2004.
11. P. Sinthong and M. J. Carey. Aframe: Extending dataframes for large-scale modern data analysis. In 2019 IEEE International Conference on Big Data (Big Data), pages 359–371. IEEE, 2019.
12. Stephan Hoyer and Joseph J. Hamman: xarray: N-d labeled arrays and datasets in python. Journal of Open Research Software, 5, April 2017.
13. Ryan Abernathy: Step-by-Step Guide to Building a Big Data Portal.
14. Adnan, K., Akbar, R.: An analytical study of information extraction from unstructured and multidimensional big data. Journal of Big Data 6(1), 91 (2019).
15. Angelova, R., Siersdorfer, S.: A neighborhood-based approach for clustering of linked document collections. In: Proceedings of the 15th ACM International Conference on Information and Knowledge Management. p. 778–779. CIKM '06, Association for Computing Machinery (2006).
16. Bach, M., Krstic, Z., Seljan, S., Turulja, L.: Text mining for big data analysis in the financial sector: A literature review. Sustainability (Switzerland) 11(5) (2019).
17. Grishin, V.: Method of analysis and search for borrowings in the text. Problems of science 7, 31 (2018)
18. Jalali, S.M.J., Park, H.W., Vanani, I.R., Pho, K.H.: Research trends on big data domain using text mining algorithms. Digital Scholarship in the Humanities (04 2020).
19. Logica, B., Magdalena, R.: Using big data in the academic environment. Procedia Economics and Finance 33, 277–286 (2015).
19. Rocklin, M.: Dask: Parallel computation with blocked algorithms and task scheduling. In: Python in Science Conference. pp. 126–132 (2015).
20. Salvador, S., Chan, P.: Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In: 16th IEEE International Conference on Tools with Artificial Intelligence. pp. 576–584 (2004).
21. Zambelli, A.: A data-driven approach to estimating the number of clusters in hierarchical clustering. F1000 Research 5 (2016).
22. Elgendy, N., Elragal, A.: Big data analytics: a literature review paper. Lect. Notes Comput. Sci. 8557, 214–227 (2014).
23. Gandomi, A., Haider, M.: Beyond the hype: Big data concepts, methods, and analytics. Int. J. Inf. Manage. 35(2), 137–144 (2015)
24. Shin, D., Shim, J.: A systematic review on data mining for mathematics and science education. Int. J. Sci. Math. Educ. 19, 639–659 (2021)
25. Worldwide statistical data provided by the United Nations // UN Electronic Portal // Electronic resource // URL: <http://data.un.org/>
26. World Bank Open Data // UN Electronic Portal // Electronic resource // URL: <https://data.worldbank.org>

МАТЕМАТИЧЕСКИЕ, СТАТИСТИЧЕСКИЕ
И ИНСТРУМЕНТАЛЬНЫЕ МЕТОДЫ В ЭКОНОМИКЕИспользование методов машинного обучения
при обработке данных: практические аспекты

Степанович Кристина Сергеевна
магистр РЭУ им. Г. В. Плеханова
E-mail: christy-stepanovich@mail.ru

Булгаков Андрей Леонидович
к. э. н., доцент МГУ имени М. В. Ломоносова;
Доцент РЭУ им. Г. В. Плеханова
Профессор Московского института
современного академического образования
E-mail: z3900207@mail.ru

Благодаря цифровизации и развитию методов машинного обучения многие процессы и алгоритмы стали осуществляться быстрее и точнее, без участия человека. Современные цифровые технологии находят применение в различных сферах, включая экологию, экономику, медицину, банковскую отрасль, транспорт и маркетинг. Методы машинного обучения позволяют анализировать большие объемы данных, выявлять скрытые закономерности и повышать качество решений. В экологии машинное обучение используется для анализа состояния окружающей среды и разработки методов решения экологических проблем. В экономике и маркетинге технологии помогают создавать эффективные инструменты анализа потребительского поведения, прогнозировать спрос и оценивать кредитный потенциал. В медицине методы машинного обучения способствуют точной диагностике, прогнозированию рисков заболеваний и персонализации лечения. В транспортной отрасли они улучшают управление трафиком и повышают безопасность. Модели машинного обучения решают задачи классификации, регрессии и кластеризации. Их успешное применение требует качественных данных, а также знаний в математике, статистике и программировании. В данной работе особое внимание уделяется использованию моделей машинного обучения для решения задачи регрессии, демонстрируя их потенциал в автоматизации сложных процессов и улучшении качества жизни.

Ключевые слова: машинное обучение, обработка данных, логистическая регрессия, линейная регрессия, датасет.

Современные темпы развития цифровизации диктуют новую реальность в деятельности многих сфер. Новые технологии, основанные на них инструменты и методы, ускоряют многие процессы, являются отличной заменой человеческим ресурсам. Благодаря этому многие рутинные процессы в разных сферах стали совершаться быстрее и качественнее. Множество современных инструментов на сегодняшний день отлично справляются с функцией обработки данных и построения прогнозов. Машинное обучение делает это качественно и быстро, избавляя человека от долгих рутинных подсчетов и анализов больших объемов информации. На сегодняшний день в сфере экологии, экономики и медицины есть интересные практики в части использования методов машинного обучения.

Так, в медицине сведения о множестве пациентов (его пол, рост, вес, наличие гипертонии, сердечных заболеваний, уровень глюкозы, индекс массы тела, наличие или отсутствие инсульта) в одном эксперименте использовали для занесения в инструменты машинного обучения. После обработки множества данных, среди которых было немало пациентов, перенесших инсульт, программа обучилась предсказывать вероятность риска инсульта. В том числе и относительно многих других болезней программа обучалась предсказывать вероятность болезни, что помогало врачам более точно определиться с диагнозом. То есть программа выявляла вероятность методом построения корреляционной связи между разными переменными базы данных со сведениями о пациентах. Для обучения использовался

метод логической регрессии [1], который представляет собой «способ определения зависимости между переменными, одни из которых категориально зависимы (инсульт), а другие независимы (индекс массы тела, уровень сахара в крови, возраст)» [2].

В экономике также можно встретить множество примеров применения машинного обучения. По мнению многих специалистов при осуществлении предпринимательской деятельности машинное обучение позволяет прийти к большему объему прибыли благодаря своему эффективному и мобильному инструментарию [3,26]. Так, в маркетинге существуют технологии, с помощью машинного обучения анализируются огромные массивы данных произвольного формата в четкий численный анализ о параметрах, предпочтениях и мнениях того или иного потребителя конкретного продукта — иными словами, формируется портрет потребителя [4]. Отличных результатов удалось добиться и в части формирования «персонализированных предложений с помощью методов машинного обучения. Программа собирает данные из разных источников, анализирует поведение на сайте, отзывы, совершенные покупки и благодаря этому разделяет клиентов на условные группы, формируя для каждой индивидуальное предложение» [5]. Торговая сеть «Перекресток» благодаря внедрению машинного обучения «учитывает множество поведенческих факторов: частоту и сумму покупок, приемлемые цены, любимые категории товаров, что позволяет ей увеличить эффективность целевого маркетинга на 5% и сократить затраты на коммуникации почти на половину» [6].

В банковском секторе также активно применяются методы машинного обучения, что позволяет, например, оперативно получить отчетность или оценить кредитоспособность заемщика. На основе анализа всех данных о выручке и доходах клиента, выплачиваемых налогах и штрафах банки могут оценить риски в кредитовании. Анализ с помощью машинного обучения позволяет производить обработку более 80 000 запросов в сутки, что значительно ускоряет процесс принятия решения и сни-

жает риск выдачи потребительских кредитов неплатежеспособным лицам [7].

Больших результатов благодаря применению методов машинного обучения удалось достичь и в сфере экологии. Так, уже на государственном уровне применяется система экологического мониторинга, когда на основании данных из разных источников делается вывод об экологической ситуации в конкретной местности — в частности контролируются параметры атмосферы, литосферы, биосферы, гидросферы и криосферы. При этом данные собираются совершенно разными методами и далее объединяются воедино для анализа и прогноза. В случае осуществления такого анализа человеком процесс бы занял очень долгое время. Например, данные собираются с помощью дистанционного зондирования (наблюдения за окружающей средой «без непосредственного контакта с помощью спутников и бортовых датчиков» [8], сенсорных технологий, находящихся в непосредственной близости от объекта наблюдения.

Данный рискоориентированный подход используется также и для определения эффективных поглотителей углекислого газа. Ученые из Ивановского Института химии растворов имени Г. А. Крестова РАН разработали «алгоритм машинного обучения, способный предсказать, насколько хорошо глубокий эвтектический растворитель того или иного состава будет поглощать углекислый газ» [9]. Машинное обучение оперативно оценило различные свойства множество жидкостей (их химический состав, строение молекул, температуру плавления). На основе анализа данных, содержащихся в нескольких сотнях смесей, ученые сделали прогноз о тех смесях, которые наиболее эффективно смогут поглощать углекислый газ.

Также модели машинного обучения могут обрабатывать большие объемы данных и выявлять в них закономерности, которые не могут быть обнаружены человеком. Например, модели машинного обучения используются в маркетинге для анализа поведения потребителей и прогнозирования спроса на товары и услуги [20].

Модели машинного обучения могут улучшать качество жизни людей. Они используются в медицине для создания персонализированных лечебных планов и управления хроническими заболеваниями. Также могут помочь в создании более удобных и безопасных транспортных средств [26–29].

Машинное обучение может решать различные типы задач, такие как классификация, регрессия и кластеризация [16]. Для работы с машинным обучением необходимы хорошие данные, на которых будет обучаться алгоритм, а также знания в области математики, статистики и программирования [12].

Далее анализируется применение моделей машинного обучения для задачи регрессии.

Постановка задачи

1. Определение подходящей для обработки данных базы данных (Dataset).
2. Определение методов машинного обучения, которые будут использоваться.
3. Проведение обработки базы данных с использованием методов машинного обучения.
4. Получение результатов и подведение итогов анализа данных с использованием методов машинного обучения.

База данных для анализа (датасет)

Анализ проводился на примере российской базы данных «Очаги распространения голландской болезни вязов», расположенного на сайте <https://classif.gov.spb.ru> [30].

В датасете представлены данные по Санкт-Петербургу по заболеванию вязов графиейзом ильмовых (голландской болезнью).

Графиейз ильмовых представляет собой болезнь деревьев, от которой погибают вязы. Дерево вяз ослабевает и может рухнуть в любой момент. Графиейз заразен и быстро распространяется на другие деревья этого вида. В Санкт-Петербурге данное заболевание впервые было зарегистрировано в 1995 году.

Эффективных мер лечения болезни не существует. Погибает 100 процентов деревьев, пораженных острой формой графиейза. Единственная мера по сохранению оставшихся деревьев — удаление пораженных вязов.

Под угрозой каждое третье дерево. Полагают, что во всем виновато изменение климата.

Переносчики графиейза — жуки короеды, которые разносят грибок. Раньше большинство насекомых не могло пережить холодную зиму. Теперь, когда стало теплее, численность вредителей увеличилась, они распространяют заболевание гораздо быстрее. Сейчас уже можно говорить о полномасштабной эпидемии. Больше всего случаев графиейза фиксируют в Калининском, Московском и Василеостровском районе.

Чтобы хоть как-то замедлить распространение грибка, пораженные вязы в Петербурге спиливают, заменяя их другими деревьями: липами, ивами, кленами и рябинами.

С 2016 по 2020 год в городе спилили 16359 вязов, наибольшее количество в Московском (1933 шт.), в Красносельском (1599 шт.) и в Калининском (1532 шт.) районах [30]. На всех территориях проведены мероприятия по компенсационным посадкам.

В Нидерландах, где эту болезнь открыли, уже научились бороться с недугом. Лекарства от графиейза там не придумали, но распространение грибка голландцы успешно предупреждают с помощью вакцинации деревьев и регулярного мониторинга их состояния.

Вязы болеют и умирают — попробуем сделать модель, которая бы предсказывала, какая будет площадь поражения, чтобы можно было вовремя помочь деревьям.

Возможности использования математических моделей для анализа данных

Линейная регрессия представляет собой модель зависимости одной переменной y (называемой объясняемой или зависимой)

от одной или нескольких других переменных (называемых факторами, регрессорами или независимыми переменными), где зависимость описывается линейной функцией [16, 19].

Модели линейной регрессии относительно просты и предоставляют легко интерпретируемую математическую формулу для создания прогнозов.

Дерево решений представляет собой инструмент поддержки принятия решений, использующийся в машинном обучении, анализе данных и статистике [16]. Структура дерева состоит из «листьев» и «веток». На ветках дерева записаны признаки, от которых зависит целевая функция, в листьях записаны значения целевой функции, а в остальных узлах — признаки, по которым различаются случаи.

Для проведения анализа данных будут использованы модели линейной регрессии и дерева принятия решений.

Преобразование категориальных признаков для использования в модели

Линейным моделям для корректной работы, в качестве входных данных, нужно передавать только численные признаки. Поэтому, необходимо преобразовывать категориальный признак в некоторое численное представление. Примеры категориальных признаков: пол, страна проживания, номер группы, категория товаров и т. п.

Самый простой способ — нумерация значений (0, 1, 2, ...). У данного подхода есть недостаток. Обычно он ведет к плохому результату, так как алгоритмы начинают учитывать бессмысленную упорядоченность значений признаков. Однако, данный метод имеет преимущество с точки зрения памяти.

Метод реализован в классе `sklearn.preprocessing.LabelEncoder` [21].

Для применения модели линейной регрессии важно оценить качество построенной модели. Для этого используют ряд показателей, которые имеют свои особенности применения [15].

Возможные метрики:

- Среднеквадратичная ошибка (MSE).
- Средняя абсолютная ошибка (MAE).
- Средняя абсолютная ошибка в процентах (MAPE).
- Коэффициент детерминации R-квадрат.
- Скорректированный коэффициент детерминации.
- Корень из среднеквадратичной ошибки (RMSE).
- Среднеквадратичная ошибка в процентах (MSPE).
- Симметричная средняя абсолютная процентная ошибка (SMAPE).
- Средняя абсолютная масштабированная ошибка (MASE)
- Средняя относительная ошибка (MRE).
- Среднеквадратичная логарифмическая ошибка (RMSLE).

Наиболее часто используют две метрики для количественной оценки соответствия модели набору данных:

1. Среднеквадратическая ошибка (RMSE).
2. R-квадрат (R^2).

RMSE показывает типичное расстояние между прогнозируемым и фактическим значениями.

R-квадрат показывает, насколько хорошо переменные-предикторы могут объяснить изменение переменной отклика.

RMSE — метрика, сообщающая квадратный корень из средней квадратичной разницы между прогнозируемыми значениями и фактическими значениями в наборе данных. Она показывает типичное расстояние между данными значениями. Чем ниже RMSE, тем лучше модель соответствует набору данных.

R-квадрат — метрика, сообщающая долю дисперсии переменной отклика регрессионной модели, которая может быть объяснена предикторными переменными. Значение находится в диапазоне от 0 до 1. Чем выше значение R-квадрат, тем лучше модель соответствует набору данных.

Для проведения исследования использовался ряд готовых библиотек Python.

Используемые библиотеки Python для анализа данных

Pandas (<https://pandas.pydata.org/>) — библиотека Python для обработки и анализа структурированных данных [17].

Matplotlib (<https://matplotlib.org/>) — библиотека на языке Python для визуализации данных двумерной и трёхмерной графикой [18].

Scikitlearn (<https://scikit-learn.org/>) — библио-

тека Python, включающая все алгоритмы и инструменты, нужные для задач классификации, регрессии и кластеризации. Также включает все методы оценки производительности модели машинного обучения [21].

Проведение исследования данных

Импортируем библиотеки (см. рис. 1).

```
#Библиотека для обработки и анализа данных. https://pandas.pydata.org
import pandas as pd
#Библиотека машинного обучения с открытым исходным кодом. https://scikit-learn.org
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.metrics import mean_squared_error, r2_score
#Библиотека визуализации данных. https://matplotlib.org
import matplotlib.pyplot as plt
#Модуль работы с числами math
import math
```

Рис. 1. Загрузка базовых библиотек Python (составлено авторами)

Из pandas возьмем метод `read_csv`, он позволит прочитать csv файл. Укажем, что у нас есть url и считывание произойдет с него.

Проверим, записалась ли датафрейм в файле, вызовем для этого метод `head` (см. рис. 2 и 3) [10–13].

```
url = "https://raw.githubusercontent.com/KristinaStepanovich/Project/main/The_nidus_of_Dutch_elm_disease.csv"
df = pd.read_csv(url)
df.head()
#Просмотр датафрейма.
```

Рис. 2. Скрипт, описывающий вывод набора данных (составлено авторами)

	Порядковый номер	Идентификационный номер	Классификация очагов	Год обследования	Периметр территории очагов,м	Площадь территории очагов,м2
0	1464	1	Действующий очаг	2017	17.840530	11.855066
1	1465	2	Возникающий очаг	2017	47.601830	175.242859
2	1466	3	Действующий очаг	2017	48.658123	95.710403
3	1467	4	Действующий очаг	2017	239.168655	595.867127
4	1468	5	Без признаков поражения	2016	1829.541016	3685.098633

Рис. 3. Вывод пяти первых элементов скачиваемого набора данных (составлено авторами)

Делаем проверку на наличие отрицательных значений (см. рис. 4).

```
#Проверяем отрицательные значения
df[df['Площадь территории очагов, м2'] < 0]
```

	Порядковый номер	Идентификационный номер	Классификация очагов	Год обследования	Периметр территории очагов, м	Площадь территории очагов, м2
200	1699	201	Действующий очаг	2017	94.086380	-293.949188
201	1700	202	Действующий очаг	2017	196.579117	-1412.491821
392	1954	393	Затухающий очаг	2017	213.583282	-6341.792969
411	2007	412	Действующий очаг	2017	79.326340	-397.736572
423	2019	424	Затухающий очаг	2017	593.938660	-25939.234375
...	...	...	...	...	...	...
2218	3853	2219	Действующий очаг	2017	205.306900	-2012.946777
2221	3856	2222	Действующий очаг	2017	417.890320	-36852.847656
2233	3868	2234	Без признаков поражения	2017	291.452728	-1303.706543
2236	3871	2237	Без признаков поражения	2017	177.443848	-411.102264
2238	3873	2239	Без признаков поражения	2017	155.339767	-105.151955

Рис. 4. Вывод данных, проверяемых на отрицательные значения (составлено авторами)

В столбце «Площадь территории очагов, м2» имеются отрицательные значения, уберем их.

```
#Убираем отрицательные значения
#Конвертер вернет модуль числа
converter = lambda x: math.fabs(x)
df['Площадь территории очагов, м2']
= df['Площадь территории очагов, м2'].apply(converter)
```

Вновь сделаем проверку на наличие отрицательных значений (см. рис. 5).

```
#Проверка на отрицательные значения
#Теперь отрицательные значения отсутствуют
df[df['Площадь территории очагов, м2'] < 0]
```

Порядковый номер	Идентификационный номер	Классификация очагов	Год обследования	Периметр территории очагов, м	Площадь территории очагов, м2
------------------	-------------------------	----------------------	------------------	-------------------------------	-------------------------------

Рис. 5. Скрипт вывода повторной проверки на наличие пропусков и отрицательных значений (составлено авторами)

Делаем проверку на наличие пустых пропущенных NaN значений, т. к. они могут повлиять на качество обучения модели [22,23]. Вызываем метод df.info, который нам выводит информацию о нашем датасете. Соответственно, можем видеть Dtype — тип данных (числовой, строки и т.д.). Можем видеть размер датасета и количество строчек. У нас нет NaN значений,

т. к. везде одинаковое количество строк (см. рис. 6).

```
#Проверка NaN значений
df.info()

#Количество строчек везде одинаково, поэтому NaN значений нет
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2414 entries, 0 to 2413
Data columns (total 6 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Порядковый номер                     2414 non-null   int64
1   Идентификационный номер             2414 non-null   int64
2   Классификация очагов                 2414 non-null   object
3   Год обследования                     2414 non-null   int64
4   Периметр территории очагов,м        2414 non-null   float64
5   Площадь территории очагов,м2        2414 non-null   float64
dtypes: float64(2), int64(3), object(1)
memory usage: 113.3+ KB
```

Рис. 6. Анализ структуры данных сохраненного датасета (составлено авторами)

Далее будем преобразовывать категориальные признаки, т. к. мы делаем регрессию и она на вход предполагает только численные значения [24]. Проверим, категориальный ли признак имеет столбец «Классификация очагов» — вызываем df и в кавычках указываем название столбца, с которым будем работать. Вызываем метод value_counts. У нас получилось всего 6 возможных уникальных значений, значит это категориальный признак. Допустим, если бы было значение больше, к примеру — 50, то это не категориальный признак, т. к. оно слишком большое. В нашем случае столбец «Классификация очагов» имеет текстовое значение и чтобы его использовать, надо преобразовать в числовое значение с помощью LabelEncoder.

#Проверка категориальных признаков —

Классификация очагов

```
df['Классификация очагов'].value_counts()
```

```
Действующий очаг          671
Возникающий очаг          653
Без признаков поражения   628
Не определено              236
Затухающий очаг           137
Вырублено                  89
Name: Классификация очагов, dtype: int64
```

#Преобразование категориальных признаков. Теперь столбец «Классификация очагов» имеет числовое значение.

```
encoder = LabelEncoder()
df['Классификация очагов'] =
encoder.fit_transform(df['Классификация очагов'])
df.head()
```

	Порядковый номер	Идентификационный номер	Классификация очагов	Год обследования	Периметр территории очагов,м	Площадь территории очагов,м2
0	1464	1	3	2017	17.840530	11.855066
1	1465	2	1	2017	47.601830	175.242859
2	1466	3	3	2017	48.658123	95.710403
3	1467	4	3	2017	239.168655	595.867127
4	1468	5	0	2016	1829.541016	3685.098633

Разобьем датасет на тренировочные и тестовые выборки. Создадим переменную X — это будут признаки, столбцы «Классификация очагов» и «Периметр территории

очагов, м». Используем метод values, чтобы их получить в виде списка. Также используем метод loc, он позволяет по названию столбца получить его значение. Указали,

значит берем все строчки из исходного датафрейма. Через запятую указываем названия нужных столбцов. Переменная Y — наша цель. Укажем размер тестовой выборки 20% `test_size=0.2`. Тренировочная выборка будет 80%.

```
#Создание тренировочных и тестовых выборок
```

```
#Признаки столбцы Классификация очагов, Периметр территории очагов, м
X = df.loc[:, ['Классификация очагов', 'Периметр территории очагов, м']].values
Y = df['Площадь территории очагов, м2']
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=25)
```

Рассматриваем линейную регрессию. Обучаем эту модель по тем данным, что уже подготовили. Обучить модель — значит подобрать вес каждого признака. У нас два признака, соответственно параметров будет три. Два веса перед признаками и один свободный член. Создаем экземпляр линейной регрессии и обучаем модель. Для обучения вызовем метод `fit` и передадим его тренировочные данные `X_train` и `Y_train`.

```
#Модель 1 — линейная регрессия
model = LinearRegression()
```

```
#Обучаем модель
model = model.fit(X_train, Y_train)
```

Попробуем сделать предсказание — сделаем `predict` по тестовым данным `X_test`.

```
model.predict(X_test)
array([-9.23489347e+02, 2.47909698e+02, -5.66025924e+02, 1.46122903e+04, 7.00411419e+03, -1.09326279e+03, 4.58596507e+03, -7.90930494e+02, 5.98846458e+02, -3.69223071e+02, 1.93222460e+03, -1.19361809e+03,
```

```
-9.80001096e+01, 1.50079638e+03, 2.91784829e+03, -7.29614325e+02, -7.40074487e+02, -1.02061517e+03, 9.29958316e+02, -1.30604995e+03, -1.31329766e+03, -1.49348247e+03, -1.01920096e+03, 1.57434605e+04, 4.62106390e+03, -1.17770791e+03, -9.38413811e+02, -1.77707526e+02, 4.81877873e+03, -3.23092533e+02, -1.39324941e+03, -6.11083822e+02, -1.71104751e+03, 5.15262641e+03, 1.07599198e+03, 1.47012929e+02, -1.46084407e+03, 2.60023248e+03, -4.68701740e+02, -1.41687520e+03, 4.29167774e+03, 3.61770209e+03, 1.25330344e+03, -9.93528880e+02, -3.56630239e+02, 4.28106953e+03, -7.08706093e+02, -1.58565318e+03, -3.60367600e+02, 5.96279897e+03, -1.02126606e+03, 7.43043855e+03, 5.70061551e+02, 1.25816439e+03, 7.44697492e+03, 1.19463717e+04
```

Для наглядности визуализируем всё. Отобразим на графике один признак и одну цель. Посмотрим, как от классификаций очагов и периметров территории очагов зависит площадь территории (см. рис. 7 и 8).

```
Y_pred = model.predict(X_train)
```

```
#Первый признак — Классификация очагов
```

```
#Не все данные совпали, значит, в среднем, наша модель предсказывает верно и она не переобучилась.
```

```
plt.scatter(X_train[:,0], Y_train, label='Исходные данные')
plt.plot(X_train[:,0], Y_pred, 'ro', label='Предсказанные данные')
plt.xlabel('Классификация очагов')
plt.ylabel('Площадь территории очагов, м2')
plt.legend()
plt.show()
```

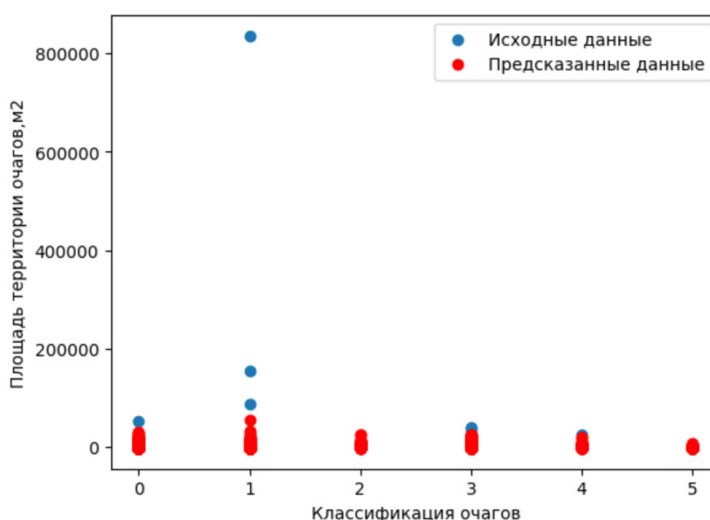


Рис. 7. Распределение заболеваемости вязов (составлено авторами)

#Второй признак – Периметр территории очагов

#Получилась практически прямая красная линия – это наша регрессия. Она усредняет значения, не переобучилась, есть некоторые выбросы.

```
plt.scatter(X_train[:,1], Y_train,
            label='Исходные данные')
```

```
plt.plot(X_train[:,1], Y_pred, 'ro',
         label='Предсказанные данные')
plt.xlabel('Периметр территории очагов, м')
plt.ylabel('Площадь территории очагов, м2')
plt.legend()
plt.show()
```

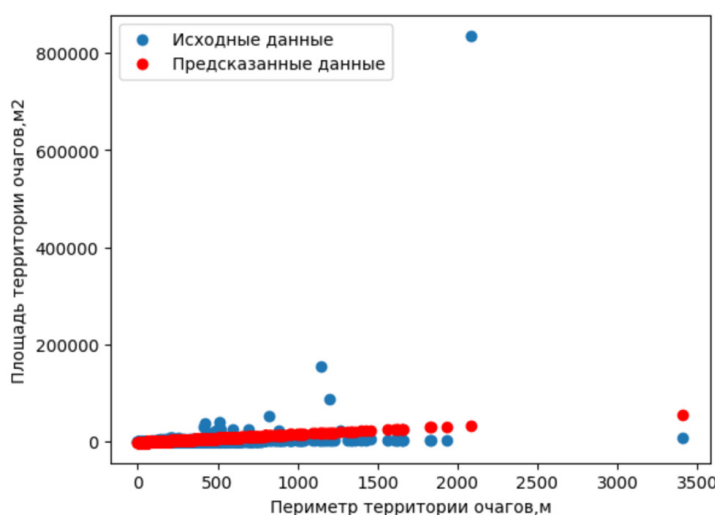


Рис. 8. График распределения заболеваемости вязов (составлено авторами)

Теперь сделаем оценку с помощью метрик, это позволит численно оценить. Будем использовать метрику *rmse*, она позволяет находить

усредненное значение. $rmse = \sqrt{1 / n * \sum (y_{pred} - y)^2}$, где *n* – количество признаков, умножается на сумму разностей предсказанного

и исходного значений, и делится на количество использованных строчек. Из полученного извлечем корень. Также используем метрику r^2 — коэффициент детерминации. $r^2 = 1 - \frac{\sum(y_{\text{pred}} - y)^2}{\sum(y_{\text{sred}} - y)^2}$, значения по аналогии.

Коэффициент детерминации позволяет вычислять, насколько наш признак зависит от своей дисперсии (отклонения значения).

```
# rmse - sqrt(1 / n * sum(y_pred - y)^2)
# r2 - 1 - sum(y_pred - y)/sum(y_sred - y) коэффициент детерминации

# rmse -> 0 должно стремиться к 0
# r2 -> 1 должно стремиться к 1

#Тестовая выборка
#Получились достаточно неплохие значения. Выбросы не удалены, поэтому значение rmse достаточно большое.
Y_pred_test = model.predict(X_test)
rmse = math.sqrt(mean_squared_error(Y_test, Y_pred_test))
r2 = r2_score(Y_test, Y_pred_test)

print(f"rmse = {rmse}, r2 = {r2}")
```

```
rmse = 10838.057269046842, r2 =
0.22413877106382374
```

Рассмотрим вторую модель — дерево принятия решений. Используем `model_tree` дерева, импортируем класс `DecisionTreeRegressor`, так как задача регрессии (см. рис. 9 и 10).

```
#Модель 2 - дерево принятия решений без параметров
model_tree = DecisionTreeRegressor()
model_tree = model_tree.fit(X_train, Y_train)

Y_pred = model_tree.predict(X_train)

#Первый признак - Классификация очагов
plt.scatter(X_train[:,0], Y_train, label='Исходные данные')
plt.plot(X_train[:,0], Y_pred, 'ro', label='Предсказанные данные')
plt.xlabel('Классификация очагов')
plt.ylabel('Площадь территории очагов, м2')
plt.legend()
plt.show()
#Модель переобучена
```

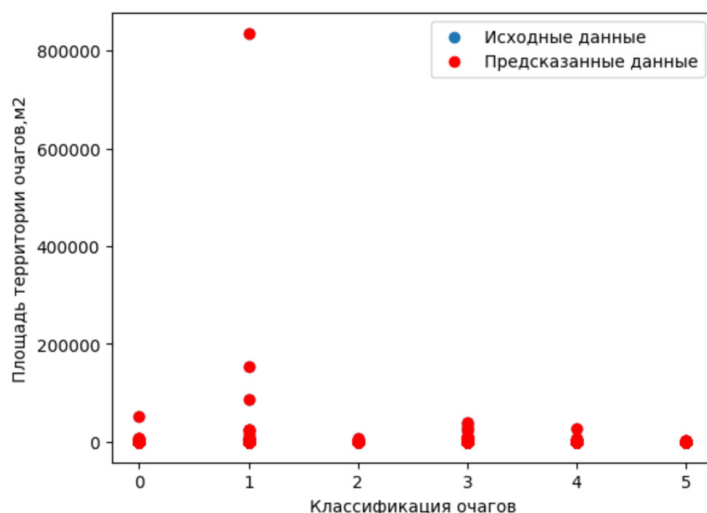


Рис. 9. Визуализация использования модели — дерево принятия решений

(составлено авторами)

```
#Второй признак – Периметр террито-
рии очагов
```

```
plt.scatter(X_train[:,1], Y_train,
label='Исходные данные')
plt.plot(X_train[:,1], Y_pred, 'ro',
label='Предсказанные данные')
```

```
plt.xlabel('Периметр территории оча-
гов, м')
```

```
plt.ylabel('Площадь территории оча-
гов, м2')
```

```
plt.legend()
```

```
plt.show()
```

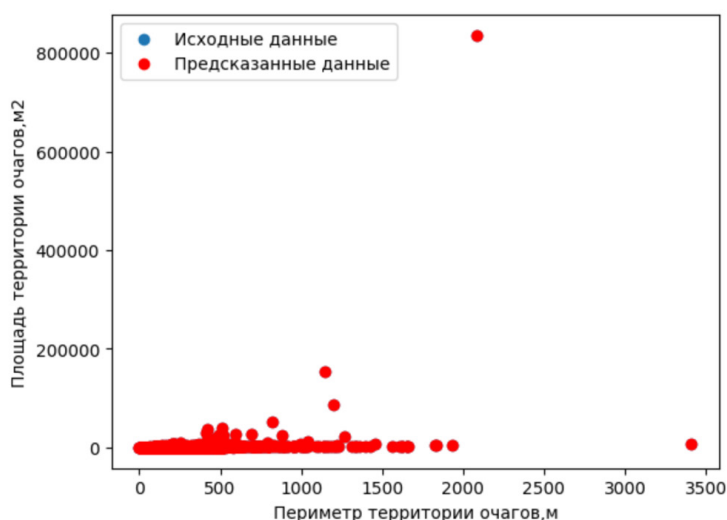


Рис. 10. Визуализация использования модели — дерево принятия решений
(составлено авторами)

Модель 2 переобучена, ограничим глубину дерева, этого будет достаточно для регрессии. Возьмём целое значение 5.

```
#Модель 2 – дерево принятий реше-
ний с ограничением по глубине, max_
depth=5
```

```
model_tree =
DecisionTreeRegressor(max_depth=5)
model_tree = model_tree.fit(X_train,
Y_train)
```

```
#Делаем предсказание
```

```
model_tree.predict(X_train)
array([83.3305007, 83.3305007,
83.3305007, ..., 1021.95105064,
83.3305007, 83.3305007])
```

Построим графики по первому и второ-
му признаку (см. рис. 11 и 12).

```
Y_pred = model_tree.predict(X_train)
```

```
#Первый признак – Классификация очагов
```

```
plt.scatter(X_train[:,0], Y_train,
label='Исходные данные')
```

```
plt.plot(X_train[:,0], Y_pred, 'ro',
label='Предсказанные данные')
```

```
plt.xlabel('Классификация очагов')
plt.ylabel('Площадь территории оча-
гов, м2')
```

```
plt.legend()
```

```
plt.show()
```

```
#Второй признак – Классификация очагов
```

```
plt.scatter(X_train[:,1], Y_train,
label='Исходные данные')
```

```
plt.plot(X_train[:,1], Y_pred, 'ro',
label='Предсказанные данные')
```

```
plt.xlabel('Периметр территории оча-
гов, м')
```

```
plt.ylabel('Площадь территории оча-
гов, м2')
```

```
plt.legend()
```

```
plt.show()
```

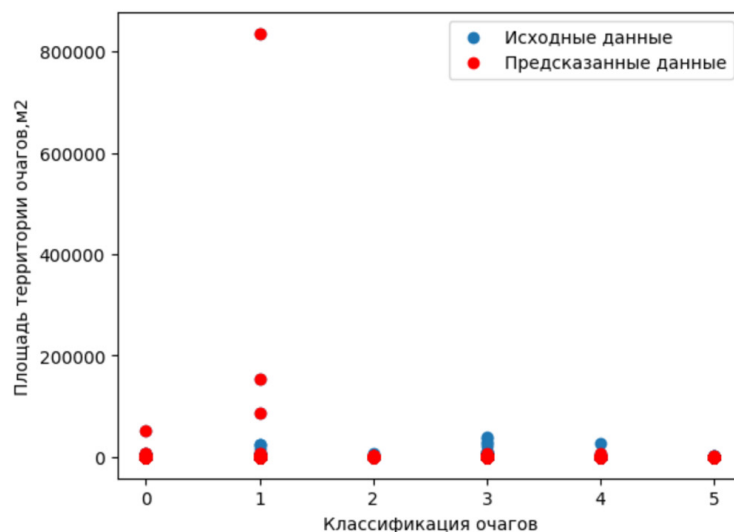


Рис. 11. Обобщенный график заболеваемости вязов по двум признакам (составлено авторами)

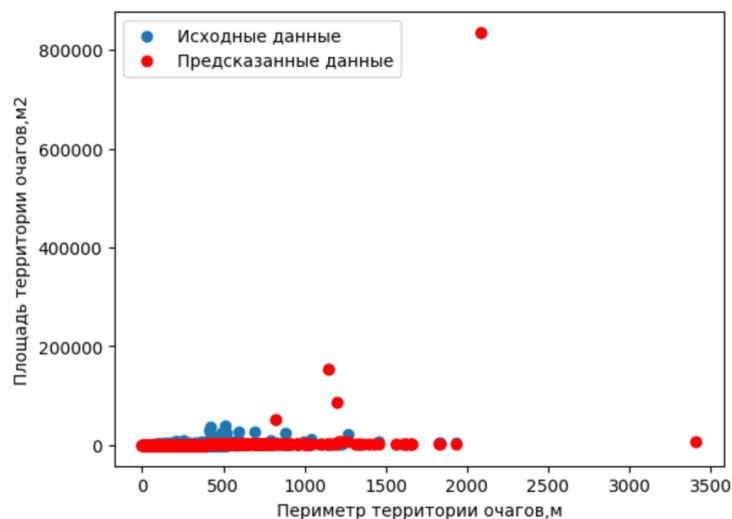


Рис. 12. Линейный график заболеваемости вязов по двум признакам (составлено авторами)

Посчитаем метрики для непереобученной модели. Нужно, чтобы они не были слишком идеальными. $r^2 = -15$, отрицательное значение, что очень плохо. И $rmse$ тоже гораздо больше получилась.

#Тестовая выборка

```
Y_pred_test = model_tree.predict(X_test)
rmse = math.sqrt(mean_squared_error(Y_test, Y_pred_test))
r2 = r2_score(Y_test, Y_pred_test)
```

```
print(f»rmse = {rmse}, r2 = {r2}»)
rmse = 49635.896926922425, r2
=-15.27319823967644
```

Мы сравнили линейную регрессию и дерево принятия решений. Для конкретного набора данных, модель линейной регрессии оказалась лучше. Потому что у неё метрики лучше и она не переобучается, как дерево принятия решений. Дерево принятия решений вообще в данном случае не подходит, т. к. метрика r^2 имеет отрицательное значение.

Вывод

Проведено практическое исследование использования методов машинного обучения при обработке данных.

В рамках выполненной работы, на сайте <https://classif.gov.spb.ru> была найдена российская база данных с соответствующими критериями (датасет) «Очаги распространения голландской болезни вязов». База данных была преобразована, в том числе были изменены категориальные признаки в численные. Была убрана отрицательность признаков. Были опробованы две модели машинного обучения — модель линейной регрессии и дерево принятия решений для одной и той же задачи. Для них рассчитаны метрики и нарисованы графики.

Для данного датасета лучшей оказалась модель линейной регрессии, так как у неё лучше метрики, и она не переобучается, как дерево принятия решений.

Полученные результаты исследования размещены на платформе GitHub [31].

Литература

1. Рекурсивный алгоритм построения регрессионных моделей сложных вероятностных объектов / Мокшин В. В., Сайфудинов И. Р., Кирпичников А. П. Вестник Технологического университета. 2017. Т. 20. № 9. С. 112–116 // Электронный ресурс // URL: <https://cyberleninka.ru/article/n/rekursivnyy-algoritm-postroeniya-regressionnyh-modeley-slozhnyh-veroyatnostnyh-obektov>.
2. Якимов И. М. Структурное моделирование бизнес-процессов в системах BPMN EDITOR, ELMA, RUNAWFE /Якимов И.М., Кирпичников А. П., Мокшин В. В., Махмутов М. Т., Пейсахова М. Л., Валиева А. Х., Низамиев Б. А.// Вестник Казанского технологического университета. 2014. Т. 17. № 10. С. 249–256 // Электронный ресурс // URL: <https://cyberleninka.ru/article/n/mashinnoe-obuchenie-v-biznese/viewer>
3. Кораблев А. Ю., Булатов Р. Б. Машинное обучение в бизнесе // Азимут научных исследований: экономика и управление. 2018. Т. 7. № 2(23).
4. Лобин М. А., Филиппова И. А. Машинное обучение в экономике // Вестник УлГТУ 3 (2019) // Электронный ресурс // URL: <https://cyberleninka.ru/article/n/mashinnoe-obuchenie-v-ekonomike>
5. 4 практических совета по применению машинного обучения в маркетинге // Электронный ресурс // URL: <http://datareview.info/article/4-prakticheskix-soveta-po-primeneniyu-mashinnogo-obucheniya-v-marketinge/>
6. X5 автоматизировала целевой маркетинг в «Перекрестке» с помощью технологии машинного обучения // Электронный ресурс // URL: <https://www.vedomosti.ru/press-releases/2017/11/01/x5-avtomatizirovalatsелеvoy-marketing-v-perekrestke-s-pomoschyu-tehnologiiimashinnogo-obucheniya>
7. Нейросети для транзакций: как на деле работают «большие данные» в российских банках?// Электронный ресурс//URL: <http://www.forbes.ru/tehnologii/341517-neyroseti-dlya-tranzakciy-kak-na-delerabotayut-bolshie-dannye-v-rossijskih-bankah>
8. Lychak, A. I., Bobra T. V. GIS-maintenance of environmental management of territories // Геополитика и экогеодинамика регионов. 2011. Т. 7. № 1–2. С. 126–138.
9. Алгоритм машинного обучения позволит вычислить эффективные поглотители углекислого газа [Электронный ресурс] Экология производства: научно-практический портал // Электронный ресурс // URL: <https://www.ecoindustry.ru/news/view/62922.html>
10. Постолита А. Python, Django и PyCharm для начинающих // БХВ, 2021. — 465с
11. Руководство по Python // Электронный ресурс // URL: <https://metanit.com/python/tutorial/>
12. Материалы из Википедии о Машинном обучении// Электронный ресурс // URL: https://ru.wikipedia.org/wiki/Машинное_обучение
13. Гвидо, Мюллер: Машинное обучение с помощью Python. Руководство для специалистов по работе с данными // Вильямс, 2022. — 480с

14. ГОСТ 7.32-2017 Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления. М.: Стандартинформ, 2018. — 33 с.
15. Метрики и оценки: количественная оценка качества прогнозов // Электронный ресурс // URL: <https://scikit-learn.ru/3-3-metrics-and-scoring-quantifying-the-quality-of-predictions/>
16. LinearRegression // Электронный ресурс // URL: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
17. Pandas // Электронный ресурс // URL: <https://pandas.pydata.org/>
18. Matplotlib // Электронный ресурс // URL: <https://matplotlib.org/>
19. Курс «Введение в машинное обучение» // Электронный ресурс // URL: <https://www.kaggle.com/learn/intro-to-machine-learning>
20. Платонов, А. В. Машинное обучение: учебное пособие для вузов / А. В. Платонов. — Москва: Издательство Юрайт, 2023. — 85 с.
21. О. Жерон «Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow» // Диалектика, 2020 г. — 1040 с.
22. С. Рашка «Python и машинное обучение» // Диалектика, 2020 г. — 848 с.
23. Ф. Шолле «Глубокое обучение на Python» // Питер, 2018 г. — 400 с.
24. Н. Гифт «Прагматичный ИИ» // Питер, 2019 г. — 304 с.
25. Н. Шакла «Машинное обучение и TensorFlow» // Питер, 2019 г. — 336 с.
26. Х. Брикс, Дж. Ричардс, М. Феверолф «Машинное обучение» // Питер, 2023 г. — 336 с.
27. А. Бурков «Машинное обучение без лишних слов» // Питер, 2020 г. — 192 с.
28. П. Домингос «Верховный алгоритм» // Манн, Иванов и Фербер, 2016 г. — 336 с.
29. Шалев-Шварц, Бен-Давид: Идеи машинного обучения // ДМК-Пресс, 2019 г. — 436 с.
30. Система классификаторов и открытых данных Санкт-Петербурга // Электронный ресурс // URL: <https://classif.gov.spb.ru>.
31. Результаты анализа данных проекта // Электронный ресурс // URL: <https://github.com/KristinaStepanovich/Project> [31].

Using Machine Learning Methods in Data Processing: Practical Aspects

Stepanovich K. S., Bulgakov A. L.

Lomonosov Moscow State University, Plekhanov Russian University of Economics, Moscow Institute of Modern Academic Education

Machine learning (ML) models are widely used in various fields, automating complex processes, processing large amounts of data and improving the quality of people's lives. They are used in the banking industry to detect fraud, in medicine — for diagnostics and personalization of treatment, in transport — for traffic management and improving safety. ML is also actively used in marketing to analyze consumer behavior and forecast demand. Machine learning models solve problems of classification, regression and clustering, identifying patterns that are inaccessible to human perception. For their successful application, a high-quality data set is required, as well as knowledge of mathematics, statistics and programming. The goal of this work is to use machine learning models to solve a regression problem.

Keywords: machine learning, data processing, logistic regression, linear regression, dataset

References

1. Recursive algorithm for constructing regression models of complex probabilistic objects / Mokshin V. V., Saifutdinov I. R., Kirpichnikov A. P. Bulletin of the Technological University. 2017. Vol. 20. No. 9. Pp. 112–116 // Electronic resource // URL: <https://cyberleninka.ru/article/n/rekursivnyy-algoritm-postroeniya-regressionnyh-modeley-slozhnyh-veroyatnostnyh-obektov>
2. Yakimov I. M. Structural modeling of business processes in the BPMN EDITOR, ELMA, RUNAWFE systems / Yakimov I. M., Kirpichnikov A. P., Mokshin V. V., Makhmutov M. T., Peisakhova M. L., Valieva A. Kh., Nizamiyev B. A. // Bulletin of the Kazan Technological University. 2014. Vol. 17. No. 10. Pp. 249–256 // Electronic resource // URL: <https://cyberleninka.ru/article/n/mashinnoe-obucheniye-v-biznese/viewer>
3. Korablyov A. Yu., Bulatov R. B. Machine learning in business // Azimuth of scientific research: economics and management. 2018. Vol. 7. No. 2(23).
4. Lobin M. A., Filippova I. A. Machine learning in economics // Bulletin of USTU 3 (2019) // Electronic resource // URL: <https://cyberleninka.ru/article/n/mashinnoe-obucheniye-v-ekonomike>
5. 4 practical tips for applying machine learning in marketing [Electronic resource] // Electronic resource // URL: <http://datareview.info/article/4-prakticheskix-soveta-po-primeneniyu-mashinnogo-obucheniya-v-marketinge/>
6. X5 automated target marketing in // Electronic resource // URL: <https://www.vedomosti.ru/press-releases/2017/11/01/x5-avtomatiziroval-atsevoi-marketing-v-perekrestke-s-pomoschyu-tehnologii-mashinnogo-obucheniya>
7. Neural networks for transactions: how do «big data» actually work in Russian banks? // Electronic resource // URL: <http://www.forbes.ru/tehnologii/341517-neyroseti>

- dlya-tranzakciy-kak-na-delerabotayut-bolshie-dannye-v-rossiyskih-bankah
8. Lychak, A. I., Bobra T. V. GIS-maintenance of environmental management of territories // Geopolitics and ecogeodynamics of regions. 2011. Vol. 7. No. 1–2. Pp. 126–138.
 9. Machine learning algorithm will allow calculating effective carbon dioxide absorbers [Electronic resource] Ecology of production: scientific and practical portal // Electronic resource // URL: <https://www.ecoindustry.ru/news/view/62922.html>
 10. Postolita A. Python, Django, and PyCharm for Beginners // BHV, 2021. — 465 p
 11. Python Guide // Electronic resource // URL: <https://metanit.com/python/tutorial/>
 12. Wikipedia materials on Machine learning // Electronic resource // URL: https://ru.wikipedia.org/wiki/Машинное_обучение
 13. Guido, Müller: Machine learning with Python. A guide for data scientists // Williams, 2022. — 480 p
 14. GOST 7.32–2017 System of standards for information, librarianship and publishing. Report on research work. Structure and rules of registration. M.: Standartinform, 2018. — 33 p.
 15. Metrics and evaluations: quantitative assessment of the quality of predictions // Electronic resource // URL: <https://scikit-learn.ru/3-3-metrics-and-scoring-quantifying-the-quality-of-predictions/>
 16. LinearRegression // Electronic resource // URL: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
 17. Pandas // Electronic resource // URL: <https://pandas.pydata.org/>
 18. Matplotlib // Electronic resource // URL: <https://matplotlib.org/>
 19. Course «Introduction to machine learning» // Electronic resource // URL: <https://www.kaggle.com/learn/intro-to-machine-learning>
 20. Platonov, A. V. Machine learning: a tutorial for universities / A. V. Platonov. — Moscow: Yurait Publishing House, 2023. — 85 p.
 21. O. Zheron «Applied Machine Learning with Scikit-Learn and TensorFlow» // Dialectics, 2020. — 1040 p.
 22. S. Rashka «Python and Machine Learning» // Dialectics, 2020. — 848 p.
 23. F. Chollet «Deep Learning in Python» // Piter, 2018. — 400 p.
 24. N. Gift «Pragmatic AI» // Piter, 2019. — 304 p.
 25. N. Shakla «Machine Learning and TensorFlow» // Piter, 2019. — 336 p.
 26. H. Brix, J. Richards, M. Feverolf «Machine Learning» // Peter, 2023. — 336 p.
 27. A. Burkov «Machine Learning Without Further Worth» // Peter, 2020. — 192 p.
 28. P. Domingos «The Master Algorithm» // Mann, Ivanov and Ferber, 2016. — 336 p.
 29. Shalev-Schwartz, Ben-David: Machine Learning Ideas // DMK-Press, 2019. — 436 p.
 30. The system of classifiers and open data of St. Petersburg // Electronic resource // URL: <https://classif.gov.spb.ru>
 31. Project data analysis results // Electronic resource // URL: <https://github.com/KristinaStepanovich/Project> [31].

Повышение эффективности промышленных предприятий в условиях санкций

Дмитриева Людмила Владиславовна

к. э. н., первый заместитель Председателя
Правительства Ивановской области,
доцент ФГБОУ ВО «Ивановский
государственный университет»
E-mail: ludmilavd@yandex.ru

Алешина Анна Валентиновна

к. э. н., доцент, экономический факультет МГУ
имени М. В. Ломоносова, г. Москва,
Российская Федерация
E-mail: annaaaleshina@mail.ru

Обострение политической, военной и экономической ситуации в мире после введения США и рядом крупных западных держав санкций против России негативно отразилось на развитии промышленных предприятий и всей отечественной экономики в целом. Новые условия требуют выработки новых стратегий повышения эффективности и обеспечения стабильности функционирования компаний промышленного комплекса страны. Одним из вариантов решения возникших проблем является внедрение принципов умного управления, основанных на использовании современных цифровых технологий и поиске эффективных логистических цепочек поставок сырья, оборудования, комплектующих.

Ключевые слова: экономические санкции, риски, задержки поставок, оптимизация, стратегическое управление, планирование.

Введение

В 2020 году российская экономика столкнулась с пандемией коронавируса, которая повлекла необходимость введения карантинных мер, приостановки работы ряда предприятий, что вызвало структурный экономический кризис в глобальном масштабе. Позднее в 2022 г. к последствиям Covid-19 добавились введенные против РФ рядом зарубежных стран экономические санкции. В результате новых

вызовов перед российскими предприятиями встал вопрос повышения эффективности их работы в сложных условиях. В современных условиях промышленные предприятия должны уделять повышенное внимание оптимизации издержек, это позволит им остаться на рынке и удержать имеющиеся конкурентные преимущества. Одно из возможных направлений повышения эффективности деятельности компаний промышленного сектора связано с внедрением передовых автоматизированных систем.

Анализ существующих проблем развития промышленности России

Банк России в ноябре-декабре 2022 г. провел мониторинг предприятий обрабатывающей промышленности. Всего было опрошено порядка 2 тыс. организаций. Обследование носило разовый характер и касалось адаптации российских предприятий к условиям внешнеэкономического санкционного давления. Организаторы опроса поставили цель выяснить, как санкции со стороны многих зарубежных стран повлияли на производственную деятельность российских компаний. Результаты мониторинга были получены следующие:

- 79% промышленных предприятий РФ столкнулись с поставками импортного сырья и комплектующих;
- 70% респондентов отметили дефицит запчастей;
- 48% компаний испытали недостаток в связи с недопоставкой оборудования и машин [1].

Благодаря довольно быстрой адаптации российских предприятий к новым экономическим условиям речь шла не о полном прекращении импортных поставок, а всего лишь о задержках и перебоях в закупках. Подстройка под режим ограничений происходила с параллельным восстановлением импорта, который происходил сразу в нескольких направлениях:

1. инерция поставок по контрактам, срок действия которых еще не закончился;
2. восстановление деловых партнерских связей с прежними поставщиками;
3. легализация параллельного импорта из других стран;
4. переключение на сырье и продукцию иных иностранных производителей [2].

Сформированный до введения санкций товарный запас по комплектующим, материалам и сырью позволил минимизировать производственные простои в связи с задержкой импортных поставок. Определенный объем возникшего дефицита был ликвидирован благодаря импортозамещению на позиции, которые производятся отечественными предприятиями.

Если говорить в целом, то на конец 2022 г. по результатам мониторинга Банка России адаптация российских промышленных предприятий к санкциям еще не была завершена. Примерно 50% опрошенных Банком России компаний находилась на этапе оптимизации производственных цепочек. Можно констатировать появление проинфляционных рисков на фоне производственных ограничений. По мере адаптации к новым экономическим условиям данные риски будут минимизироваться и их влияние на экономический рост и развитие предприятий ослабнет.

Перестройка экспортных и импортных потоков вызвала у подавляющего числа промышленных предприятий РФ рост производственных затрат и транзакционных издержек, связанных с логистикой. В свою очередь, эти факторы негативно повлияли на цены. Второй немаловажный аспект — простои некоторых производственных компаний из-за дефицита или задержек поставок сырья негативно влияет на объемы продаж и насыщение рынка

готовой продукцией. Перечисленные факторы выступают источниками инфляционного экономического давления [3]. Одновременный рост спроса на товары, ушедшие с рынка, и падение потребительского интереса к отечественным аналогам неблагоприятно сказывается на внутренней рыночной конъюнктуре.

На фоне перечисленных событий происходит снижение инвестиционной активности, вызванное общей неопределенностью. Предприятия, работающие на импортном оборудовании, вынуждены сегодня переходить на альтернативные варианты, подчас с худшими производственно-техническими параметрами. В отдаленной перспективе с высокой долей вероятности это создаст риск накопления технологического отставания.

Автоматизация произведенных процессов как инструмент оптимизации издержек

Когда повышение эффективности производственных процессов через экстенсивный рост становится невозможным, пристальное внимание стоит уделять совокупности внутренних факторов:

- оптимизация ресурсного потребления;
- качественное улучшение бизнес-процессов;
- рост производительности труда;
- сокращение производственных простоев;
- уменьшение времени на настройку оборудования и т. д.

Одним из вариантов разрешения указанных «узких мест» может быть внедрение автоматизированных передовых систем [4]. Если рассматривать современные условия, то в России цифровизация распространяется чаще всего на две производственные области:

1. управление материальным и техническим обеспечением (распоряжение ресурсами);
2. планирование производственных бизнес-процессов.

Остановимся подробнее на планировании производства. Оптимизация данного участка связана с рядом задач, касающихся поиска внутренних резервов роста эффективности. И для решения данных задач промышленное

предприятие может использовать систему класса APS (Advanced Planning and Scheduling). Системный подход обеспечивает автоматизацию процессов операционного, календарного и объемного планирования с учетом имеющихся ограничений. Удобство пользования системами класса APS уже доказано практическим опытом. После введения экономических санкций многие IT-технологии зарубежного производства стали недоступны российским компаниям, но полноценное импортозамещение в данной сфере обусловлено созданием отечественных автоматизированных систем, решающих аналогичные задачи.

Существуют разные примеры российских платформ, которые решают задачи математической оптимизации производственных процессов. Апробация внедрения таких платформ промышленными компаниями по мнению экспертов может дать следующие результаты:

- срок окупаемости до 1 года;
- повышение производительности труда от 5 до 10%;
- сокращение периода оборота запасов и полуфабрикатов от 20 до 30%;
- увеличение времени полезной работы станков и оборудования до 3,5–5% [4].

При внедрении платформы APS-класса внутри промышленного предприятия должен быть усилен контроль исполнения плановых показателей, сформированных системой. Это означает, что специалисты четко должны выполнять установленные для них задачи, не отвлекаясь и не переключаясь на другие проекты. В противном случае логика системного планирования будет нарушена. Поддержка организационных изменений в сфере мотивации сотрудников возможна посредством введения дополнительного KPI с критерием эффективности в виде успешно завершенных производственных поручений, которые получил работник при планировании сменной нагрузки.

Второй областью производственной деятельности, которая хорошо отзывается на цифровизацию, является управление ресурсным обеспечением. Многие крупные промышленные предприятия РФ уделяют недостаточно

внимания производственному снабжению и автоматизации складского учета, что в корне неправильно и является фактором риска, т. к. может нарушить ритмичность поставок сырья или материалов. На сегодняшний день лишь 20% складов материально-технического обеспечения в РФ автоматизировано [4].

Но именно в области производственной логистики часто скрыты резервы оптимизации, которые при оценке экономической выгоды исчисляются миллионами рублей. Для повышения эффективности управления складами специалисты рекомендуют внедрять классические системы класса WMS (Warehouse Management System) с параллельной организацией регламентирования всех операций. Это обеспечит прозрачную оцифровку всех производственных процессов. Можно отметить, что производственная оптимизация в условиях введения экономических санкций позволит промышленным предприятиям РФ не утратить рыночных позиций и выйти на качественно новый уровень развития.

Умное управление для минимизации рисков

До недавнего времени общемировая практика умного управления территориями или отдельными предприятиями базировалась на двух принципах: создание и развитие данных объектов управления. Санкции в отношении большого количества российских промышленных предприятий внесли существенные коррективы в отработанные управленческие схемы и алгоритмы. Поэтому к управлению промышленными предприятиями в измененных политических и экономических условиях стоит подходить комплексно.

От эффективности функционирования отдельных субъектов хозяйствования зависит экономическая стабильность региона и всего государства. Промышленные предприятия в России формируют более 30% валового национального продукта.

Целью оперативного и стратегического управления промышленными предприятиями в России является максимизация рентабель-

ности. Промышленность имеет специфические черты, когда перспективное развитие невозможно без вложений в материально-техническую базу, грамотного управления и планирования. Когда руководство промышленного предприятия не беспокоится о долгосрочном развитии, постепенно происходит утрата конкурентных преимуществ и продукция становится неликвидной. Поэтому ни одно предприятие сегодня не может функционировать автономно, в отрыве от происходящих в общенациональном масштабе событий [5].

Обострение глобальной экономической нестабильности и антироссийские санкции стали тем катализатором, который показывает уровень адаптации предприятий. Перспективные направления повышения эффективности деятельности промышленных российских компаний — использование новейших цифровых технологий и методов цифровизации бизнес-процессов. При разработке нового подхода к управлению любым хозяйствующим субъектом важно сочетать оперативные и стратегические методы, учитывать изменение ключевых факторов. Современная экономическая ситуация характеризуется высокой степенью нестабильности, поэтому факторы прямого воздействия на компании отходят на второй план. Более информативными сегодня считаются факторы, воздействующие на бизнес косвенно.

Разработанная для конкретного предприятия стратегия управления должна быть целостным и хорошо продуманным алгоритмом, сочетающим в себе традиционный подход и возможности цифровых технологий. При ограничении импорта российским промышленным компаниям важно выработать новые эффективные схемы поставок сырья и оборудования.

Выводы

Для повышения эффективности российских промышленных предприятий к решению возникших после введения антироссийских санкций проблем стоит подходить комплексно.

В первую очередь речь идет об оценке возможного импортозамещения из-за ограничения или прекращения импортных поставок. Также при стратегическом и оперативном управлении руководство компаний должно учитывать все риски и действовать с учетом новых условий обдуманно, четко и с ориентацией на перспективу. Использование цифровых технологий для автоматизации различных бизнес-процессов позволяет выявить скрытые резервы роста производительности труда и оптимизировать издержки. Важную роль в преодолении экономической ситуации с санкциями играет государственная поддержка предприятий промышленности как флага на национальной экономики РФ.

Литература

1. Карлова Н., Пузанова Е. Российская обрабатывающая промышленность в условиях санкций. Результаты опроса предприятий. Официальный сайт Банка России: www.cbr.ru. 2023. // Электронный ресурс // URL: https://cbr.ru/Content/Document/File/154320/analytic_note_20230926_dip.pdf.
2. Логиновский О. В., Голлай А. В., Дранко О. И. и др. Эффективное управление организационными и производственными структурами: моногр. / О. В. Логиновский, А. В. Голлай, О. И. Дранко и др. М.: ИНФРА-М, 2020. — С. 125.
3. Коренная К. А., Голлай А. В., Логиновский О. В. Модели управления промышленными предприятиями в условиях нестабильности внешней среды и необходимости технологического перевооружения // Проблемы управления. 2021. № 4. С. 40–49. DOI: 10.25728/pu.2021.4.4. // Электронный ресурс // URL: <https://cyberleninka.ru/article/n/modeli-upravleniya-promyshlennymi-predpriyatiyami-v-usloviyah-nestabilnosti-vneshney-sredy-i-neobhodimosti-tehnologicheskogo>.
4. Воронин С., Сидоров А. Оптимизация производства в условиях санкций: как повысить эффективность в работе // Электронный

ресурс, 18.11.2022 // URL: <https://finance.rambler.ru/economics/49720565-optimizatsiya-proizvodstva-v-usloviyah-sanktsiy-kak-povysit-effektivnost-v-rabote/>.

5. Логиновский О. В., Голлай А. В., Шестаков А. Л., Коренная К. А. Развитие умных технологий в управлении регионами России в современных условиях / О. В. Логиновский, А. В. Голлай, А. Л. Шестаков, К. А. Коренная // Вестник ЮУрГУ. Серия «Компьютерные технологии, управление, радиоэлектроника». 2022. Т. 22, № 3. С. 80–92. DOI: 10.14529/ctcr220308. // Электронный ресурс // URL: <https://cyberleninka.ru/article/n/razvitie-umnyh-tehnologiy-v-upravlenii-regionami-rossii-v-sovremennyh-usloviyah/viewer>.

Improving the efficiency of industrial enterprises under sanctions

Dmitrieva L. V., Aleshina A. V.

Ivanovo State University, Lomonosov Moscow State University

The aggravation of the political, military and economic situation in the world after the introduction of sanctions against Russia by the United States and a number of major Western powers has had a negative impact on the development of industrial enterprises and the entire domestic economy as a whole. The new conditions require the development of new strategies to improve efficiency and ensure the stability of the functioning of companies in the country's industrial complex. One of the options for solving the problems that have arisen is the introduction of smart management principles based on the use of modern digital technologies and the search for effective logistics supply chains for raw materials, equipment, and components.

Keywords: economic sanctions, risks, delivery delays, optimization, strategic management, planning.

References

1. Karlova N., Puzyanova E. Russian manufacturing industry under sanctions. Results of a survey of enterprises. Official website of the Bank of Russia: www.cbr.ru. 2023. — //Electronic resource// URL: https://cbr.ru/Content/Document/File/154320/analytic_note_20230926_dip.pdf
2. Loginovsky O. V., Gollai A. V., Dranko O. I. et al. Effective management of organizational and production structures: monograph / O. V. Loginovsky, A. V. Gollai, O. I. Dranko et al. Moscow: INFRA-M, 2020. — P. 125.
3. Korennaya K. A., Gollai A. V., Loginovsky O. V. Models of industrial enterprise management in the context of environmental instability and the need for technological re-equipment // Problems of Management. 2021. No. 4. P. 40–49. DOI: 10.25728/pu.2021.4.4. //Electronic resource// URL: <https://cyberleninka.ru/article/n/modeli-upravleniya-promyshlennymi-predpriyatiyami-v-usloviyah-nestabilnosti-vneshney-sredy-i-neobhodimosti-tehnologicheskogo>
4. Voronin S., Sidorov A. Optimization of production under sanctions: how to increase efficiency at work // Electronic resource, 11/18/2022 // URL: <https://finance.rambler.ru/economics/49720565-optimizatsiya-proizvodstva-v-usloviyah-sanktsiy-kak-povysit-effektivnost-v-rabote/>
5. Loginovsky O. V., Gollai A. V., Shestakov A. L., Korennaya K. A. Development of smart technologies in the management of Russian regions in modern conditions / O. V. Loginovsky, A. V. Gollai, A. L. Shestakov, K. A. Korennaya // Bulletin of SUSU. Series «Computer technologies, management, radio electronics». 2022. Vol. 22, No. 3. Pp. 80–92. DOI: 10.14529/ctcr220308. //Electronic resource// URL: <https://cyberleninka.ru/article/n/razvitie-umnyh-tehnologiy-v-upravlenii-regionami-rossii-v-sovremennyh-usloviyah/viewer>

Финтех и децентрализованные финансы: тенденции, проблемы и роль в финансовом секторе

Воронкова Светлана Николаевна

Аспирант экономического факультета Воронежского государственного лесотехнического университета
E-mail: 9651021720@mail.ru

Алешина Анна Валентиновна

к. э. н., доцент кафедры финансов и кредита
Экономического факультета МГУ
имени М. В. Ломоносова
E-mail: annaaleshina@mail.ru

Финтех и децентрализованные финансы являются современным трендом развития финансовых технологий. Децентрализованные финансы занимали первоначально очень узкую нишу финансовых технологий и представляли собой достаточно рисковый сектор, в котором процветали мошеннические финансовые схемы. В последние годы ситуация кардинально поменялась и децентрализованные финансы стали внедрять даже классические финансовые институты: банки, биржи, страховые компании. Появление новых финансовых технологий влечет за собой как новые возможности, в первую очередь связанные со снижением транзакционных издержек на проведение финансовых операций и увеличение скорости совершения финансовых сделок, так и новые типы рисков, которые ранее не были присущи финансовым институтам. Изменяющаяся среда на финансовых рынках вызывает необходимость новых подходов к регулированию финансовых операций.

Ключевые слова: финтех, децентрализованные финансы (DeFi), кибербезопасность, цифровая трансформация.

Введение

Внедрение цифровых технологий во многие сферы жизни породило возникновение новых цифровых инструментов как в сфере промышленности, в сельском хозяйстве, в сфере услуг. Финансовая сфера не стала исключением. В качестве альтернативы традиционных субъектам финансовой сферы и централизованным финансам набирают популярности финансовые технологии и децентрализованные финансы. Их преимуществом является высокая степень автономности и отсутствие посредников при осуществлении финансовых операций.

Появившиеся сервисы в области финтех набирают большую популярность, однако относительная новизна практики их использования порождает ряд проблем. На сегодняшний день существует потребность в улучшении уровня кибербезопасности при использовании финтеха, в силу наличия геополитических проблем сфера финтеха не является устойчивой и нуждается в управлении рисками. При этом, учитывая большое количество достоинств данных технологий, многими учеными и практиками высказывается точка зрения о перспективе сбалансированного сочетания традиционных централизованных и децентрализованных инструментов, что, по их мнению, приведет новому взрывному росту в финансовом секторе.

Повсеместная цифровая трансформация видоизменила многие традиционные процессы в большинстве сфер и отраслей. Развитие современных технологий стало причиной

появления новых инструментов и явлений и в финансовом секторе — появляются новые сервисы и локальные экосистемы, увеличивается количество участников рынка, возникают новые интеграции.

Как следствие, в финансовом секторе сокращаются издержки, на смену отделениям обслуживания приходят доступные онлайн-сервисы. Финансовые технологии (далее — финтех) становятся новым драйвером развития экономики. Финтех предоставляет множество технических возможностей, повышающих качество предоставляемых услуг, что, повышает конкурентоспособность организаций, которые их используют, и положительно влияет на спрос. По данным EY Global FinTech Adoption Index, средний уровень проникновения финтех-услуг на различные рынки составляет 33% [1].

В качестве основных преимуществ финтеха называют мобильность предоставления информации и ее доступность, скорость предоставления услуг [2]. Данные технологии автоматизируют традиционные операции и процесс коммуникации в системах «клиент — банк» и «клиент — финтех-провайдер — банк» [3].

Современный отечественный финтех можно характеризовать как довольно конкурентную отрасль, где по сравнению с зарубежными странами сохраняется невысокая стоимость интеллектуального труда и разработок. Это повышает привлекательность финтеха для инвестиций, что увеличивает коммерческий потенциал разрабатываемых в данной сфере технологий. На сегодняшний день сфера финтеха привлекательна и в силу снижения уровня доверия к традиционным субъектам финансового рынка, а также в силу наличия возможности сохранить анонимность в операциях.

Появление и стремительное развитие финтеха стало настоящей революцией в финансовом секторе. В первую очередь, финтех компании стали основными конкурентами традиционному банковскому сектору, обменным пунктам и страховым организациям. Это происходит в силу более низких транзакцион-

ных издержек, более слабого регулирования данного сектора со стороны финансовых регуляторов (в России — Центрального банка), в связи с чем финтех-компаниям не требуется соблюдать большое количество нормативных требований, которые вынуждены соблюдать классические коммерческие банки. Это влечет появление новой парадигмы развития экономики и финансовой сферы и дает возможность появления новых финансовых институтов и качественно новой финансовой инфраструктуры.

Государство поддерживает появление и развитие финтеха в России. Согласно Распоряжению Правительства РФ от 29.12.2022 N 4355-р «Об утверждении Стратегии развития финансового рынка РФ до 2030 года» «на российском рынке будут расширяться возможности для участников финансового рынка по совмещению регулируемых и нерегулируемых видов деятельности на финансовом рынке, включая оптимизацию соответствующих лицензионных процедур, в том числе временный ограниченный упрощенный доступ для финтех-компаний» [4]. Благодаря такой позиции государства могут быть снижены издержки, а также устранены барьеры для полноценного развития и внедрения финтеха в современную экономику.

Одной из последних тенденций в сфере финтеха является появление новых платежных систем — Apple Pay, Google Pay; китайских платежных сервисов Alipay и WeChat Pay. Таким образом повысилась конкуренция платежных сервисов — наряду с привычными всем платежными системами Visa и MasterCard стали появляться новые конкурентоспособные системы, которые предлагают услуги с более низкими комиссиями, что влечет заинтересованность бизнеса во внедрении новых платежных систем.

Децентрализованные финансы являются одним из видов финтеха. По мнению Фабиана Шера децентрализованные финансы представляют собой «финансовую инфраструктуру, построенную на общедоступных платформах смарт-контрактов ... основанную на открытых протоколах и децентрализованных приложе-

ниях» [5]. То есть в отличие от классических банковских операций в моделях, предлагаемых финтех компаниями отсутствует финансовый посредник, а действия не имеют централизованного характера. В рамках Всемирного экономического форума Вартонским университетом был подготовлен проект, в котором под децентрализованными финансами понимается «отдельная новая отрасль, сочетающая технологию блокчейна, цифровые активы и финансовые сервисы, работа которых строится в рамках децентрализованных приложений и исключает наличие централизованного механизма и участие финансовых посредников» [5].

Децентрализованные финансы имеют ряд достоинств, в том числе более низкие комиссии за операции (которые на порядок ниже комиссий классических финансовых институтов), более скоростное проведение финансовой операции, прозрачную систему проведения сделок без возможности вмешательства третьих лиц и внесения исправлений постфактум, что делает сделки, совершаемые с их помощью, надежными.

О росте популярности децентрализованных финансов свидетельствуют статистические данные — по данным портала DeFi Pulse [7] в сервисах, использующих децентрализованные финансы, аккумулируются большие суммы, и с каждым годом стоимость цифровых активов растет. Если в 2020 году стоимость достигала 15 миллиардов долларов, то к середине 2021 года показатели возросли до 80 млрд долларов. Далее, согласно множеству прогнозов экспертов, спрос на услуги сервисов с финтехом будет стабильно расти.

Однако наряду с упомянутыми достоинствами, вызвавшими высокую популярность децентрализованных финансов, следует упомянуть и ряд недостатков. При всей своей мобильности и технологичности децентрализованные финансы несут в себе ряд рисков. Более того, еще не сформирован рискоориентированный подход ко всем негативным последствиям использования децентрализованных финансов, поэтому

пользователи слабо защищены от существующих рисков.

Существует риск банкротства криптобирж и регулятивные механизмы предотвращения таких ситуаций еще не разработаны. Но даже при отсутствии инцидентов, связанных с банкротством, риски неустойчивого функционирования все равно сохраняются.

Причиной неустойчивости децентрализованных финансов могут выступать и мировые геополитические события. Например, в конце февраля и в конце сентября 2022 года много перспективных разработчиков финтеха перестали функционировать на территории Российской Федерации. Некоторые из них вообще прекратили работу в области финтеха из-за риска попадания под санкции. Существуют также высокие риски кибератак на сервисы, что может повлечь потерю денежных средств у пользователей. Примеры таких кибератак с неприятными последствиями для пользователей децентрализованных сервисов можно привести множество.

2022 год стал годом с самым крупным ущербом от действий криптохакеров, так как за 2022 год у криптовалютных компаний было украдено 3,8 миллиарда долларов (см. рис. 1) [9].

При этом резко растут хакерские атаки именно на протоколы децентрализованных финансов. В 2022 году над долю децентрализованных протоколов приходится «82,1% всех украденных хакерами криптовалют — в общей сложности 3,1 миллиарда долларов — по сравнению с 73,3% в 2021 году» [9], см. рис. 2.

Многие страны предпринимают меры по увеличению новых финансовых технологий и снижению криминальности данной сферы. В некоторых странах расширилась сфера государственного регулирования финтех-компаний, стали появляться требования о лицензировании криптобирж. Однако и при таком подходе появляется ряд проблем, препятствующих глобализации сферы, так как получение лицензии на операции в одной из стран не предполагает возможность предоставления услуг в других странах.

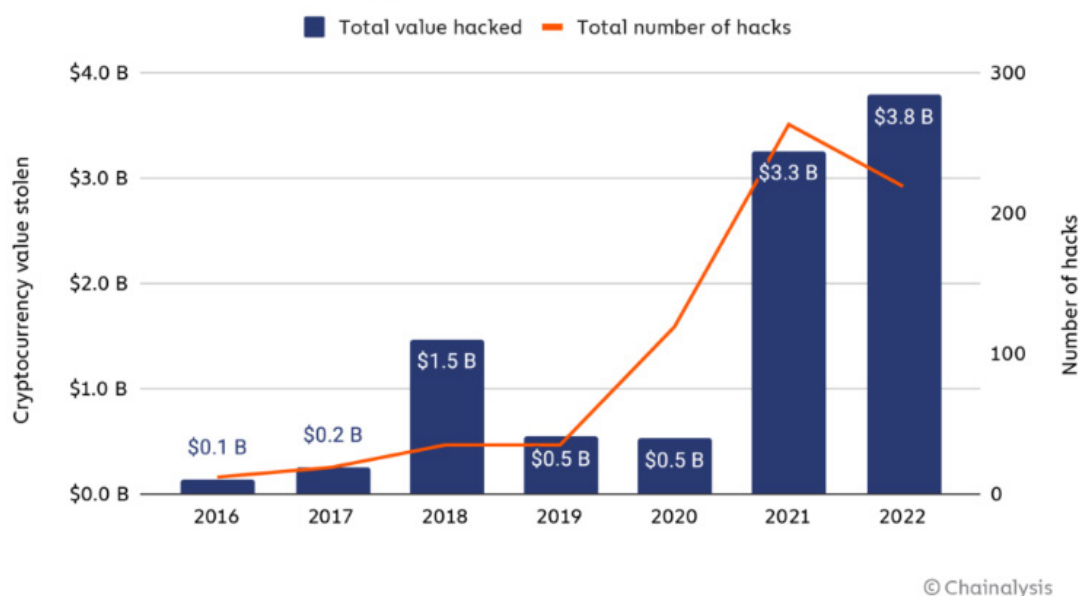


Рис. 1. Количество украденной криптовалюты и количество хакерских атак на криптовалютные кошельки с 2016 по 2022 годы [9]

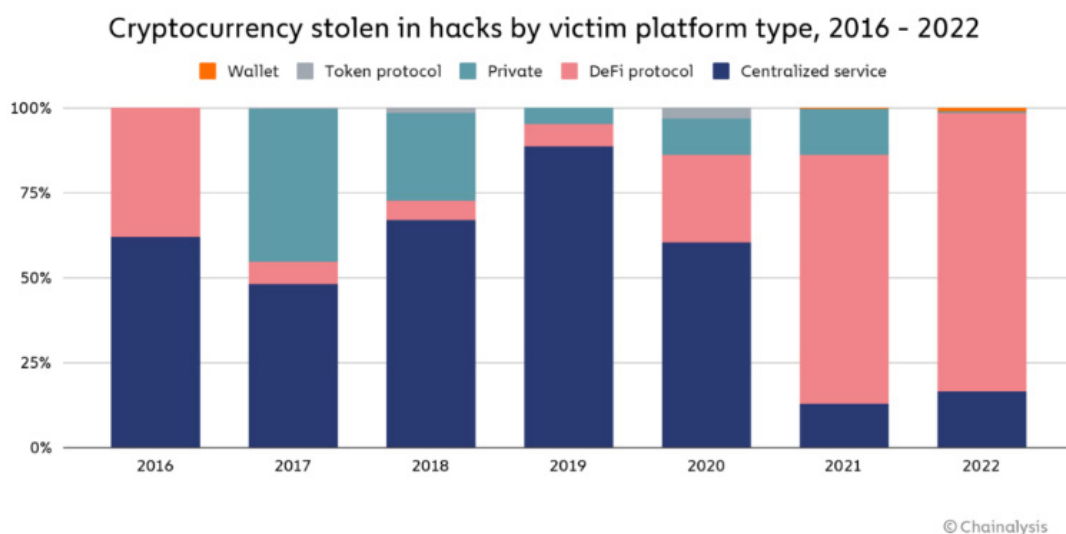


Рис. 2. Распределение похищенных средств с криптокошельков в зависимости от вида платформы за период 2016–2022 годы [9]

Такие меры снижают гибкость финтех-компаний, которые изначально были созданы для увеличения мобильности и упрощения процессов. Введение более жесткого регулирования для криптобирж в ряде стран, в том числе требования об обязательном их лицен-

зировании, увеличивает барьеры для вхождения на рынок разных стран. Криптобиржам придется создавать дочерние компании в других странах в соответствии с регулятивными требованиями этих стран. Это создает излишние препятствия для финтеха, основным

достоинством которого изначально являлось стирание границ между государствами.

Государственное регулирование финтеха даже в развитых странах находится на начальных стадиях развития. В большинстве случаев требования о лицензировании, существующие в ряде стран, носят рекомендательный, а не обязательный характер, что делает даже этот инструмент государственного регулирования довольно точечным.

На фоне названных выше тенденций и проблем большой интерес представляют дискуссии о перспективах финтеха в части полной замены традиционных финансов, имеющих централизованный характер.

Очевидно, на сегодняшний день финтех довольно сильно влияет на работу макрорегуляторов, спрос и предложение в части функционирования децентрализованных финансов стремительно растет, но многие ученые и практики довольно скептически относятся к возможности полной ликвидации традиционных инструментов и централизованных финансов в пользу финтеха.

Даже несмотря на передовой характер децентрализованных финансов их стоимость до сих пор выражается в традиционных финансовых единицах (фиатных валютах), что усложняет процесс их использования в силу необходимости точного установления их самоценности [8].

Параллельное существование централизованных и децентрализованных финансов может повысить стабильность экономики в эпоху участившихся геополитических конфликтов.

Традиционная сфера финансов характеризуется сложной многоуровневой системой с разнообразными составляющими — в их числе резервные валюты, инструменты регулирования курсов валют, универсальные трансграничные системы платежей. Это придает традиционной системе устойчивость.

Избегая крайностей и монополизации централизованных либо децентрализованных финансов, на наш взгляд, можно стремиться к балансу между ними, сочетая стабильность и современность, традиции и новации. В перспективе возможно осуществить рациональ-

ное совмещение централизованных и децентрализованных решений.

Роль финтеха на внутреннем и внешнем финансовом рынке на сегодняшний день трудно переоценить. Последние годы финтех растет все более быстрыми темпами и становится оптимальным и удобным инструментом. Особенно популярность данного инструмента может вырасти в случае, если в данной сфере будут предприняты попытки управления рисками (противодействие мошенничеству и укрепление уровня кибербезопасности).

Выводы

Финтех и децентрализованные финансы являются последним трендом развития финансовых инструментов и технологий как в России, так и за рубежом. Они являются альтернативой традиционным финансовым инструментам. Однако в настоящее время децентрализованные финансы могут нести финансовые риски и не до конца являются безопасными для пользователей. Полагаем, что сочетание традиционных и новых инструментов в финансовой сфере позволит снизить транзакционные издержки на сделки и увеличить скорость совершения финансовых операций.

Литература

1. Проникновение финансово-технологических услуг в мегаполисах России и в мире // Электронный ресурс // URL: [http://www.ey.com/Publication/vwLUAssets/EY-fintech-index-russia-rus2017/\\$FILE/EY-fintech-index-russia-rus-2017.pdf](http://www.ey.com/Publication/vwLUAssets/EY-fintech-index-russia-rus2017/$FILE/EY-fintech-index-russia-rus-2017.pdf).
2. Поддубная М. Н. Особенности и факторы развития системы финтех / М. Н. Поддубная, Е. Я. Волков // Международный журнал гуманитарных и естественных наук. — 2021. — № . 2–1. — С. 202–205 <https://cyberleninka.ru/article/n/osobennosti-i-factory-razvitiya-sistemy-finteh/viewer>.
3. Алешина А. В. Финтех (Fintech) и новые вопросы регулирования / А. В. Алешина,

К. С. Михайлов, А. П. Падалко // Финансовые рынки и банки. — 2021. — № 4. — С. 112–119 <https://cyberleninka.ru/article/n/finteh-fintech-i-novye-voprosy-regulirovaniya/viewer>.

4. Распоряжение Правительства РФ от 29.12.2022 N 4355-р (ред. от 21.12.2023) «Об утверждении Стратегии развития финансового рынка РФ до 2030 года».
5. Schär, Fabian, Decentralized Finance: On Blockchain- and Smart Contract-based Financial Markets (March 8, 2020). // Электронный ресурс // URL: <https://ssrn.com/abstract=3571335>.
6. DeFi Beyond the Hype. The Emerging World of Decentralized Finance // Research produced by the Wharton Blockchain and Digital Asset Project, in collaboration with the World Economic Forum, May 2021 // Wharton University Portal // Электронный ресурс // URL: <https://wifpr.wharton.upenn.edu/wp-content/uploads/2021/05/DeFi-Beyond-the-Hype.pdf>.
7. Consumer warning on Binance Markets Limited and the Binance Group // Financial Conduct Authority (FCA) portal // Электронный ресурс // URL: <https://www.fca.org.uk/news/news-stories/consumer-warning-binance-markets-limited-and-binancegroup>.
8. Павлов А. А. Современный финтех в российском банковском секторе // Научный вестник: Финансы, банки, инвестиции — 2022 — № 4 — с. 61 <https://cyberleninka.ru/article/n/sovremennyy-finteh-v-rossiyskom-bankovskom-sektore>.
9. 2022 Biggest Year Ever For Crypto Hacking with \$3.8 Billion Stolen, Primarily from DeFi Protocols and by North Korea-linked Attackers // Электронный портал chainalysis.com // Электронный ресурс 1 февраля 2023 года // URL: <https://www.chainalysis.com/blog/2022-biggest-year-ever-for-crypto-hacking/>

Fintech and decentralized finance: trends, problems and role in the financial sector

Voronkova S. N., Aleshina A. V.

Voronezh State Forestry University, Lomonosov Moscow State University

Fintech and decentralized finance are a modern trend in the development of financial technologies. Decentralized finance initially occupied a very narrow niche of financial technologies and represented a rather risky sector in which fraudulent financial schemes flourished. In recent years, the situation has changed dramatically and even classical financial institutions have begun to implement decentralized finance: banks, exchanges, insurance companies. The emergence of new financial technologies entails both new opportunities, primarily related to the reduction of transaction costs for financial transactions and the increase in the speed of financial transactions, and new types of risks that were not previously inherent in financial institutions. The changing environment in financial markets necessitates new approaches to regulating financial transactions.

Keywords: fintech, decentralized finance (DeFi), cybersecurity, digital transformation.

References

1. Penetration of financial and technological services in Russian megacities and in the world // Electronic resource // URL: [http://www.ey.com/Publication/vwLUAssets/EY-fintech-index-russia-rus2017/\\$FILE/EY-fintech-index-russia-rus-2017.pdf](http://www.ey.com/Publication/vwLUAssets/EY-fintech-index-russia-rus2017/$FILE/EY-fintech-index-russia-rus-2017.pdf).
2. Poddubnaya M. N. Features and factors of development of the fintech system / M. N. Poddubnaya, E. Ya. Volkov // International journal of humanitarian and natural sciences. — 2021. — No. 2–1. — P. 202–205 <https://cyberleninka.ru/article/n/osobennosti-i-factory-razvitiya-sistemy-finteh/viewer>.
3. Aleshina A. V. Fintech and new regulatory issues / A. V. Aleshina, K. S. Mikhailov, A. P. Padalko // Financial markets and banks. — 2021. — No. 4. — P. 112–119 <https://cyberleninka.ru/article/n/finteh-fintech-i-novye-voprosy-regulirovaniya/viewer>.
4. Order of the Government of the Russian Federation of December 29, 2022 N 4355-r (as amended on December 21, 2023) «On approval of the Strategy for the Development of the Financial Market of the Russian Federation until 2030».
5. Schär, Fabian, Decentralized Finance: On Blockchain- and Smart Contract-based Financial Markets (March 8, 2020). // Electronic resource // URL: <https://ssrn.com/abstract=3571335>.
6. DeFi Beyond the Hype. The Emerging World of Decentralized Finance // Research produced by the Wharton Blockchain and Digital Asset Project, in collaboration with the World Economic Forum, May 2021 // Wharton University Portal // Electronic resource // URL: <https://wifpr.wharton.upenn.edu/wp-content/uploads/2021/05/DeFi-Beyond-the-Hype.pdf>.

7. Consumer warning on Binance Markets Limited and the Binance Group // Financial Conduct Authority (FCA) portal // Electronic resource // URL: <https://www.fca.org.uk/news/news-stories/consumer-warning-binance-markets-limited-and-binancegroup>.
8. Pavlov A. A. Modern fintech in the Russian banking sector // Scientific Bulletin: Finance, banks, investments — 2022 — No. 4 — p. 61 <https://cyberleninka.ru/article/n/sovremennyy-finteh-v-rossiyskom-bankovskom-sektore>.
9. 2022 Biggest Year Ever For Crypto Hacking with \$3.8 Billion Stolen, Primarily from DeFi Protocols and by North Korea-linked Attackers // Electronic portal chainalysis.com // Electronic resource February 1, 2023 // URL: <https://www.chainalysis.com/blog/2022-biggest-year-ever-for-crypto-hacking/>