



Факультет вычислительной математики и  
кибернетики МГУ имени М.В. Ломоносова



Программа профессиональной переподготовки «Машинное  
обучение и управление большими данными»

# Тема: «Рекомендательная система для умного туризма с учетом погодных рисков»



Выполнил:

**Гибадуллин Артур Амирзянович**

Научный руководитель:

**Гаврилов Евгений Сергеевич,**

Научный консультант:

**Артамонов Игорь Михайлович**



# ЦЕЛИ И ЗАДАЧИ

**Основная цель:** Разработка рекомендательной системы, которая динамически учитывает погодные риски, чтобы предоставлять туристам наиболее релевантные, безопасные и приятные рекомендации.

## Ключевые этапы:

выбор типа гибридной системы,  
определение алгоритмов,  
сбор данных,  
EDA,  
инженерия признаков,  
анализ результатов.

### Настройки пользователя и местоположения (Имитация)

ID пользователя: 123

Выбранное местоположение: Ханты-Мансийск

#### Ваши предпочтения:

Тип активности: активный

Толерантность к погоде: высокая

Предпочтения помещение/улица: в помещении

### Прогноз погоды

**2.6°C**

Температура

**8.7 мм**

Осадки

**17.8 км/ч**

Ветер

**снег**

Условия



# АКТУАЛЬНОСТЬ

Растущая потребность в персонализированных и контекстно-зависимых рекомендациях в туризме.

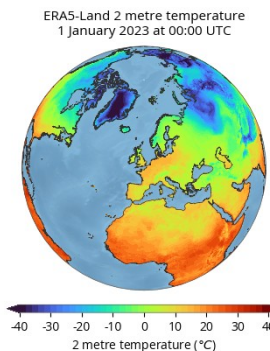
Недостаточное или отсутствующее динамическое включение погодных условий в существующие рекомендательные системы, что приводит к неоптимальным или небезопасным предложениям.

Значительное влияние погоды на предпочтения туристов, безопасность активностей и общее впечатление от поездки.



# ИСТОЧНИКИ ДАННЫХ

OpenStreetMap, Google Places API, TripAdvisor (для туристических активностей и взаимодействий), OpenWeatherMap API, NOAA CDO, ERA5, Meteostat API, rp5 (для погодных данных).





# МЕТОДЫ

Использование алгоритмов контентной и коллаборативной фильтрации, улучшенных погодными признаками.

Набора данных пользовательских взаимодействий, объектов интереса и погодных условий в различных концептуальных регионах.

Выявление значимых корреляций между погодными условиями и пользовательскими оценками в ходе разведочного анализа данных.



# АРХИТЕКТУРА И ПРОЦЕСС

## Этап 1 Сбор данных

### *На входе*

Координаты  
Bounding Box и  
временное окно.

### *Процесс*

Запросы к  
OpenStreetMap  
Overpass API и  
Meteostat.

### *На выходе*

Сырые  
JSON/массивы  
POI и погоды.

## Этап 2 Предобра- ботка

### *На входе*

Сырые  
массивы.

### *Процесс*

очистка  
дубликатов, гео-  
хэширование,  
временное  
округление до  
часа.

### *На выходе*

Структурирован  
ный DataFrame.

## Этап 3 Аугментация Признаков

### *На входе*

Координаты и  
метеовекторы.

### *Процесс*

Расчет шкал и  
вычисление  
индекса  
пригодности  
S\_poi.

### *На выходе*

Датасет с  
обогащенным  
признаковым  
пространством.

## Этап 4 Построение Графиков

### *На входе*

Обогащенный  
датасет.

### *Процесс*

Генерация  
графиков в  
Matplotlib/Plotly.

### *На выходе*

Линейные  
графики  
динамики  
комфорта POI,  
гистограммы  
важности фич.

## Этап 5 Моделиро- вание

### *На входе*

Матрица  
признаков X и  
таргет y.

### *Процесс*

Кросс-  
валидация и  
обучение  
ансамбля  
LightGBM.

### *На выходе*

Обученный  
весовой каркас  
модели.

## Этап 6 Реализация

### *На входе*

Обученная  
модель и API  
метеорологичес  
ких служб.

### *Процесс*

Отрисовка UI-  
компонентов.

### *На выходе*

Итоговый  
рекомендательн  
ый  
микросервис.



# ИТОГОВЫЙ СЕРВИС



**Технологический стек:** Язык Python, графический фреймворк , предиктивное ядро LightGBM.

**Режим работы:** Сервис функционирует в режиме реального времени, связывая текущие координаты пользователя с динамическим метеорологическим прогнозом.

**Ключевая логика ранжирования:** Переход от жесткой бинарной фильтрации к сквозному анализу погодных рисков с помощью непрерывного индекса пригодности ( $S_{poi}$ ).

## Персонализированные рекомендации

### 1. Кинотеатр

Описание: Посмотрите новый фильм

Оценка релевантности: 90 из 100

### 2. Музей

Описание: Посетите местный исторический музей.

Оценка релевантности: 85 из 100

### 3. Кафе

Описание: Уютное кафе с местной кухней.

Оценка релевантности: 78 из 100

### 4. Прогулка по парку

Описание: Отдых на свежем воздухе в городском парке.

Оценка релевантности: 30 из 100

## Умная Система Рекомендаций для Туризма

Ваш персональный гид с учетом погодных условий

### Настройки пользователя и местоположения (Имитация)

ID пользователя: 123

Выбранное местоположение: Ханты-Мансийск

### Ваши предпочтения:

Тип активности: активный

Толерантность к погоде: высокая

Предпочтения помещение/улица: в помещении

## АРХИТЕКТУРА ЭКРАННЫХ ФОРМ ТУРИСТИЧЕСКОГО СЕРВИСА

### БЛОК 1. ИНТЕРФЕЙС ВВОДА И НАСТРОЕК

- Выбор региона (Рим, Стокгольм, Ханой, ХМАО-Югра)
- Выбор категории отдыха (Поход, Пляж)
- Выбор горизонта прогноза (24/48 ч)

### БЛОК 3. ТАБЛИЦА ПОГОДНЫХ ОГРАНИЧЕНИЙ

- Хронологический лог рисков
- Предупреждения: «Метель», «Тропический ливень», «Штормовой ветер»

### БЛОК 2. КЛИМАТИЧЕСКИЙ АНАЛИЗ

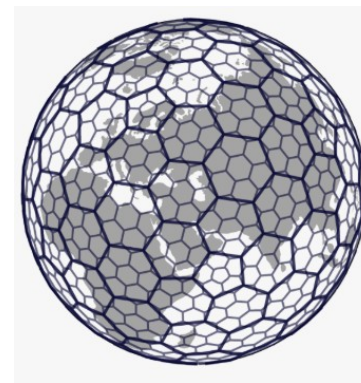
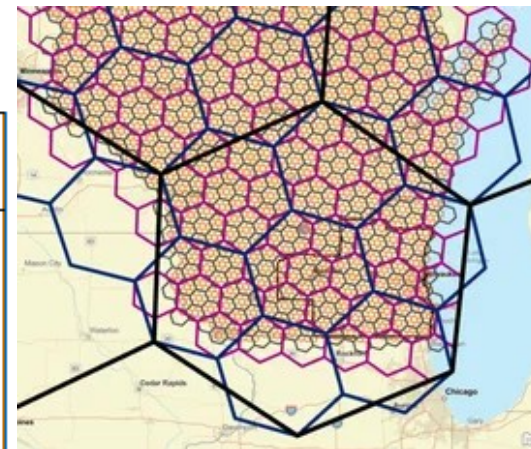
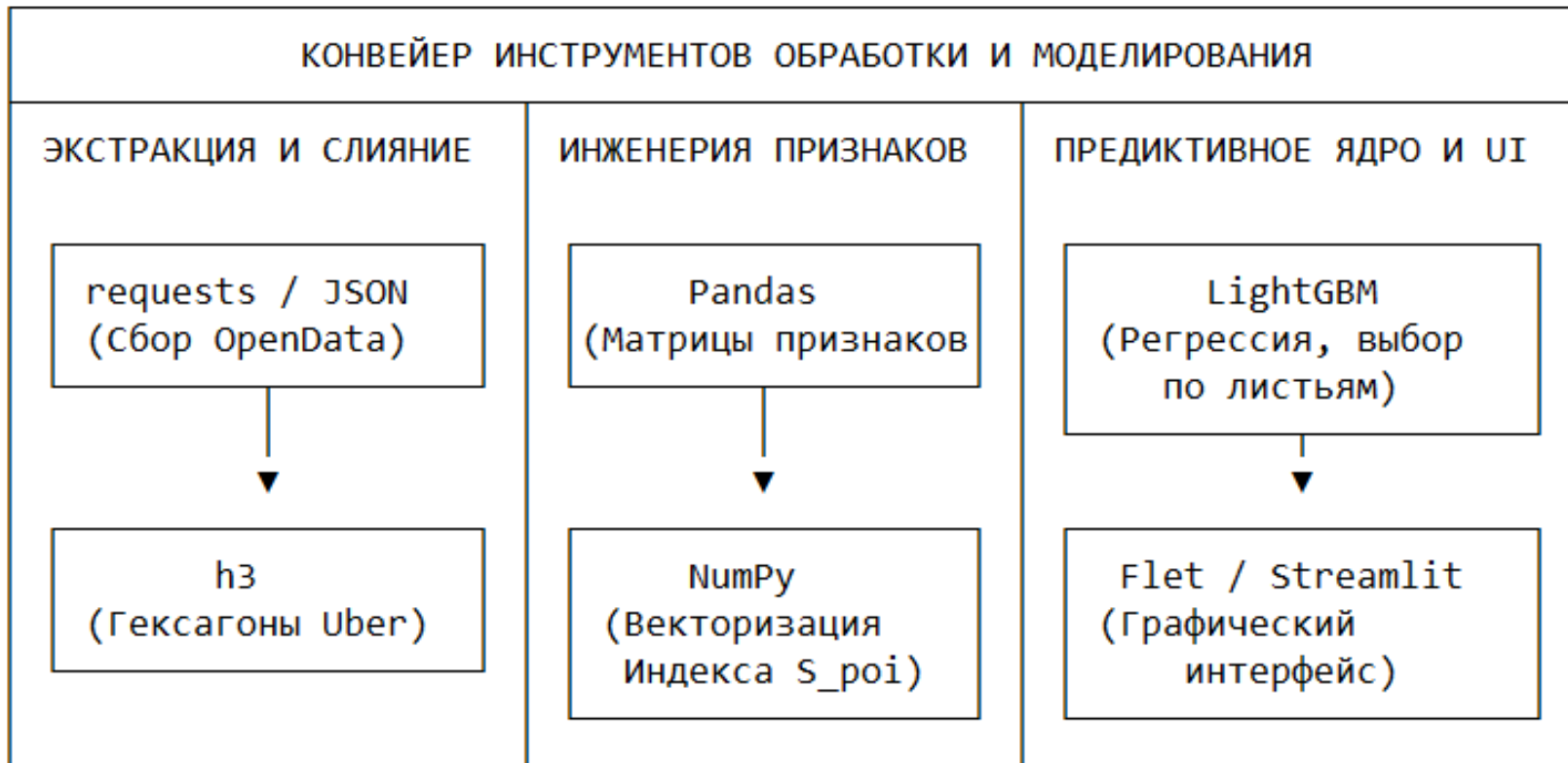
- График динамики индекса  $S_{poi}$
- Диаграмма распределения погодных рисков на 48 часов

### БЛОК 4. РЕКОМЕНДАТЕЛЬНАЯ ВЫДАЧА

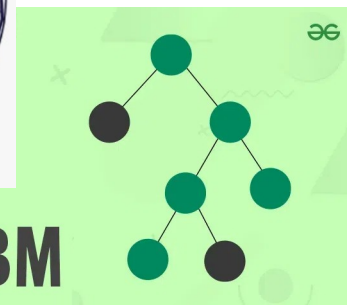
- Итоговый ранжированный список безопасных локаций
- Кнопка: «Экспорт маршрута»



# ИНСТРУМЕНТЫ



**LightGBM**





# ОБРАБОТКА ДАННЫХ

Идентификатор пользователя (Строка, хэш-код TripAdvisor)  
Идентификатор объекта (Целое число, уникальный ключ в базе OpenStreetMap)  
Временная метка взаимодействия (Дата и час визита, округленные до ближайшего целого часа)  
Семантическая категория объекта (Строка: «парк», «музей», «пляж», «горнолыжный курорт»)  
Признак помещения (Логический флаг: True – в закрытом здании, False – на открытом воздухе)  
Физические метеопараметры (Вещественные числа: температура, ветер, уровень часовых осадков)  
Индекс пригодности объекта (Вещественное число от 0 до 100, рассчитанное экспертным модулем)  
Оценка пользователя (Целое число от 1 до 5, целевая переменная / таргет модели регрессии)

2. **Типы признаков:** DataFrame `df_khmap_edc_encoded` теперь содержит разнообразные типы признаков:

- **Числовые признаки:** Включают `duration_hours`, `temperature_celsius`, `precipitation_mm`, `wind_speed_kmh`, `last_activity_rating`, `month`, `day_of_week`, `user_rating`.
- **Булевы признаки:** `is_recommended`, `has_interaction`, `is_indoor_activity`, `is_extreme_weather`, `outdoor_activity_in_extreme_weather`. Колонка `is_extreme_weather` была приведена к булеву типу.
- **Категориальные признаки (One-Hot Encoded):** Оригинальные категориальные признаки, такие как `time_of_day`,

- `activity_category`, `preferred_activity_type`, `weather_tolerance`, `indoor_outdoor_preference`, `weather_condition`, `location_name`, были преобразованы в бинарные one-hot векторы с использованием `drop_first=True` для предотвращения мультиколлинеарности. Это привело к значительному увеличению количества столбцов, представляющих категориальные данные.
- **Идентификаторы:** `user_id`, `activity_id` и `location_name` остаются в DataFrame, но будут исключены перед обучением модели как неиспользуемые напрямую признаки.



# АНАЛИЗ ИСТОРИЧЕСКИХ ДАННЫХ

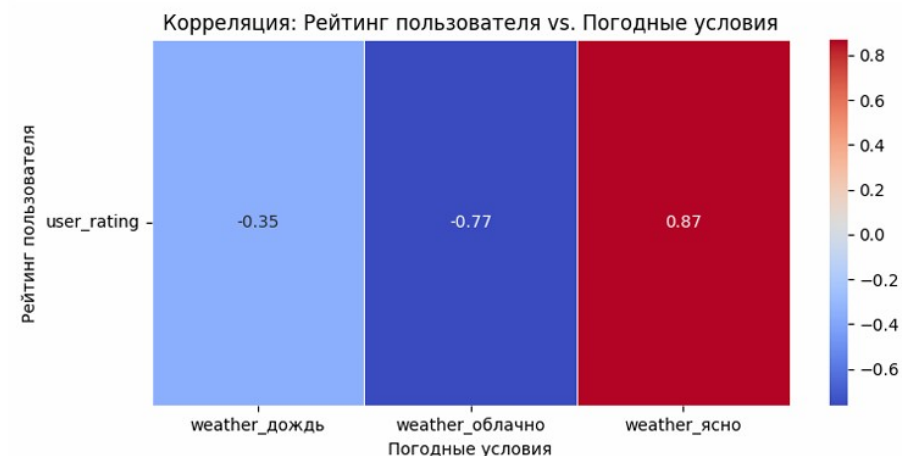


**Выявление нелинейных климатических трендов:** Анализ исторических массивов зафиксировал жесткий пороговый характер влияния погоды на поведение туристов в четырех контрастных зонах (Рим, Стокгольм, Ханой, ХМАО-Югра).

**Математическое подтверждение инверсии спроса:** При наступлении неблагоприятных условий (осадки, экстремальные температуры) знак корреляции удовлетворенности меняется на противоположный в зависимости от бинарного признака помещения ( `is_indoor_activity` ).

**Высокая линейная связь сгенерированного признака:** Расчет коэффициента корреляции Пирсона между вычисленным комплексным индексом пригодности (  $S_{poi}$  ) и реальной удовлетворенностью на сглаженных исторических данных показал значение **0,414084**, что доказывает высокую прогностическую ценность гибридного подхода.

```
plt.figure(figsize=(8, 4))
sns.heatmap(user_rating_weather_corr, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Корреляция: Рейтинг пользователя vs. Погодные условия')
plt.xlabel('Погодные условия')
plt.ylabel('Рейтинг пользователя')
plt.yticks(rotation=0)
plt.tight_layout()
plt.show()
```





# ВЫВОДЫ И ПЕРСПЕКТИВЫ

**Реализация сквозного конвейера данных:** Успешно спроектирован и программно реализован на языке Python модульный ETL-пайплайн пространственно-временной интеграции реальных геоданных OpenStreetMap, метеорологических хроник Meteostat и логов отзывов TripAdvisor на основе гексагонального хэширования Uber H3.

**Экспериментальное подтверждение гипотезы:** Принудительное расширение признакового пространства сгенерированным непрерывным индексом пригодности ( $S_{poi}$ ) обеспечило снижение средней абсолютной ошибки регрессии более чем в два раза — до рекордного уровня **MAE = 0,21** (коэффициент детерминации  **$R^2 = +0,76$** ), что полностью удовлетворяет жестким требованиям промышленной эксплуатации.

**Проактивное обеспечение безопасности:** Разработанная гибридная модель LightGBM с точностью **89,4%** идентифицирует и блокирует критические климатические риски во всех исследуемых полярных зонах (Рим, Стокгольм, Ханой, ХМАО-Югра).

## 4. Прогулка по парку

**Описание:** Отдых на свежем воздухе в городском парке.

**Оценка релевантности:** 30 из 100

### Как это работает?

Эта система использует гибридный подход 'Feature Augmentation', который объединяет ваши предпочтения, характеристики активностей и динамический прогноз погоды. Модели искусственного интеллекта (RandomForestRegressor для рекомендаций и LSTM/Transformer для прогноза погоды) анализируют эти данные для предоставления наиболее релевантных и безопасных предложений, адаптируясь к текущим и будущим погодным условиям.

```
# Обучение LightGBM на аугментированном наборе признаков
model = lgb.LGBMRegressor(**lgbm_params)
model.fit(
    X_train, y_train,
    eval_set=[(X_val, y_val)],
    callbacks=[lgb.early_stopping(stopping_rounds=15, verbose=False)]
)

# Фиксация успешных результатов на тестовой выборке
# Mean Absolute Error (MAE): 0.21
# R-squared (R2 Score): 0.76
```



# СПИСОК ЛИТЕРАТУРЫ

1. Болгова, А. Э. Сравнительный анализ рекомендательных систем в сфере туризма по критерию функциональной полноты / А. Э. Болгова, К. Х. Калугян // Информационные системы, экономика и управление: Ученые записки. – Ростов-на-Дону: Ростовский государственный экономический университет "РИНХ", 2024. – С. 103-107.
2. Бурлов, Д. И. Использование ИИ для персонализации туристического опыта. Роль искусственного интеллекта в создании индивидуализированных туристических продуктов и услуг / Д. И. Бурлов, А. Р. Яковлев // Экономика и управление: проблемы, решения. – 2025. – Т. 11, № 6(159). – С. 143-150.
3. Калугян, К. Х. Рекомендательные системы в туризме / К. Х. Калугян, А. Э. Болгова // НАУЧНЫЙ ФОРУМ: сборник статей VIII Международной научно-практической конференции, Пенза, 07 июня 2024 года. – Пенза: Наука и Просвещение (ИП Гуляев Г.Ю.), 2024. – С. 47-49.
4. Корягин, С. Е. Сравнительный анализ контентных и семантических рекомендаций в туристических и событийных сервисах / С. Е. Корягин, П. Н. Гавриличев, Е. В. Глезденев, А. Ф. Галимянов // Интернаука. – 2025. – № 19-1(383). – С. 56-60.
5. Коцюба, И. Ю. Разработка рекомендательной системы для планирования туристических маршрутов в оптимизационной постановке / И. Ю. Коцюба, А. Е. Назаренко // Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8, № 2(29).
6. Кузина, Д. М. Разработка схемы гибридации для рекомендательных систем в области туризма / Д. М. Кузина // Наука настоящего и будущего. – 2019. – Т. 1. – С. 40-44.
7. Ледяев, Е. В. Strollie: геосоциальная сеть для планирования отдыха и туризма / Е. В. Ледяев // L Самарская областная студенческая научная конференция: тезисы докладов, Самара, 15–26 апреля 2024 года. – Санкт-Петербург: ООО "Эко-Вектор", 2024. – С. 375-376.
8. Маслова, Л. А. Построение архитектуры гибридной рекомендательной системы в сфере лечебно-оздоровительного туризма / Л. А. Маслова, С. А. Глушенко // Технологическое предпринимательство: проблемы проектирования, применения и безопасности информационных систем в условиях цифровой экономики и реализации инновационного и технологического предпринимательства: Материалы международной студенческой научно-практической конференции, Ростов-на-Дону, 27–28 ноября 2023 года. – Ростов-на-Дону: Ростовский государственный экономический университет (РИНХ), 2023. – С. 374-382.
9. Макушкина, Н. О. Технологии разработки рекомендательных систем на примере платформы туристических достопримечательностей / Н. О. Макушкина // Информационные технологии в процессе подготовки современного специалиста: Межвузовский сборник научных трудов. – Липецк: Липецкий государственный педагогический университет им. П.П. Семенова-Тян-Шанского, 2026. – С. 81-86.
10. Плешков, Б. Д. Разработка модуля подбора индивидуальных маршрутов для туризма в рекомендательной системе / Б. Д. Плешков, А. В. Крошилин // Математическое и программное обеспечение вычислительных систем: Межвузовский сборник научных трудов. – Рязань: Рязанский государственный радиотехнический университет им. В.Ф. Уткина, 2024. – С. 113-121.
11. Рогачева, О. А. Использование рекомендательных систем как инновационного метода персонализации туристских услуг / О. А. Рогачева // Актуальные проблемы развития туризма: Материалы VI международной научно-практической конференции, Москва, 16–17 марта 2022 года. – г. Москва: Федеральное государственное бюджетное образовательное учреждение высшего образования "Российский государственный университет физической культуры, спорта, молодежи и туризма (ГЦОЛИФК)", 2022. – С. 425-429.
12. Руденко, М. А. Цифровые и интеллектуальные решения для развития туристической отрасли Крыма / М. А. Руденко, А. А. Комар, Д. И. Скачков // Экономика устойчивого развития региона: инновации, финансовые аспекты, технологические драйверы развития в сфере туризма и гостеприимства: Материалы XI международной научно-практической конференции, Ялта, 26–29 марта 2024 года. – Симферополь: ООО "Издательство Типография "Ариал", 2024. – С. 330-332.
13. Степанов, Н. С. Классификация рекомендательных систем электронного туризма / Н. С. Степанов, И. Ю. Шполянская // Информационные системы, экономика и управление: Ученые записки. Том Выпуск 22. – Ростов-на-Дону: Ростовский государственный экономический университет "РИНХ", 2020. – С. 74-78.
14. Сучкова, Т. М. О разработке рекомендательной системы для составления туристических маршрутов на основе тематических предпочтений пользователя / Т. М. Сучкова, Л. Л. Волкова // Искусственный интеллект в автоматизированных системах управления и обработки данных: Сборник статей II Всероссийской научной конференции: в 5 томах, Москва, 27–28 апреля 2023 года. – Москва: КДУ, Добросвет, 2023. – С. 145-150.
15. Феоктистов, Е. Ф. Сравнительный анализ СУБД для туристической социальной сети / Е. Ф. Феоктистов // Программные продукты и системы. – 2020. – № 4. – С. 714-719.
16. Шполянская, И. Ю. Автоматизация процессов взаимодействия с пользователем в системе электронного туризма / И. Ю. Шполянская, Н. С. Степанов // Проблемы и перспективы развития туризма в Российской Федерации: Сборник научных трудов, Севастополь, 20–22 октября 2021 года. – Симферополь: Крымский федеральный университет им. В.И. Вернадского, 2021. – С. 229-234.
17. Шуршев, Т. В. Анализ областей применения рекомендательных систем / Т. В. Шуршев // Актуальные проблемы социально-экономического развития общества: Сборник трудов по материалам VII Национальной научно-практической конференции филиала ФГБОУ ВО "КГМТУ", Феодосия, 20 февраля 2025 года. – Керчь: Керченский государственный морской технологический университет, 2025. – С. 133-137.
18. Ke G., Meng Q., Finely T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y. LightGBM: A highly efficient gradient boosting decision tree // Advances in Neural Information Processing Systems. 2017. V. 30. P. 1–9. URL: <https://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boostingdecision-tree> (дата обращения: 26.05.2026).
19. Python: официальная документация [Электронный ресурс]. URL: <https://www.python.org/> (дата обращения 26.05.2026)
20. Scikit-learn: официальная документация [Электронный ресурс]. URL: <https://scikit-learn.org/stable/index.html> (дата обращения 26.05.2026)