

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

по выполнению лабораторных работ
по дисциплине «Технологии анализа больших данных»
для студентов специальности 38.05.01 – Экономическая безопасность
специализация «Экономико-правовое обеспечение экономической
безопасности»

Ставрополь
2026

Методические указания по дисциплине «технологии анализа больших данных» предназначены для выполнения лабораторных работ по изучаемой дисциплине.

Выполнение заданий направлено на формирование у студентов системного представления о технологиях больших данных (Big Data) и выработке практических навыков их применения для решения профессиональных задач в сфере экономической безопасности, включая выявление мошенничества, оценку рисков, анализ финансовых потоков и противодействие легализации преступных доходов.

Содержание

1. Цели и задачи освоения дисциплины	4
2. Компетенции обучающегося, формируемые в результате освоения дисциплины	4
3. Задания к лабораторным работам	5
Лабораторная работа по теме 1. Введение в Big Data. Роль в системе экономической безопасности.	5
Лабораторная работа по теме 2. Технологический стек Big Data. Инструменты и платформы.	6
Лабораторная работа по теме 3. Основы работы с данными на Python	7
Лабораторная работа по теме 4. Предварительная обработка и очистка данных (ETL)	8
Лабораторная работа по теме 5. Визуализация данных	9
Лабораторная работа по теме 6. Основы анализа данных и машинного обучения	11
Лабораторная работа по теме 7. Прикладные аспекты Big Data в экономической безопасности	12
Лабораторная работа по теме 8. Правовые и этические основы работы с Big Data	13
Список рекомендуемой литературы	15

1. Цель и задачи освоения дисциплины

Целью освоения дисциплины является формирование части общепрофессиональных (ОПК-6) и профессиональных (ПК-1) компетенций будущего специалиста по направлению подготовки 38.05.01 Экономическая безопасность специализация «Экономико-правовое обеспечение экономической безопасности». Основными задачами изучения дисциплины «Технологии анализа больших данных» являются:

- формирование навыков системного представления о технологиях больших данных (Big Data);
- приобретение умений решения профессиональных задач в сфере экономической безопасности, включая выявление мошенничества, оценку рисков, анализ финансовых потоков и противодействие легализации преступных доходов;
- изучение основных методов и алгоритмов анализа данных, применимых в задачах экономической безопасности;
- формирование понимания процессов ETL (Extract, Transform, Load) и работы с различными типами данных (структурированные, неструктурированные).

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Код, формулировка компетенции	Код, формулировка индикатора	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
ОПК-6. Способен использовать современные информационные технологии и программные средства при решении профессиональных задач	ИД-1 ОПК-6. Использует современные информационные технологии и программные средства при решении профессиональных задач	Применяя современные информационные технологии, программные средства, автоматизированные способы обработки данных осуществляет загрузку, очистку, трансформацию и анализ структурированных и неструктурированных данных; использует библиотеки визуализации для создания графиков и диаграмм, выявляющих закономерности и аномалии в данных; разрабатывает интерактивные дашборды в BI-системах для мониторинга ключевых показателей деятельности организации.
ПК-1 способность подготавливать исходные данные и выполнять расчеты экономических показателей для осуществления планово-отчетной работы организации, разработки	ИД-1 ПК-1. Подготавливает исходные данные для модельного описания конкретных количественных взаимосвязей, обусловленных общими экономическими	Применяя методы идентификации, сбора и верификации исходных данных из различных внутренних и внешних источников, процедуры предварительной обработки данных (ETL). методики расчета ключевых экономических показателей на основе агрегированных данных рассчитывает экономические показатели деятельности организации в разрезе различных

проектных решений, разделов текущих и перспективных планов экономического развития организации, бизнес-планов, смет, учетно-отчетной документации, нормативов затрат и соответствующих предложений по реализации разработанных проектов, планов, программ	закономерностями, для разработки проектных решений, разделов текущих и перспективных планов экономического развития организации, бизнес-планов и выполняет расчеты экономических показателей для осуществления планово-отчетной работы организации, подготовки учетно-отчетной документации	сегментов с использованием группировок и агрегаций, обоснует репрезентативность и достаточность исходных данных для модельного описания.
---	---	--

3. Задания к лабораторным работам

Лабораторная работа по теме

1. Введение в Big Data. Роль в системе экономической безопасности.

Цель: освоить классификацию источников Big Data и научиться связывать их с задачами экономической безопасности.

Лабораторная работа 1. Классификация источников и угроз. Работа в малых группах.

Задание:

1. Получите список из 5 отраслей (банкинг, ритейл, телеком, госсектор, логистика).

2. Для каждой отрасли определите не менее 5 источников Big Data (транзакции, логи, соцсети, сенсоры, открытые реестры, геоданные).

3. По каждому источнику опишите:

- тип данных (структурированные/неструктурированные);
- какие характеристики V (Volume, Velocity, Variety) преобладают;
- какие угрозы экономической безопасности можно выявить (мошенничество, утечки, недобросовестные контрагенты, репутационные риски).

Оформите в виде таблицы и представьте к публичной защите.

Лабораторная работа 2. Сбор данных из открытых источников (API). Работа в паре. Используется Python (библиотеки requests, pandas).

Задание:

1. Зарегистрируйтесь на сервисе, предоставляющем открытые данные

(например, API Центробанка РФ, API newsapi.org, публичные данные ЕГРЮЛ через любой открытый портал или датасеты с Kaggle).

2. Напишите скрипт, который получает:

- курсы валют за последний месяц.
- заголовки новостей по ключевому слову «банкротство» за неделю;
- список компаний с определенным ОКВЭД.

3. Сохраните полученные данные в формате CSV/JSON.

4. Проведите первичный анализ: сколько записей, какие поля, есть ли пропуски.

Лабораторная работа 3. Построение карты рисков по новостному фону. Работа с данными из лабораторной работы 2.

Задание:

1. Используя полученные новостные заголовки, проведите контент-анализ на предмет упоминаний конкретной компании (выберите публичную компанию).

2. Классифицируйте упоминания на позитивные, негативные, нейтральные.

3. Постройте временной ряд тональности новостей.

4. Сделайте вывод: как изменение тональности может коррелировать с изменением стоимости акций или кредитным рейтингом (гипотетически).

Лабораторная работа по теме

2. Технологический стек Big Data. Инструменты и платформы.

Цель: изучить архитектуры хранения и обработки данных, научиться выбирать инструменты под задачу.

Лабораторная работа 1. Развертывание тестового Hadoop-кластера в Docker. Работа на локальных машинах с Docker.

Задание:

1. Установите Docker и docker-compose.

2. Используя готовый образ (например, sequenceiq/hadoop-docker или big-data-europe/docker-hadoop), поднимите кластер из одной NameNode и двух DataNode.

3. Зайдите в контейнер NameNode, выполните команды HDFS:

- `hdfs dfs -mkdir /user/data;`
- загрузите любой текстовый файл в HDFS;
- выведите список файлов и содержимое.

4. Напишите простейший MapReduce на Python (WordCount) и запустите его на загруженном файле через `hadoop jar`.

5. Соберите скриншоты успешного выполнения.

Лабораторная работа 2. Работа с облачным объектным хранилищем. Использование Yandex Object Storage / AWS S3.

Задание:

1. Создайте бакет (bucket) в облаке.
2. Настройте политики доступа: сделайте один объект публичным, другой - приватным.
3. Загрузите набор файлов (несколько CSV) через веб-интерфейс и через Python (boto3 для AWS / boto3-совместимый для Yandex).
4. Напишите скрипт, который вычитывает все файлы из бакета и объединяет их в один DataFrame.

Лабораторная работа 3. Проектирование архитектуры для AML-системы.

Проектная работа в группах.

Задание:

1. Представьте, что вы архитектор данных в банке. Вам нужно спроектировать систему для выявления подозрительных транзакций в реальном времени.
2. Требования - обработка 10 000 транзакций/сек, хранение истории 7 лет, возможность графового анализа связей, быстрое реагирование.
3. Выберите компоненты - какой брокер сообщений, какая stream-обработка, какое хранилище (HBase / Cassandra / ClickHouse), какой графовый движок (Neo4j / JanusGraph).
4. Нарисуйте схему взаимодействия компонентов и обоснуйте выбор каждого.

Лабораторная работа по теме

3. Основы работы с данными на Python

Цель: освоить Pandas и NumPy для анализа данных, включая работу с большими объёмами.

Лабораторная работа 1. Исследование кадровых данных. Выдаётся расширенный датасет (сотрудники + их транзакции по корпоративным картам).

Задание:

1. Загрузите данные employees.csv и spendings.csv.
2. Объедините их по идентификатору сотрудника.
3. Рассчитайте для каждого отдела:
 - средний возраст, среднюю зарплату, средний стаж.
 - суммарные траты по категориям (такси, рестораны, авиабилеты).
4. Задача по безопасности: найдите сотрудников, чьи траты в категории «рестораны» превышают 30% от зарплаты (потенциальные долги/коррупция).
5. Постройте сводную таблицу (pivot_table) с суммой трат по отделам и категориям.

Лабораторная работа 2. Работа с временными рядами. Дается датасет с ежедневными объемами продаж и инцидентами безопасности (сбои, атаки).

Задание:

1. Преобразуйте столбец с датой в формат `datetime` и установите его как индекс.
2. Выполните ресемплинг: найдите недельные и месячные суммы продаж.
3. Наложите на график даты инцидентов безопасности. Видна ли корреляция (падение продаж после инцидента)?
4. Рассчитайте скользящее среднее для сглаживания шума.

Лабораторная работа 3. Оптимизация памяти при работе с большими данными. Искусственно созданный датасет большого размера (> 1 млн строк).

Задание:

1. Загрузите датасет обычным способом (`pd.read_csv`). Зафиксируйте время и потребление памяти.
2. Оптимизируйте загрузку:
 - укажите типы столбцов (`dtype`) для снижения памяти.
 - загружайте данные чанками (`chunksize`) и агрегируйте на лету.
3. Сравните время выполнения и использование памяти до и после оптимизации. Напишите вывод.

Лабораторная работа 4. Работа с API и парсинг. Индивидуальное задание.

Задание:

1. Используя API Центробанка, получите курсы валют за последние 30 дней.
2. Постройте график изменения курса USD/RUB.
3. Найдите день, когда курс изменился максимально за сутки. Попробуйте найти новостное событие, которое могло это вызвать.

Лабораторная работа по теме

4 Предварительная обработка и очистка данных (ETL)

Цель: овладеть методами очистки, трансформации и интеграции данных из разных источников.

Лабораторная работа 1. Очистка и стандартизация данных контрагентов. Несколько исходных «грязных» файлов: `customers.csv`, `contracts.xlsx`, `blacklist.json`.

Задание:

1. Загрузите все файлы.
2. В `customers.csv`:
 - приведите телефоны к единому формату (например, +7 (XXX) XXX-

XX-XX).

- исправьте опечатки в названиях городов (используйте словарь или нечёткое сравнение).

- удалите строки, где ИНН меньше 10 или 12 цифр.

3. В contracts.xlsx:

- переименуйте столбцы в единый стиль (snake_case).

- замените все пропуски в сумме контракта на 0.

4. Объедините данные из двух таблиц по идентификатору контрагента, добавив флаг «в чёрном списке» из JSON.

Лабораторная работа 2. Обработка выбросов и нормализация.
Используем очищенный датасет.

Задание:

1. Найдите выбросы в столбце «сумма контракта» методом межквартильного размаха (IQR). Визуализируйте боксплотом.

2. Примените winsorizing (замените выбросы на граничные значения).

3. Выполните нормализацию (MinMaxScaler) и стандартизацию (StandardScaler) для числовых признаков.

4. Сравните распределения до и после. Какой метод лучше подходит для последующей кластеризации?

Лабораторная работа 3. Извлечение данных из PDF-отчётов. Даются несколько PDF-файлов с финансовыми отчётами компаний.

Задание:

1. Используя библиотеки tabula-py или camelot, извлеките табличные данные из PDF.

2. Преобразуйте извлечённые данные в структурированный DataFrame.

3. Объедините данные из нескольких файлов в один.

4. Очистите от лишних символов (пробелы, точки).

Лабораторная работа 4. Обработка пропусков. Работа с пропусками в разных колонках.

Задание:

1. Определите долю пропусков в каждом столбце.

2. Для категориальных признаков заполните пропуски модой.

3. Для числовых — попробуйте разные стратегии: среднее, медиана, заполнение предыдущим значением (ffill).

5. Оцените, как изменились основные статистики (среднее, стандартное отклонение) при разных подходах.

Лабораторная работа по теме

5. Визуализация данных

Цель: научиться создавать наглядные отчёты и дашборды для мониторинга рисков.

Лабораторная работа 1. Разведывательный анализ с Matplotlib/Seaborn. Используем очищенный датасет контрагентов.

Задание:

1. Постройте матрицу корреляций между числовыми признаками. Выделите сильно коррелирующие пары.
2. Создайте pairplot с цветовым кодированием по признаку «наличие в чёрном списке».
3. Постройте распределение суммы контрактов в логарифмической шкале.
4. Создайте столбчатую диаграмму средних сумм контрактов по регионам.
5. Подпишите оси, добавьте заголовки, настройте легенду.

Лабораторная работа 2. Геовизуализация с Folium. Данные о местоположении контрагентов (координаты или адреса).

Задание:

1. Преобразуйте адреса в координаты с помощью геокодера (например, geopy).
2. Создайте карту с маркерами для каждого контрагента.
3. Размер маркера сделайте пропорциональным сумме контракта, цвет — по риску (например, красный — чёрный список).
4. Добавьте слой тепловой карты (HeatMap) для концентрации компаний.

Лабораторная работа 3. Интерактивный дашборд в Power BI / Tableau. Работа в BI-инструменте.

Задание:

1. Импортируйте датасет.
2. Создайте дашборд для службы экономической безопасности с листами:
 - обзор - карточки с KPI (кол-во контрагентов, общая сумма контрактов, доля сомнительных).
 - анализ по отраслям - круговая диаграмма и таблица с фильтром по отрасли.
 - динамика - линейный график заключённых контрактов по месяцам.
 - карта рисков - географическая карта с цветовой индикацией.
3. Добавьте срез (slicer) по дате и по региону.
4. Опубликуйте дашборд в веб-доступ (если возможно) или сохраните как PDF.

Лабораторная работа 4. Визуализация графов. Данные о связях между контрагентами (кто с кем заключал контракты).

Задание:

1. Постройте граф с помощью NetworkX.

2. Выделите узлы с высокой степенью (центральные игроки).
3. Найдите сообщества (кластеры) и раскрасьте их.
4. Сделайте вывод: есть ли подозрительные группы компаний, которые часто взаимодействуют друг с другом.

Лабораторная работа по теме

6. Основы анализа данных и машинного обучения

Цель: применить базовые алгоритмы ML для выявления аномалий, сегментации и прогнозирования.

Лабораторная работа 1. Кластеризация контрагентов. Использование нормализованных признаков.

Задание:

1. Выберите признаки для кластеризации (например, сумма контрактов, частота участия в тендерах, задолженность).
 2. Определите оптимальное число кластеров методом «локтя» (elbow method).
 3. Примените K-Means.
 4. Интерпретируйте кластеры:
 - кластер 0: «Крупные игроки»
 - кластер 1: «Мелкие разовые»
 - кластер 2: «Проблемные» (высокая задолженность)
- Визуализируйте кластеры в двухмерном пространстве (PCA или t-SNE).

Лабораторная работа 2. Поиск аномалий методом Isolation Forest. Датасет из лабораторной работы 1 с добавлением искусственных аномалий.

Задание:

1. Обучите модель Isolation Forest.
2. Получите предсказания (1 - норма, -1 - аномалия).
3. Посмотрите на описание аномальных точек. Какие у них характеристики? (например, очень маленькая выручка, но огромный долг).
4. Оцените качество, если есть размеченные аномалии (матрица ошибок).
5. Сравните с методом One-Class SVM.

Лабораторная работа 3. Регрессия для прогнозирования риска дефолта. Датасет с историческими данными о кредитах (сумма, срок, доход, возраст, просрочка).

Задание:

1. Целевая переменная: был ли дефолт (0/1). Используйте логистическую регрессию.
2. Разделите выборку на обучающую и тестовую.
3. Обучите модель, оцените точность, полноту, F1-меру.
4. Постройте ROC-кривую и посчитайте AUC.

5. Интерпретируйте коэффициенты модели: какие факторы больше всего влияют на риск дефолта.

6. Попробуйте улучшить модель с помощью RandomForestClassifier, сравните метрики.

Лабораторная работа 4. Поиск мошеннических транзакций. Датасет с транзакциями (несбалансированный).

Задание:

1. Обучите модель классификации для выявления фрода.
2. Используйте методы борьбы с дисбалансом классов (взвешивание классов, SMOTE).
3. Оцените качество с упором на полноту (recall) — важно найти как можно больше мошеннических транзакций даже ценой ложных срабатываний.

Лабораторная работа по теме

7. Прикладные аспекты Big Data в экономической безопасности

Цель: Решение реальных кейсов с использованием полученных навыков.

Лабораторная работа 1. Выявление подозрительных транзакций (AML). Дается большой датасет транзакций (отправитель, получатель, сумма, время, страна).

Задание:

1. Реализуйте набор правил (rule-based detection):
 - сумма > 600 000 руб. (порог ЦБ);
 - транзакции в ночное время (0-5 утра);
 - дробление крупной суммы на несколько мелких в течение короткого времени (например, суммарно > 1 млн за час).
2. Для выявления дробления используйте оконные функции (groupby + rolling sum).
3. Пометьте подозрительные транзакции и визуализируйте их количество по дням.

Лабораторная работа 2. Анализ госзакупок на предмет картельного сговора. Датасет госзакупок (lot_id, winner_inn, winner_name, price, date, okpd).

Задание:

1. Рассчитайте долю побед для каждого участника.
2. Найдите пары компаний, которые часто побеждают в одних и тех же лотах по очереди.
3. Постройте граф соучастников (ребро, если компании участвовали в одном лоте). Выделите кластеры, где участников мало, а лотов много.
4. Проверьте гипотезу: есть ли компании, которые выигрывают лоты с ценой, близкой к начальной (возможно, сговор с заказчиком). Рассчитайте отношение цены контракта к начальной цене.

5. Напишите отчёт с подозрительными участниками.

Лабораторная работа 3. Прогнозирование банкротств. Финансовые показатели компаний за несколько лет (активы, обязательства, прибыль, денежный поток).

Задание:

1. Рассчитайте коэффициент Альтмана (Z-score) для каждой компании за каждый год.
2. Определите пороговые значения (красная, серая, зеленая зоны).
3. Постройте временной ряд изменения Z-score для компаний, которые обанкротились, и для успешных. Есть ли паттерн?
4. Обучите модель классификации (например, градиентный бустинг) для прогнозирования банкротства на год вперёд. Оцените важность признаков.

Лабораторная работа 4. Анализ цифрового следа сотрудников. Логи активности сотрудников в корпоративной сети (входы, выходы, объём переданных данных, посещённые сайты).

Задание:

1. Постройте профиль нормального поведения для каждого сотрудника (среднее время входа, объём трафика).
2. Выявите аномалии: кто заходил в нерабочее время, кто скачал аномально много данных.
3. Визуализируйте активность по часам.

Лабораторная работа по теме

8. Правовые и этические основы работы с Big Data

Цель: освоить нормативную базу и научиться проектировать этические и законные решения.

Лабораторная работа 1. Разбор нарушений законодательства. Кейсы из реальной практики (открытые решения Роскомнадзора, судов).

Задание:

1. Прочитайте краткое описание ситуации: компания собирала биометрию без согласия, или передавала данные третьим лицам, или не удалила данные по требованию.
2. Определите, какие статьи законов нарушены (152-ФЗ, 572-ФЗ, УК РФ).
3. Какие штрафы грозят компании?
4. Предложите меры, которые нужно было принять, чтобы избежать нарушения.

Лабораторная работа 2. Разработка политики конфиденциальности. Работа в группах.

Задание:

1. Представьте, что вы запускаете мобильное приложение для анализа финансов клиентов (помощник по бюджету).

2. Напишите пользовательское соглашение и политику обработки персональных данных (краткую версию, основные пункты):

- какие данные собираете;
- с какой целью;
- кто имеет доступ;
- срок хранения;
- как пользователь может отозвать согласие.

3. Оформите в виде документа и представьте другой группе для «юридической проверки».

Лабораторная работа 3. Этическая дилемма. Дебаты.

Задание:

1. Ситуация: Банк хочет использовать систему распознавания лиц в отделениях для выявления клиентов, которые могут быть мошенниками (по базе неблагонадежных). Система также может отслеживать эмоции и «подозрительное» поведение.

2. Разделитесь на две команды: «За» (безопасность важнее приватности) и «Против» (нарушение прав, риск дискриминации).

3. Подготовьте аргументы с опорой на этические принципы (прозрачность, справедливость, минимизация данных) и законодательство.

4. Проведите дебаты. В конце попытайтесь найти компромиссное решение.

Лабораторная работа 4. Разработка контрольного списка соответствия требованиям. Индивидуальная работа.

Задание:

1. Создайте чек-лист для нового проекта по обработке данных, чтобы проверить его соответствие требованиям безопасности и закона.

2. Включите пункты:

- получено ли согласие субъекта?
- уведомлен ли Роскомнадзор?
- обеспечена ли шифровка данных при передаче?
- назначен ли ответственный за обработку?
- есть ли процедура удаления данных по запросу?
- проведена ли оценка вреда?

3. Примените чек-лист к одному из проектов, которые делали на предыдущих занятиях (например, анализ транзакций). Что нужно доработать?

Список рекомендуемой литературы

1. Основная литература:

1. Воронова, Л. И. Big Data. Методы и средства анализа Электронный ресурс : Учебное пособие / Л. И. Воронова, В. И. Воронов. - Big Data. Методы и средства анализа, 2022-04-04. - Москва : Московский технический университет связи и информатики, 2016. - 33 с.
2. Воскобойников, Ю. Е. Обработка и анализ экспериментальных данных в пакетах MathCAD и Excel : учебное пособие / Ю. Е. Воскобойников. - Обработка и анализ экспериментальных данных в пакетах MathCAD и Excel, Весь срок охраны авторского права. - Электрон. дан. (1 файл). - Новосибирск : Новосибирский государственный архитектурно-строительный университет (Сибстрин), ЭБС АСВ, 2020. - 161 с.
3. Горелов, В.И. Анализ статистических данных Электронный ресурс : практикум / Т.Н. Ледащева / В.И. Горелов. - Анализ статистических данных, 2020-06-16. - Москва : Российская международная академия туризма, Университетская книга, 2015. - 120 с.
4. Горяинова, Е. Р. Прикладные методы анализа статистических данных : учебное пособие / Е.Р. Горяинова, А.Р. Панков, Е.Н. Платонов. - Москва : Издательский дом Высшей школы экономики, 2012. - 312 с.
5. Кемаева, М. В. Количественные методы анализа экономики Электронный ресурс / Кемаева М. В. : учебно-методическое пособие. - Нижний Новгород : ННГУ им. Н. И. Лобачевского, 2017. - 49 с.
6. Кулаичев А.П. Методы и средства комплексного анализа данных: Учебное пособие / А.П. Кулаичев. — М.: Форум, 2022. — 160 с.
7. Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с.
8. Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 188 с.
9. Селиванова, И.А. Построение и анализ алгоритмов обработки данных Электронный ресурс : учебно-методическое пособие / В.А. Блинов / И.А. Селиванова. - Построение и анализ алгоритмов обработки данных, Екатеринбург : Уральский федеральный университет, ЭБС АСВ, 2015. - 108 с.
10. Шнарева, Г. В. Анализ данных : учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. - Анализ данных, 2024-12-06. - Электрон. дан. (1 файл). - Симферополь : Университет экономики и управления, 2019. - 129 с.

2. Дополнительная литература:

1. Криволапов С.Я. Введение в анализ данных. Поиск структуры данных с применение языка Python: учебное пособие / С. Я. Криволапова. — Москва : ИНФРА-М, 2024. — 177 с.
2. Лакшманан В., Тайджани Д. «Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении», Санкт Петербург: Издательский

дом «Питер». – 2021. – 496с.

3. Мхитарян В.С. Анализ данных: учебник для академического бакалавриата / В.С. Мхитаряна. – Москва : Юрайт, 2020. – 490 с.

3. Интернет-ресурсы:

1. <https://www.online-python.com/> – онлайн компилятор Python.

2. <https://powerbi.microsoft.com/> - аналитическая система Power BI

3. <https://spark.apache.org/> - Apache Spark

4. Программное обеспечение:

1. Операционная система: Microsoft Windows 8: 2013-02(3000).
Бессрочная лицензия. Договор № 01-эа/13 от 25.02.2013. Окончание
бесплатной поддержки – 2023-01

2. Базовый пакет программ Microsoft Office (Word, Excel, PowerPoint)

3. Google Colab, онлайн компилятор – требуется доступ в интернет.

4. Интерпретатор и среда разработки (IDE) Python.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

для студентов специальности 38.05.01 «Экономическая безопасность»
по организации самостоятельной работы
по дисциплине «Технологии анализа больших данных»

Ставрополь
2026

Данные методические указания предназначены для оказания помощи студентам при изучении дисциплины «Технологии анализа больших данных». Работа содержит требования к выполнению индивидуального творческого проекта, порядок выполнения и защиты, рекомендации по оформлению.

Методические указания предназначены для студентов специальности 38.05.01 «Экономическая безопасность».

Индивидуальный творческий проект (ИТП) - самостоятельная работа студента, основной целью и содержанием которой является развитие навыков теоретических исследований, обобщений, оценка результатов исследований.

Индивидуальный творческий проект должен отразить умение студента с помощью системного анализа и осмысления фактического материала исследовать избранную тему и грамотно изложить результаты исследования.

Целью подготовки ИТП по дисциплине является глубокое изучение и усвоение предмета по одной из тем.

ИТП должен представлять собой самостоятельное изложение существа (содержания) основных вопросов избранной темы. Выполнение ИТП должно помочь студенту:

- научиться подбирать литературу по определенной теме;
- содействовать развитию и закреплению навыков теоретического и практического использования различных источников: учебной литературы, научных работ, нормативно-законодательных актов, статистических материалов;
- на основе полученных знаний выполнять теоретическую разработку избранной темы;
- выработать умение самостоятельно формулировать свои мысли и выводы, логично и последовательно доказывать их, уметь анализировать;
- научиться: правильно оформлять научно-справочный материал;
- выработать умение публичной защиты подготовленной работы (сделать доклад, дать ответы на вопросы, отстаивать свое мнение и т. д.).

1 ФОРМУЛИРОВКА ЗАДАНИЯ И ЕГО ОБЪЕМ

Наиболее важным и ответственным моментом, предопределяющим успех в подготовке ИТП, является выбор темы.

Темы ИТП охватывают наиболее актуальные проблемы в области изучения дисциплины. Тематика ИТП составлена в соответствии с программой по дисциплине и выдается студентам, в начале семестра. Тема ИТП выбирается студентами самостоятельно, по желанию, или выдается преподавателем. Прежде чем приступить к написанию ИТП, необходимо ознакомиться с заданием и подобрать необходимую литературу. После выбора темы составляется список литературы, опубликованных статей, необходимых справочных источников. Подбранную литературу следует зафиксировать согласно ГОСТ по библиографическому описанию произведений печати. Подбранная литература изучается в следующем порядке:

- знакомство с литературой, просмотр ее и выборочное чтение с целью общего представления проблемы и структуры будущей научной работы;
- исследование необходимых источников, сплошное чтение отдельных работ, их изучение, конспектирование необходимого материала;
- обращение к литературе для дополнений и уточнений на этапе написания ИТП.

Для разработки ИТП достаточно изучение 10 - 12 важнейших источников по избранной проблеме. При изучении литературы необходимо выбирать материал, не только подтверждающий позицию автора ИТП, но и материал для полемики.

Тематика ИТП разрабатывается, утверждается кафедрой и предоставляется студентам для ознакомления.

2 ОБЩИЕ ТРЕБОВАНИЯ К НАПИСАНИЮ ИТП

ИТП должен отвечать следующим требованиям. По содержанию:

- 1) в ИТП должны присутствовать характерные поисковые признаки:
 - раскрытие содержания основных концепций;
 - цитирование мнений некоторых специалистов по данной проблеме;
 - материал следует излагать кратко, не допуская дословного переписывания текста из литературных источников;
- 2) при обработке полученного материала автор должен:
 - систематизировать его по разделам;
 - выдвинуть и обосновать свои гипотезы;
 - определить свою позицию, точку зрения по рассматриваемой проблеме;
 - уточнить объем и содержание понятий, которыми приходится оперировать при разработке темы;
 - сформулировать определения и основные выводы, характеризующие результаты исследования;
 - окончательно уточнить структуру ИТП.
- 3) при использовании цитат, статистических и исторических данных и т. п. необходимо делать сноски с указанием источника: автор, название книги, место издания, год издания, страница цитируемого материала;
- 4) ИТП должен включать в себя:
 - введение;

- основная часть;
- заключение (выводы и предложения);
- список информационных источников;
- приложения (при необходимости).

Во *введении* следует раскрыть теоретическое и практическое значение избранной темы проекта, обосновать ее актуальность, определить цель и задачи, объект и предмет проекта. Объем введения не должен превышать двух страниц машинописного текста.

Основная часть состоит из разделов, подразделов (при необходимости) в соответствии с логической структурой изложения. Основная часть проекта посвящается анализу предметной области, эскизному и концептуальному проектированию, разработке и тестированию программного продукта и платформ. Разрабатывается документация к техническому заданию.

Объем основной части составляет не менее 40 страниц машинописного текста.

1. Исследовательская часть: теоретические основы анализа предметной области; методология исследования; обоснование выбора конкретных методов и инструментов

2. Аналитическая и реализационная глава: применение языков программирования для решения профессиональных задач; анализ больших данных в проектах; построение аналитических моделей и визуализация результатов, анализ массивов данных с помощью технологий больших данных или СУБД, предложения для принятия бизнес-решений.

3. Представление результатов. Описание процесса обработки данных и примененных аналитических методов. Результаты исследования в форме схем, графиков, таблиц, выводов и интерпретаций.

4. Обсуждение достоверности и значимости полученных выводов.

5. Критическая оценка методологии и возможных ограничений исследования.

Заключение содержит выводы и предложения с их кратким обоснованием в соответствии с поставленной целью и задачами, раскрывает значимость полученных результатов. Объем страниц заключения не должен превышать трех страниц машинописного текста.

Список использованных источников должен содержать не менее 10 источников. Список использованных источников отражает перечень источников, которые использовались при написании проектной работы, показывает глубину и широту изучаемой темы.

Объем проекта составляет *не менее 45 страниц машинописного текста*, не включая приложения.

ИТП должен полностью соответствовать утвержденной теме проекта, содержать элементы новизны, отражать теоретическую и практическую значимость, а описательная часть ИТП должен иметь следующую структуру: титульный лист, содержание, введение, основной текст, заключение, список использованных источников; приложения (при необходимости).

Приложения помещают после списка литературы и располагают в порядке появления на них ссылок в тексте ИТП. Они призваны облегчить восприятие содержания работы, и могут включать:

~ дополнительные материалы, иллюстрации вспомогательного характера, схемы, графики, программные коды;

~ документы, материалы, содержащие первичную информацию для анализа, таблицы статистических данных и т.д.

3 РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ РАБОТ НАД ЗАДАНИЕМ, ПРИМЕРНЫЙ КАЛЕНДАРНЫЙ ПЛАН ЕГО ВЫПОЛНЕНИЯ

Прежде чем приступить к написанию ИТП, студенту необходимо по выбранной теме задания изучить рекомендуемую как основную, так и дополнительную литературу, а также периодическую печать. Затем приступить к написанию работы.

Подготовка ИТП - это особый вид работы с литературой, прежде всего научной. Изучение литературы по теме следует начинать с учебников и учебных пособий, которые более доступны и позволяют определить круг рассматриваемых вопросов избранной темы. Однако эти источники не могут быть использованы в качестве теоретической основы ИТП. Являясь самостоятельным исследованием, он должен опираться на научную, монографическую литературу, а также журнальные статьи. Работа с монографиями включает ознакомление с заглавиями, аннотациями, предисловиями, в которых содержатся данные о структуре издания, его актуальности и отличие от других работ.

Подобный подход позволяет за максимально короткий срок определить ценность библиографических источников и составить их список.

Предварительное общее знакомство с литературой чрезвычайно полезно. Проведя систематизацию изученных работ, студент одновременно определяет способы будущей обработки того или иного материала. Изучение вопросов темы лучше начинать с основных обобщающих работ, постепенно переходя к специальным исследованиям. При изучении научной литературы надо обязательно обращать внимание на научно-справочный аппарат: примечания, подстрочные ссылки, библиографию, приложение и т. д.

Изучение специальной литературы может производиться двумя способами. В каждой научной работе материал, относящийся к теме, прочитывается и осмысливается сразу по всем вопросам. В этом случае после изучения каждого такого источника у студента складывается более или менее цельное представление о содержании ИТП. Но при таком способе труднее акцентировать внимание на каждом его вопросе в отдельности. Поэтому применяем и другой способ - изучение литературы по каждому отдельному вопросу.

При изучении литературы нужно записывать все встречающиеся спорные вопросы, выявлять точки зрения различных авторов, с тем чтобы, сопоставив их, получить правильное представление.

Изучив необходимую по данной проблеме литературу, автор должен разобраться в имеющихся точках зрения, показать умение обобщать, систематизировать, сопоставлять, кратко и концептуально излагать и оценивать идеи различных авторов и на этой основе формулировать собственную точку зрения или присоединяться к одной из изложенных в научных работах.

После ознакомления с литературой студент составляет план ИТП. План работы есть отображение ее структуры, под которой понимается порядок построения и взаимосвязь ее частей. Следует отметить, что часто встречается такой недостаток, как несоразмерность глав, разделов темы, что объясняется увлечением какой-то одной проблемой в ущерб другим.

ИТП рекомендуется выполнять в соответствии с графиком выполнения, предложенным преподавателем. На основе графика студент сам должен планировать ход своего исследования, которое включает индивидуальное составление календарного плана научного исследования и плана предполагаемого ИТПе.

Календарный план исследования включает следующие элементы:

- выбор и формулирование проблемы, разработка плана исследования и предварительного плана ИТП;
- сбор и изучение неходкого материала, подбор литературы;
- анализ собранного материала, теоретическая разработка проблемы;
- сообщение о предварительных результатах исследования;
- литературное оформление исследовательской проблемы;
- обсуждение работы (на семинаре, в студенческом Научном обществе, на конференции и т. п.).

4 ОФОРМЛЕНИЕ ГРУППОВОЙ ПРОЕКТНОЙ РАБОТЫ

4.1 Общие требования

Текст ИТП выполняют с использованием компьютера на одной стороне листа белой бумаги, формата А4, шрифт – Times New Roman 14-го размера, межстрочный интервал – 1,5.

Страницы ИТП следует нумеровать арабскими цифрами, соблюдая сквозную нумерацию по всему тексту работы. Номер страницы проставляют в центре нижней части листа без точки. Титульный лист включают в общую нумерацию страниц отчета. Номер страницы на титульном листе не проставляют. Иллюстрации и таблицы, расположенные на отдельных листах, включают в общую нумерацию страниц отчета.

Текст отчета следует печатать, соблюдая следующие размеры полей: левое - 30 мм, правое - 15 мм, верхнее и нижнее - 20 мм. Абзацный отступ должен быть одинаковым по всему тексту отчета и равен 1,25 см. Сокращения слов и словосочетаний на русском и иностранных европейских языках оформляют в соответствии с требованиями ГОСТ 7.11, ГОСТ 7.12.

Выравнивание основного текста – по ширине. Выравнивание заголовков разделов, подразделов и пунктов – по центру. При применении выравнивания по центру необходимо убрать красную строку (отступ) абзаца.

Рисунки и таблицы оформляются в строгом соответствии с общими требованиями; они должны содержать соответственно подрисуночные подписи и названия, иметь порядковый номер, располагаться по мере обращения к ним в тексте, сопровождаться ссылками, пояснениями, выводами и т. п.

Плотность текста должна быть равномерной (без разрежений и уплотнений). Опечатки, помарки, графические неточности, повреждение листов в работе не допускаются. Работа должна быть сброшюрована плотно, без перекосов, смятия листов и т. п.

4.2 Изложение текста

Изложение текста ИТП следует вести от третьего лица.

Текст ИТП должен быть кратким, четким и не допускать различных толкований. В тексте должны применяться научные термины, обозначения и определения, установленные соответствующими стандартами, а при их отсутствии – общепринятые в научно-технической и методической литературе.

В тексте не допускается применять:

- ~ обороты разговорной речи;
- ~ для одного и того же понятия различные научно-технические термины, близкие по смыслу (синонимы);

иностранные слова при наличии равнозначных слов и терминов на русском языке;
сокращения понятий без предварительной расшифровки.

4.3 Построение ИТП

Заголовки структурных элементов ИТП следует располагать в середине строки без точки в конце и печатать прописными буквами, не подчеркивая.

Основную часть ИТП следует делить на разделы и подразделы. Каждый раздел должен содержать законченную информацию.

Разделы должны иметь порядковые номера в пределах всей ИТП, обозначенные арабскими цифрами.

Заголовки и нумерация разделов и подразделов, а также номера страниц, указанные в структурном элементе «Содержание» должны полностью соответствовать тексту работы.

Иллюстрации, таблицы и распечатки с ЭВМ, расположенные на отдельных листах, учитывают, как одну страницу и включают в общую нумерацию страниц.

Внутри подразделов могут быть приведены перечисления. Перед каждым перечислением следует ставить дефис или, при необходимости ссылки в тексте документа на одно из перечислений, строчную букву, после которой ставится скобка.

4.4 Оформление иллюстраций

Иллюстрации (чертежи, графики, схемы, компьютерные распечатки, диаграммы, фотоснимки) следует располагать в работе непосредственно после текста, в котором они упоминаются впервые, или на следующей странице. Иллюстрации могут быть в компьютерном исполнении, в том числе и цветные. На все иллюстрации должны быть даны ссылки в работе.

Чертежи, графики, диаграммы, схемы, иллюстрации должны соответствовать требованиям государственных стандартов Единой системы конструкторской документации (ЕСКД). Допускается выполнение чертежей, графиков, диаграмм, схем посредством использования компьютерной печати.

Иллюстрации, за исключением иллюстраций приложений, следует нумеровать арабскими цифрами сквозной нумерацией.

Если рисунок один, то он обозначается «Рисунок 1». Слово «Рисунок» и его наименование располагают посередине строки. Допускается нумеровать иллюстрации в пределах раздела. В этом случае номер иллюстрации состоит из номера раздела и порядкового номера иллюстрации, разделенных точкой. Например, Рисунок 1.1.

Иллюстрации, при необходимости, могут иметь наименование и пояснительные данные (подрисуночный текст). Слово «Рисунок» и наименование помещают после пояснительных данных и располагают следующим образом: Рисунок 1 – Этапы проектирования. Пример оформления приведен на рисунке 1.



Рисунок 1 – Этапы проектирования

Иллюстрации каждого приложения обозначают отдельной нумерацией арабскими цифрами с добавлением перед цифрой обозначения приложения. Например, Рисунок А.1. При ссылках на иллюстрации следует писать «...в соответствии с рисунком 2» при сквозной нумерации и «...в соответствии с рисунком 1.2» при нумерации в пределах раздела.

4.5 Оформление таблиц

Таблицы применяют для лучшей наглядности и удобства сравнения показателей. Наименование таблицы, при его наличии, должно отражать ее содержание, быть точным, кратким. Наименование таблицы следует помещать над таблицей слева, без абзацного отступа в одну строку с ее номером через тире.

Таблицу следует располагать в ИТП непосредственно после текста, в котором она упоминается впервые, или на следующей странице. На все таблицы должны быть ссылки в работе. При ссылке следует писать слово «таблица» с указанием ее номера. Таблицу с большим числом строк допускается переносить на другой лист (страницу). При переносе части таблицы на другой лист (страницу) слово «Таблица», ее номер и наименование указывают один раз слева над первой частью таблицы, а над другими частями также слева пишут слова «Продолжение таблицы» и указывают номер таблицы.

Таблицу с большим количеством граф допускается делить на части и помещать одну часть под другой в пределах одной страницы. Если строки и графы таблицы выходят за формат страницы, то в первом случае в каждой части таблицы повторяется головка, во втором случае – боковик. При делении таблицы на части допускается ее головку или боковик заменять соответственно номером граф и строк. При этом нумеруют арабскими цифрами графы и (или) строки первой части таблицы.

Если повторяющийся в разных строках графы таблицы текст состоит из одного слова, то его после первого написания допускается заменять кавычками; если из двух и более слов, то при первом повторении его заменяют словами «То же», а далее – кавычками. Ставить кавычки вместо повторяющихся цифр, марок, знаков, математических и химических символов не допускается. Если цифровые или иные данные в какой-либо строке таблицы не приводят, то в ней ставят прочерк. Цифровой материал, как правило, оформляют в виде таблиц в соответствии с рисунком 2.

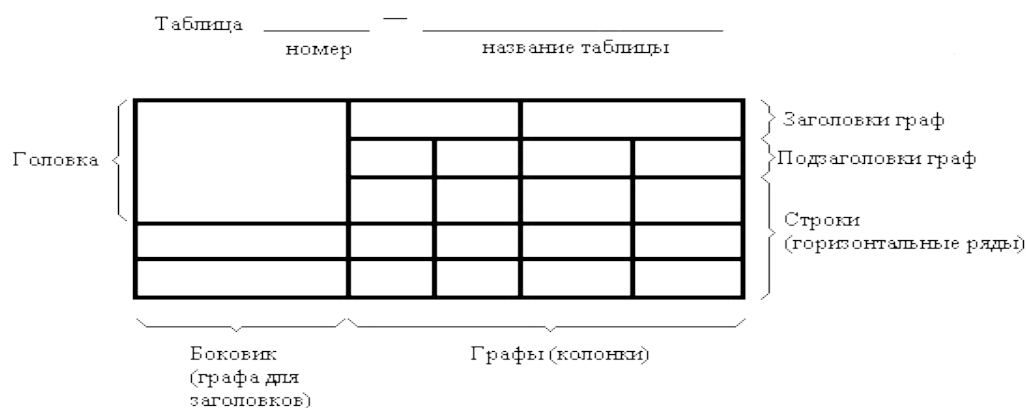


Рисунок 2 – Оформление таблицы

Таблицы, за исключением таблиц приложений, следует нумеровать арабскими цифрами сквозной нумерацией.

Допускается нумеровать таблицы в пределах раздела. В этом случае номер таблицы состоит из номера раздела и порядкового номера таблицы, разделенных точкой.

Таблицы каждого приложения обозначают отдельной нумерацией арабскими цифрами с добавлением перед цифрой обозначения приложения. Если в отчете одна таблица, то она должна быть обозначена «Таблица 1» или «Таблица В.1», если она приведена в приложении.

Заголовки граф и строк таблицы следует писать с прописной буквы в единственном числе, а подзаголовки граф – со строчной буквы, если они составляют одно предложение с заголовком, или с прописной буквы, если они имеют самостоятельное значение. В конце заголовков и подзаголовков таблиц точки не ставят.

Таблицы слева, справа и снизу, как правило, ограничивают линиями. Допускается применять размер шрифта в таблице меньший, чем в тексте. Разделять заголовки и подзаголовки боковика и граф диагональными линиями не допускается.

Горизонтальные и вертикальные линии, разграничивающие строки таблицы, допускается не проводить, если их отсутствие не затрудняет пользование таблицей.

Заголовки граф, как правило, записывают параллельно строкам таблицы. При необходимости допускается перпендикулярное расположение заголовков граф.

Головка таблицы должна быть отделена линией от остальной части таблицы.

4.6 Оформление приложения

Приложение оформляют как продолжение данного документа на последующих его листах или выпускают в виде самостоятельного документа.

В тексте ИТП на все приложения должны быть даны ссылки. Приложения располагают в порядке ссылок на них в тексте ИТП.

Каждое приложение следует начинать с новой страницы с указанием наверху посередине страницы слова «Приложение», его обозначения. Приложение должно иметь заголовок, который записывают симметрично относительно текста с прописной буквы отдельной строкой.

Приложения обозначают заглавными буквами русского алфавита, начиная с А. После слова «Приложение» следует буква, обозначающая его последовательность. Если в ИТП одно приложение, оно обозначается «Приложение А».

Текст каждого приложения, при необходимости, может быть разделен на разделы, подразделы, пункты, подпункты, которые нумеруют в пределах каждого приложения. Перед номером ставится обозначение этого приложения.

Приложения должны иметь общую с остальной частью документа сквозную нумерацию страниц. Приложение или несколько приложений могут быть оформлены в виде отдельной книги отчета, при этом на титульном листе под номером книги следует писать слово «Приложение». При необходимости такое приложение может иметь раздел «Содержание».

4.7 Ссылки на источники

Библиографические ссылки входят в состав примечаний, которые являются обязательными для проектной работы. Примечания, помимо библиографических ссылок, могут содержать авторский текст.

По месту расположения в работе библиографические ссылки могут быть:

- внутритекстовыми, помещенными в основной текст работы;
- подстрочными, вынесенными из основного текста работы вниз полосы работы (в сноску);
- затекстовыми, вынесенными за текст всей работы или ее части (в выноску).

5 ПОРЯДОК ВЫПОЛНЕНИЯ ГРУППОВОЙ ПРОЕКТНОЙ РАБОТЫ И ПОРЯДОК ПОДГОТОВКИ ГРУППОВОЙ ПРОЕКТНОЙ РАБОТЫ К ЗАЩИТЕ

Подготовка ИТП состоит из нескольких этапов:

~ выбор темы, обоснование ее актуальности, формулировка цели и задач проектной работы;

~ поиск фактического материала, анализ методов и технологий для выполнения проекта;

- ~ обработка и анализ полученной информации;
- ~ практическая часть;
- ~ оформление ИТП в соответствии с установленными требованиями.

В формулировке темы ИТП должна прослеживаться основная идея проекта.

Руководитель ИТП оказывает студентам помощь в разработке проекта весь период выполнения ИТП, рекомендует необходимую литературу, справочные материалы и другие источники по теме, проводит систематические консультации, проверяет выполнение проектной работы по частям и в целом, составляет задания для выполнения ИТП. В период работы над ИТП руководитель систематически заслушивает промежуточные отчеты студентов о степени готовности работы, контролирует работу, помогает студентам подготовиться к защите ИТП.

6 ЗАЩИТА ИНДИВИДУАЛЬНОГО ТВОРЧЕСКОГО ПРОЕКТА

Индивидуальный творческий проект в электронном и распечатанном виде сдается преподавателю не менее, чем за 3 дня до защиты.

К защите студенты подготавливают презентацию и доклад по итогам выполненного проекта.

Защита ИТП происходит публично на практических занятиях в соответствии со следующим порядком защиты ИТП:

- ~ представление темы и студента, работающего над проектом;
- ~ краткое (не более 10 мин.) сообщение студента с использованием наглядных материалов и компьютерной техники об основных результатах выполненного проекта;
- ~ вопросы преподавателя и присутствующих студентов;
- ~ ответы студента на заданные вопросы.

Продолжительность защиты одного ИТП, как правило, не должна превышать 20 минут.

В своем сообщении студент должен отразить следующие вопросы:

- ~ актуальность темы, цель, решаемые задачи и выбранные технологии /методы;
- ~ обоснование теоретических и практических решений, используемых в процессе выполнения группового проекта;
- ~ представление, полученных результатов выполнения ИТП;
- ~ основные выводы и практические рекомендации, связанные с возможностью реализации полученных результатов в профессиональной области.

При оценке ИТП учитываются следующее:

- ~ актуальность и обоснованность темы, полученных практических результатов исследования;
- ~ самостоятельность и глубина проработки темы ИТП, причем учитывается и обсуждается вклад каждого слушателя группы;
- ~ глубина анализа научно-технической литературы, используемой для решения проектной задачи;
- ~ выбор цифровых технологий, их соответствие сформулированной проблеме, целям и задачам проектной работы;
- ~ степень практической значимости;
- ~ убедительное и аргументированное выступление, продемонстрированное слушателями в процессе публичной защиты ИТП.

Результаты защиты ИТП определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

7 ТЕМАТИКА ИНДИВИДУАЛЬНЫХ ТВОРЧЕСКИХ ПРОЕКТОВ

ИТП (сроки выполнения 1-14 недели, выдача задания – 1 неделя) выполняется по

следующей тематике:

1. Проект: «Прогнозирование оттока (churn rate) клиентов мобильного оператора и оценка экономического эффекта от мероприятий по их удержанию»

Суть: на основе данных о длительности клиентской жизни, тарифах, обращениях в поддержку построить модель, предсказывающую вероятность ухода клиента. Рассчитать, какие клиенты наиболее «ценны» для удержания и какие бонусы им предложить.

Ключевые задачи: построение прогнозной модели (логистическая регрессия, случайный лес) на Python; SQL-запросы для агрегации данных по клиентам; расчет LTV и экономического обоснования программ удержания.

2. Проект: «Анализ рынка и потенциала запуска нового SaaS-продукта для автоматизации email-маркетинга малого бизнеса»

Суть: провести анализ конкурентов (от крупных игроков до нишевых), оценить объем рынка и спрос, проанализировать тематические сообщества для выявления «боли» потенциальных клиентов. Спрогнозировать возможную долю рынка.

Ключевые задачи: парсинг данных о конкурентах (Python, BeautifulSoup); анализ тональности обсуждений на форумах; построение модели TAM/SAM/SOM; визуализация конкурентного ландшафта в DataLens.

3. Проект: «Оптимизация рекламного бюджета: анализ окупаемости инвестиций (ROI) в различные каналы привлечения трафика для онлайн-кинотеатра»

Суть: построить атрибуционную модель (например, Data-Driven Attribution или Markov Chains) на Python, чтобы понять реальный вклад каждого канала (контекст, социальные сети, email) в конверсию. Перераспределить бюджет для максимизации ROI.

Ключевые задачи: сбор данных из Яндекс.Метрики/Google Analytics и рекламных кабинетов; построение и сравнение атрибуционных моделей; создание дашборда для мониторинга ROI по каналам.

4. Проект: «Причинно-следственный анализ влияния акций и скидок на ключевые маркетинговые показатели в розничной сети»

Суть: выявить, как проведение промо-акций влияет не только на продажи акционного товара, но и на средний чек, трафик и продажи сопутствующих товаров. Оценить, нет ли «каннибализации» продаж обычных товаров.

Ключевые задачи: Анализ временных рядов (Python, Pandas); регрессионный анализ для установления причинно-следственных связей; анализ ассоциативных правил (Apriori algorithm) для поиска паттернов в корзинах.

1. Проект: «Создание прототипа системы для автоматизированного анализа отзывов о продукте на маркетплейсе и выявления ключевых проблем пользователей»

Суть: разработать алгоритм на Python, который парсит отзывы, проводит тональный анализ (sentiment analysis) и тематическое моделирование (Topic Modeling, например, LDA), чтобы автоматически выявлять основные позитивные и негативные темы.

Ключевые задачи: парсинг отзывов (Python); NLP-обработка текста (библиотеки NLTK/spaCy); визуализация ключевых тем в виде облака тегов и дашборда.

2. Проект: «Геомаркетинговый анализ для оптимизации сети офлайн-точки (кофейни) и зоны доставки»

Суть: проанализировать открытые данные о транспортной и пешеходной доступности, плотности конкурирующих точек, демографическом и социально-экономическом профиле районов для выбора оптимального места открытия новой кофейни.

Ключевые задачи: работа с геоданными (Python, библиотеки Geopandas, Folium);

анализ горячих точек; построение интерактивных карт в Yandex DataLens; расчет потенциального трафика и емкости территории.

3. Проект: «Анализ воронки продаж мобильного приложения и поиск «узких мест», приводящих к потере пользователей»

Суть: построить и проанализировать воронку от установки приложения до целевого действия (покупка, подписка). Методами анализа последовательностей (Sequence Analysis) выявить типичные пути «успешных» и «неуспешных» пользователей.

Ключевые задачи: агрегация данных логов событий (SQL); построение воронок и карт путей (Yandex DataLens, Python); когортный анализ для оценки оттока на разных этапах.

4. Проект: «Оценка влияния активности бренда в социальных сетях на трафик сайта и объем продаж»

Суть: провести причинно-следственный анализ, чтобы определить, как публикации постов, запуск рекламных кампаний и упоминания бренда влияют на ключевые бизнес-показатели. Выявить лаг между публикацией и ростом трафика.

Ключевые задачи: сбор данных через API соцсетей (Python); анализ временных рядов с учетом лагов (Python, Statsmodels); корреляционный и регрессионный анализ; дашборд с интеграцией данных из соцсетей и веб-аналитики.

5. Проект: «Построение прогнозной модели спроса на сезонный товар (например, товары для зимних видов спорта) для оптимизации складских запасов и логистики»

Суть: на основе исторических данных о продажах, маркетинговой активности, макроэкономических показателей и погодных данных построить модель, которая за 2-3 месяца спрогнозирует спрос на следующий сезон.

Ключевые задачи: прогнозное моделирование (временные ряды, ARIMA, Prophet) на Python; анализ внешних факторов; оценка экономического эффекта от снижения затрат на хранение и недополученной прибыли из-за дефицита.

6. Проект: «Прогнозирование оттока (churn rate) клиентов мобильного оператора и оценка экономического эффекта от мероприятий по их удержанию»

Суть: на основе данных о длительности клиентской жизни, тарифах, обращениях в поддержку построить модель, предсказывающую вероятность ухода клиента. Рассчитать, какие клиенты наиболее «ценны» для удержания и какие бонусы им предложить.

Ключевые задачи: построение прогнозной модели (логистическая регрессия, случайный лес) на Python; SQL-запросы для агрегации данных по клиентам; расчет LTV и экономического обоснования программ удержания.

7. Проект: «Анализ рынка и потенциала запуска нового SaaS-продукта для автоматизации email-маркетинга малого бизнеса»

Суть: провести анализ конкурентов (от крупных игроков до нишевых), оценить объем рынка и спрос, проанализировать тематические сообщества для выявления «болей» потенциальных клиентов. Спрогнозировать возможную долю рынка.

Ключевые задачи: парсинг данных о конкурентах (Python, BeautifulSoup); анализ тональности обсуждений на форумах; построение модели TAM/SAM/SOM; визуализация конкурентного ландшафта в DataLens.

8. Проект: «Оптимизация рекламного бюджета: анализ окупаемости инвестиций (ROI) в различные каналы привлечения трафика для онлайн-кинотеатра»

Суть: построить атрибуционную модель (например, Data-Driven Attribution или Markov Chains) на Python, чтобы понять реальный вклад каждого канала (контекст,

социальные сети, email) в конверсию. Перераспределить бюджет для максимизации ROI.

Ключевые задачи: сбор данных из Яндекс.Метрики/Google Analytics и рекламных кабинетов; построение и сравнение атрибуционных моделей; создание дашборда для мониторинга ROI по каналам.

9. Проект: «Причинно-следственный анализ влияния акций и скидок на ключевые маркетинговые показатели в розничной сети»

Суть: выявить, как проведение промо-акций влияет не только на продажи акционного товара, но и на средний чек, трафик и продажи сопутствующих товаров. Оценить, нет ли «каннибализации» продаж обычных товаров.

Ключевые задачи: Анализ временных рядов (Python, Pandas); регрессионный анализ для установления причинно-следственных связей; анализ ассоциативных правил (Apriori algorithm) для поиска паттернов в корзинах.

10. Проект: «Анализ эффективности контент-маркетинга: определение тем и форматов, которые генерируют максимальный лидогенерационный потенциал»

Суть: связать данные из CRM (лиды, продажи) с данными о поведении пользователей на сайте (просмотр статей, вебинаров). Определить, какой контент ведет к максимальной конверсии в покупателя.

Ключевые задачи: SQL для соединения данных из разных источников; кластерный анализ путей пользователей; построение воронок в Yandex DataLens; регрессионный анализ для определения значимости разных типов контента.

11. Проект: «Анализ туристического потока и сезонности в регионе [Название региона] для оптимизации маркетинговой активности и ценовой политики»

Суть: проанализировать открытые данные о въездном турпотоке, бронированиях отелей и поисковых запросах для выявления точных пиков, спадов и «долинных» сезонов. Построить прогнозную модель и предложить календарь маркетинговых активностей и гибкого ценообразования.

Ключевые задачи: анализ временных рядов (Python, Pandas); сбор данных с сайтов бронирования (парсинг, с осторожностью) и Яндекс.Wordstat; прогнозирование (Prophet, ARIMA); визуализация сезонности в Yandex DataLens.

12. Проект: «Сегментация туристов, посещающих [Название региона], и разработка персонализированных турпакетов»

Суть: на основе анализа данных из соцсетей, опросов и поведения на туристических сайтах выделить ключевые сегменты (например, «молодые экотуристы», «семьи с детьми на пляжный отдых», «люди старшего возраста на культурные туры»). Для каждого сегмента предложить целевой турпродукт и каналы коммуникации.

Ключевые задачи: кластеризация (K-Means, методом главных компонент для снижения размерности) на Python; анализ тональности обсуждений региона; создание интерактивных портретов сегментов в DataLens.

13. Проект: «Конкурентный анализ рынка гостиничного размещения в [Город-курорт]: определение рыночных ниш и точек роста»

Суть: провести глубокий анализ предложения отелей и гостиниц по ценовым сегментам, набору услуг, локациям и отзывам. Выявить неудовлетворенные потребности туристов и предложить концепции для нового размещения или ребрендинга существующих.

Ключевые задачи: парсинг данных с Tripadvisor, Booking.com и Островка (Python, BeautifulSoup/Selenium); анализ ассортимента и цен конкурентов; сетевой анализ для

выявления кластеров конкурентов; экономическое обоснование предложений.

14. Проект: «Выявление драйверов спроса на внутренний туризм в регионе: причинно-следственный анализ на основе макроэкономических и медийных данных»

Суть: проанализировать, как такие факторы, как курс рубля, стоимость авиаперелетов за границу, упоминаемость региона в СМИ и инфлюенсерами, влияют на объем бронирований в местных отелях.

Ключевые задачи: множественный регрессионный анализ (Python, Statsmodels); сбор данных из открытых источников (ЦБ РФ, Aviasales, медиаметрии); визуализация корреляций и выводов.

15. Проект: «Разработка геоинформационной системы для анализа туристической инфраструктуры и логистики [Региона]»

Суть: создать интерактивную карту, на которую нанесены достопримечательности, отели, точки общепита, транспортные узлы. Проанализировать транспортную доступность, выявить «белые пятна» и предложить маршруты для улучшения логистики туристов.

Ключевые задачи: работа с геоданными (Python, Geopandas, Folium); анализ изохрон доступности; построение карт в Yandex DataLens; использование API Яндекс.Карт.

16. Проект: «Оптимизация ассортимента услуг туристического сервиса [например, прокат оборудования, экскурсии] на основе анализа рыночного спроса и рентабельности»

Суть: проанализировать, какие дополнительные услуги наиболее востребованы туристами и приносят максимальную маржинальность. Предложить план по оптимизации ассортимента (что добавить, что убрать, на чем сделать акцент в продвижении).

Ключевые задачи: ABC-анализ; анализ ассоциативных правил (алгоритм Apriori в Python) для поиска связей между услугами; расчет рентабельности; экономическое обоснование решений.

17. Проект: «Оценка влияния мероприятий (фестивалей, спортивных соревнований) на туристическую привлекательность региона»

Суть: количественно оценить, как проведение крупного события повлияло на турпоток, занятость в гостиницах, упоминаемость в СМИ и общую известность региона. Разработать методику оценки эффективности инвестиций в события.

Ключевые задачи: сравнительный анализ временных рядов «до/после»; анализ медиаполя; опросы; расчет ROI мероприятий для регионального бюджета.

18. Проект: «Оптимизация логистики и управление запасами скоропортящейся продукции (FMCG) в региональной дистрибьюторской компании»

Суть: проанализировать данные о продажах, остатках на складах и сроках годности товаров, чтобы построить прогнозную модель спроса. Цель — минимизировать уровень Out-of-Stock (дефицита) и потери от списаний (шорт-даты), оптимизировав график пополнения запасов и логистические маршруты.

Ключевые задачи: Анализ временных рядов для прогнозирования спроса по каждой SKU (Python, prophet, sktime). Расчет оптимального уровня страхового запаса с учетом сезонности и lead time (времени поставки). Причинно-следственный анализ факторов, ведущих к возникновению шорт-даты и дефицита. Разработка в SQL прототипа системы мониторинга критических остатков. Создание дашборда в Yandex DataLens для отслеживания KPI: уровень сервиса, коэффициент оборачиваемости, объем списаний. Расчет экономического эффекта от внедрения модели (снижение потерь, увеличение оборачиваемости).

19. Проект: «Анализ лояльности к локальным продовольственным брендам региона

и разработка мероприятий по ее повышению на основе данных цифровых активностей»

Суть: исследовать узнаваемость и восприятие локальных брендов (например, молочный комбинат, хлебокомбинат) по сравнению с федеральными. Проанализировать вовлеченность аудитории в соцсетях и реакцию на промо-акции. Предложить данные-драйверы стратегии по усилению лояльности.

Ключевые задачи: Парсинг и тональный анализ упоминаний брендов в социальных сетях и на форумах региона (Python, nltk, vaderSentiment). Анализ эффективности контента и рекламных активностей брендов в соцсетях (сравнение ER, CTR и т.д.). Проведение и анализ онлайн-опроса (можно смоделировать данные) для сегментации потребителей по уровню лояльности (NPS-методология). Сравнительный анализ с федеральными конкурентами. Разработка персонализированной программы лояльности и контент-стратегии, основанной на выводах анализа, с экономическим обоснованием затрат.

20. Проект: «Разработка системы сквозной аналитики для интернет-магазина малого бизнеса: от рекламного клика до повторной покупки»

Суть: настроить сбор данных с сайта, рекламных кабинетов и CRM. Построить единую воронку продаж, чтобы определить, какие каналы привлечения дают самых лояльных и прибыльных клиентов, а не просто первые заказы.

Ключевые задачи: Настройка API-интеграций (Python), работа с Яндекс.Метрикой/Google Analytics, построение атрибуционных моделей, создание дашборда в Yandex DataLens для контроля CAC, LTV и ROI.

21. Проект: «Сегментация клиентов локального сервиса (например, кофейни, салона красоты, фитнес-студии) для повышения эффективности программ лояльности»

Суть: проанализировать данные о покупках и посещениях, чтобы выделить сегменты (например, «утренние постоянные клиенты», «редкие гости с большим средним чеком»). Разработать для каждого сегмента персонализированные предложения в программе лояльности.

Ключевые задачи: RFM-анализ и кластеризация на Python, расчет LTV клиента, разработка и экономическое обоснование различных механик лояльности.

22. Проект: «Анализ ценовой конкуренции на локальном рынке услуг для малого бизнеса (например, репетиторства, ремонта)»

Суть: собрать и проанализировать данные по ценам конкурентов, их позиционированию и отзывам. Определить «ценовой коридор» и оптимальную точку для конкуренции: ценовой лидер или премиум-сегмент.

Ключевые задачи: Парсинг сайтов и агрегаторов услуг (Python), анализ тональности отзывов, построение карты позиционирования «цена/качество», расчет экономики для разных сценариев.

23. Проект: «Прогнозирование спроса для малого производства (например, пекарня, кондитерская, свечной цех) для оптимизации закупок и производства»

Суть: построить модель, которая на основе исторических данных о продажах, дня недели, погоды и маркетинговой активности прогнозирует спрос на каждый вид продукции. Это позволит сократить объем нереализованной продукции и дефицит.

Ключевые задачи: Прогнозное моделирование временных рядов (Python, prophet), анализ факторов, влияющих на спрос, создание в DataLens дашборда с прогнозом на следующий день/неделю.

24. Проект: «Оптимизация рекламного бюджета малого бизнеса в социальных сетях на основе A/B тестирования и анализа окупаемости»

Суть: провести серию A/B тестов рекламных креативов, таргетинга и рекламных форматов. Построить систему анализа, которая в реальном времени показывает, какие объявления приносят лиды, а не просто клики.

Ключевые задачи: Работа с API соцсетей (Python), планирование и статистическая проверка A/B тестов, расчет CPO (Cost Per Order), создание дашборда для оперативного контроля рекламных кампаний.

25. Проект: «Выявление продуктов-доноров и продуктов-локомотивов в ассортименте малого розничного магазина»

Суть: провести ABC-XYZ анализ ассортимента, анализ ассоциативных правил (какие товары покупают вместе) и расчет маржинальности. Перераспределить мерчандайзинг и маркетинговые усилия, чтобы стимулировать покупки высокомаржинальных и связанных товаров.

Ключевые задачи: Глубокий анализ данных о продажах (SQL, Python, библиотеки для анализа ассоциативных правил), визуализация результатов в виде схем выкладки товаров и таблиц приоритетов.

26. Проект: «Автоматизация конкурентного анализа для малого бизнеса: создание прототипа парсера и дашборда»

Суть: разработать скрипт на Python, который автоматически собирает с сайтов конкурентов данные о ценах, акциях, ассортименте и отзывах. Визуализировать собранные данные в простом и понятном дашборде для владельца бизнеса.

Ключевые задачи: Веб-скрейпинг (Python, BeautifulSoup/Selenium), очистка и структурирование данных, настройка регулярного сбора, создание интуитивного дашборда в Yandex DataLens.

27. Проект: «Анализ оттока клиентов салона красоты или фитнес-клуба и разработка системы удержания»

Суть: проанализировать данные о посещениях и покупках, чтобы выявить паттерны поведения, предшествующие уходу клиента. Построить модель, предсказывающую вероятность оттока, и разработать механики для их возврата (персональные предложения, напоминания).

Ключевые задачи: Построение прогнозной модели (логистическая регрессия, случайный лес), когортный анализ, расчет стоимости удержания vs привлечения, разработка триггерных email- или SMS-рассылок.

28. Проект: «Расчет unit-экономики и поиск точек роста для малого SaaS-бизнеса или онлайн-школы»

Суть: построить детальную финансовую модель бизнеса, основанную на данных. Определить, за счет каких каналов привлечения, изменения цены или снижения оттока можно быстрее достичь точки безубыточности и увеличить прибыль.

Ключевые задачи: Расчет LTV, CAC, Retention Rate; сценарное моделирование в Python; анализ воронки от лида до платежа; построение дашборда с ключевыми метриками стартапа.

29. Проект: «Геомаркетинговый анализ для обоснования открытия новой точки малого бизнеса (кофейня, пекарня, пункт выдачи)»

Суть: проанализировать открытые данные о транспортной и пешеходной доступности, плотности населения, наличии конкурентов и арендных ставках для выбора оптимального местоположения.

Ключевые задачи: Работа с геоданными (Python, библиотека geopandas), анализ

тепловых карт трафика, построение интерактивных карт в Yandex DataLens, расчет потенциальной проходимости и емкости рынка.

30. Проект: «Оптимизация маркетинговой воронки для локальной студии йоги или творческих мастер-классов»

Суть: проанализировать все этапы пути клиента: от первого контакта с рекламой до покупки абонемента и прихода на занятие. Выявить «узкие места», где теряется большая часть потенциальных клиентов, и предложить способы их «расшивки».

Ключевые задачи: Построение и анализ воронки продаж, проведение и анализ опросов, A/B тестирование посадочных страниц и ценовых предложений, расчет конверсии на каждом этапе.

31. Проект: «Внедрение системы контроля ключевых показателей (KPI) для малого бизнеса на основе Yandex DataLens»

Суть: разработать и внедрить единый дашборд для владельца бизнеса, который в режиме реального времени отображает все ключевые метрики: выручку, трафик, средний чек, конверсию, исполнение бюджета. Это позволит принимать решения на основе данных, а не интуиции.

Ключевые задачи: Интеграция данных из разных источников (CRM, Google Sheets, сайт), проектирование и создание понятного дашборда в Yandex DataLens, формулировка релевантных KPI для конкретного бизнеса.

32. Проект: «Оптимизация логистики и прогнозирование спроса для регионального производителя молочной продукции с целью сокращения потерь и увеличения оборачиваемости»

Суть: проанализировать данные о продажах, остатках на складах, сроках годности и логистических маршрутах. Построить модель прогнозирования спроса по каждой товарной единице (SKU) для оптимизации производства и графика отгрузок, минимизируя объем просроченной продукции.

Ключевые задачи: Анализ временных рядов и прогнозное моделирование спроса (Python, prophet, sktime). Расчет оптимального уровня страховых запасов с учетом сезонности и времени доставки. Анализ причинно-следственных связей между рекламными активностями, погодой и скачками спроса. Разработка дашборда в Yandex DataLens для мониторинга KPI: уровень потерь, оборачиваемость запасов, выполнение плана отгрузок. Экономическое обоснование эффективности предлагаемой модели (расчет сокращения списаний).

33. Проект: «Оценка потенциала экспорта продукции регионального производителя (например, минеральной воды, кондитерских изделий, дикоросов) в страны СНГ и ЕАЭС»

Суть: провести комплексное исследование, чтобы оценить возможности и риски выхода на внешние рынки. Проанализировать таможенное регулирование, конкурентную среду, потребительские предпочтения и логистические издержки.

Ключевые задачи: Анализ открытых данных международной торговли (статистика ООН, ФТС России). Парсинг и тональный анализ отзывов на товары-аналоги на зарубежных маркетплейсах (Python, NLP). Исследование требований к сертификации и нормативных барьеров. Расчет полной себестоимости продукта с учетом экспортной пошлины и логистики. Построение финансовой модели и определение целевых стран для экспансии.

34. Проект: «Разработка системы мониторинга и анализа производственных потерь на предприятии по переработке овощей и фруктов»

Суть: проанализировать полный цикл от приемки сырья до отгрузки готовой продукции, чтобы выявить основные точки возникновения потерь (брак, усушка, порча). Построить дашборд для оперативного контроля ключевых показателей эффективности (ОЕЕ, уровень брака).

Ключевые задачи: Сбор и интеграция данных с различных этапов производственного цикла (SQL). Причинно-следственный анализ (например, с использованием диаграмм Исикавы) для определения корневых причин потерь. Визуализация производственного цикла и "узких мест" в Yandex DataLens. Расчет экономического эффекта от внедрения предлагаемых мероприятий по снижению потерь (например, оптимизация параметров оборудования, улучшение условий хранения).

35. Проект: «Анализ восприятия бренда регионального производителя продуктов здорового питания и разработка стратегии цифрового продвижения»

Суть: исследовать цифровой след бренда: узнаваемость, sentiment упоминаний в соцсетях и на форумах, активность и вовлеченность аудитории. Сравнить с федеральными конкурентами. На основе анализа предложить data-driven стратегию контент-маркетинга и таргетированной рекламы.

Ключевые задачи: парсинг и анализ упоминаний бренда и конкурентов в социальных сетях и на тематических ресурсах (Python). Тематическое моделирование (LDA) и тональный анализ для выявления сильных и слабых сторон продукта в глазах потребителей. Анализ эффективности текущих рекламных кампаний и целевой аудитории. Разработка персонализированной стратегии продвижения, включая контент-план и параметры таргетинга, с расчетом потенциального охвата и ROMI.

36. Проект: «Построение модели скоринга для автоматизации одобрения кредитных заявок в онлайн-банке»

Суть: на основе исторических данных о клиентах и их кредитной истории создать модель машинного обучения, которая прогнозирует вероятность дефолта. Оптимизировать баланс между кредитным риском и уровнем одобрения заявок.

Ключевые задачи: знакомство с регуляторными требованиями (например, 115-ФЗ); построение и валидация моделей (логистическая регрессия, градиентный бустинг) на Python; анализ важности признаков; создание прототипа системы для оценки новых заявок.

37. Проект: «Сегментация клиентской базы розничного банка для разработки персонализированных предложений по инвестиционным продуктам»

Суть: на основе данных о транзакциях, остатках на счетах и демографии клиентов выделить сегменты с разной склонностью к сбережениям и риску. Спрогнозировать потенциальный спрос на различные финансовые инструменты.

Ключевые задачи: кластерный анализ (K-Means, DBSCAN) с использованием Python; прогнозный анализ; создание профилей сегментов и разработка для них конкретных продуктовых пакетов; экономическая оценка потенциальной выручки.

38. Проект: «Оптимизация тарифной сетки страховых продуктов (КАСКО/ОСАГО) с применением предиктивного моделирования»

Суть: проанализировать исторические данные о страховых случаях, характеристиках автомобилей и водителей. Построить модель для более точной оценки риска и предложить дифференцированные тарифы, которые будут одновременно конкурентными и прибыльными.

Ключевые задачи: регрессионный анализ на Python; работа с категориальными данными; оценка финансового эффекта от внедрения новой тарифной политики;

визуализация результатов в DataLens.

39. Проект: «Анализ качества сети (QoE) и его влияния на отток абонентов мобильного оператора»

Суть: объединить телеметрические данные о качестве связи (скорость, дропы вызовов) с данными об оттоке клиентов. Выявить критические пороги качества, при которых клиенты с наибольшей вероятностью уходят, и рассчитать экономическую целесообразность улучшения сети в конкретных районах.

Ключевые задачи: сопряжение больших массивов данных из разных систем (SQL); причинно-следственный и регрессионный анализ (Python); построение интерактивных карт с «болевыми точками» в DataLens.

40. Проект: «Прогнозирование пиковых нагрузок на инфраструктуру облачного провайдера»

Суть: на основе данных об использовании вычислительных ресурсов клиентами, календаря событий и макротрендов построить модель прогнозирования нагрузки. Цель — оптимизировать планирование закупок оборудования и затраты на электроэнергию.

Ключевые задачи: анализ временных рядов высокой частоты (Python, `prophet`, `sktime`); выявление сезонных и событийных паттернов; создание дашборда для оперативного планирования мощностей.

41. Проект: «Прогнозирование спроса на сырье и оптимизация цепочки поставок для промышленного предприятия»

Суть: проанализировать внутренние данные о производстве и внешние — о рыночных ценах на сырье и макроэкономических индикаторах. Построить модель для прогнозирования потребности и оптимального времени закупки с целью минимизации издержек.

Ключевые задачи: многовариантное прогнозирование; оптимизационное моделирование (библиотеки `pulp`, `ortools` в Python); расчет экономии от внедрения модели.

42. Проект: «Анализ эффективности работы торговых представителей B2B-компании с помощью методов Data Science»

Суть: оценить вклад каждого представителя в выручку, учитывая «сложность» территории (плотность клиентов, активность конкурентов). Выявить факторы успеха и построить систему мотивации, справедливо учитывающую внешние условия.

Ключевые задачи: кластеризация клиентов и территорий; регрессионный анализ для нормализации KPI; разработка алгоритма расчета бонусов; визуализация результатов на картах в DataLens.

43. Проект: «Динамическое ценообразование для продавца на маркетплейсе (например, Wildberries, OZON)»

Суть: разработать алгоритм, который автоматически корректирует цену товара на основе анализа цен конкурентов, остатков на складе, сезонного спроса и рейтинга карточки.

Ключевые задачи: парсинг данных маркетплейса (Python, Selenium); разработка и бэктестирование ценовых стратегий; создание прототипа бота для управления ценами.

44. Проект: «Поиск и валидация нишевых товарных категорий для запуска на маркетплейсе»

Суть: просканировать рынок, чтобы найти категории с высоким спросом, низкой конкуренцией и хорошей маржинальностью. Провести глубокий анализ товаров-лидеров и отзывов для формирования конкурентного предложения.

Ключевые задачи: парсинг каталогов и отзывов; анализ ассортимента и цен; тональный анализ (NLP) для выявления «болей»; расчет TAM/SAM/SOM для выбранных ниш.

45. Проект: «Оптимизация логистики и запасов медицинских товаров в сети аптек»

Суть: спрогнозировать спрос на лекарства и изделия медицинского назначения с учетом сезонности (грипп, аллергии), эпидемиологической обстановки и сроков годности. Снизить уровень дефицита и списаний.

Ключевые задачи: прогнозное моделирование для сотен SKU; ABC-XYZ анализ; построение дашборда для категорийных менеджеров в DataLens.

46. Проект: «Анализ эффективности планограмм и тепловых карт движения покупателей в торговом зале»

Суть: сопоставить данные о продажах с видеоданными о движении покупателей (можно использовать синтетические или анонимизированные данные). Оценить, какие зоны зала и выкладки генерируют наибольшую выручку, и предложить корректировки планограмм.

Ключевые задачи: работа с пространственными данными; анализ ассоциативных правил; A/B тестирование новых планограмм; расчет потенциального роста выручки.

47. Проект: «Прогнозирование потребления энергии и выявление аномалий на основе данных с умных счетчиков»

Суть: построить модель для прогнозирования суточного потребления энергии домохозяйством или предприятием. Разработать алгоритм для автоматического обнаружения аномальных паттернов, которые могут указывать на неисправность оборудования или несанкционированное подключение.

Ключевые задачи: анализ временных рядов; обнаружение аномалий (Isolation Forest, Autoencoders); создание оповещений в дашборде DataLens.

48. Проект:** «Анализ данных городского трафика для оптимизации работы общественного транспорта и светофорных циклов»

Суть: на основе открытых данных о движении транспорта и пассажиропотоке выявить «узкие места» и часы пик. Предложить изменения в маршрутной сети и настройках светофоров для уменьшения заторов.

Ключевые задачи: работа с геоданными и временными рядами (Python, `geopandas`); симуляция сценариев; визуализация предложений на интерактивной карте в DataLens.

49. Проект: «Рекомендательная система для стримингового сервиса (кино или музыка) на основе коллаборативной фильтрации»

Суть: разработать прототип рекомендательного алгоритма, который анализирует поведение пользователей (просмотры, оценки) и предлагает персонализированный контент. Оценить потенциальное влияние на вовлеченность и удержание.

Ключевые задачи: применение методов коллаборативной фильтрации (библиотеки `surprise`); оценка точности модели (precision, recall); проектирование архитектуры простой рекомендательной системы.

50. Проект: «Построение модели оценки рыночной стоимости объектов недвижимости»

Суть: на основе данных об объектах (площадь, район, этаж, год постройки) и исторических цен продаж создать модель автоматической оценки (аналог appraisal). Проанализировать ключевые факторы, влияющие на стоимость.

Ключевые задачи: регрессионный анализ (линейная регрессия, градиентный

бустинг) на Python; парсинг данных с порталов недвижимости; интерпретация модели.

51. Проект: «Анализ факторов успеваемости студентов онлайн-курса и прогнозирование отсева»

Суть: проанализировать данные о взаимодействии студентов с платформой (просмотр лекций, выполнение заданий, активность на форуме). Выявить ранние признаки, предсказывающие низкую успеваемость или отсев, и предложить меры вмешательства.

Ключевые задачи: анализ образовательных данных (EDA); построение прогнозной модели; разработка системы оповещений для кураторов.

52. Проект: «Оптимизация маршрутов доставки «последней мили» для курьерской службы»

Суть: разработать алгоритм, который строит оптимальные маршруты для курьеров с учетом адресов, временных окон, пробок и приоритета заказов. Цель — снизить пробег и повысить количество выполненных заказов в смену.

Ключевые задачи: решение задачи маршрутизации (VRP) с использованием Python-библиотек ('ortools', 'vgpu'); работа с геоданными и API карт; расчет экономии топлива и времени.

53. Проект: «Предиктивное обслуживание оборудования на основе анализа данных с датчиков (IoT)»

Суть: используя данные телеметрии с промышленного оборудования, построить модель, которая предсказывает вероятность поломки. Перейти от планового обслуживания к обслуживанию по состоянию, сократив простой и затраты на ремонт.

Ключевые задачи: анализ временных рядов с датчиков; построение модели классификации; разработка прототипа системы оповещений; расчет ROI проекта.

54. Проект: «Анализ качества сервиса и выявление детерминант NPS в сети ресторанов быстрого питания»

Суть: связать операционные данные (время приготовления заказа, текучесть персонала) с данными опросов NPS и онлайн-отзывами. Определить ключевые факторы, влияющие на удовлетворенность гостя, и предложить корректирующие действия.

Ключевые задачи: создание единого аналитического слоя данных (SQL); регрессионный анализ; тональный анализ отзывов; визуализация KPI по точкам в DataLens.

55. Проект: «Прогнозирование урожайности и оптимизация использования ресурсов (прецизионное земледелие)»

Суть: на основе спутниковых снимков, данных о погоде и состоянии почвы построить модель для прогнозирования урожайности по полям. Предложить карты дифференцированного внесения удобрений и полива для максимизации результата.

Ключевые задачи: работа с растровыми данными и ГИС (Python); регрессионный анализ; расчет экономической эффективности методики.

56. Проект: «Анализ монетизации и поведения игроков в free-to-play мобильной игре»

Суть: провести когортный анализ, построить воронки конверсии игроков из бесплатных в платящих. Выявить точки, где происходит отсев потенциальных платежеспособных игроков, и предложить изменения в геймдизайне или монетизационной модели.

Ключевые задачи: когортный и юнит-экономический анализ; построение сложных воронок в DataLens; A/B тестирование игровых механик.

57. Проект: «Оптимизация коммуникаций и повышение эффективности сбора

пожертвований для благотворительного фонда»

Суть: проанализировать историю взаимодействия с донорами (рассылки, события, пожертвования). Сегментировать доноров и построить модель, предсказывающую вероятность повторного пожертвования. Разработать персонализированную стратегию коммуникаций для увеличения пожизненной ценности донора.

Ключевые задачи: RFM-анализ и кластеризация; прогнозное моделирование; экономическая оценка кампаний (ROI).

58. Проект: «Предиктивная аналитика текучести кадров в крупной компании»

Суть: на основе анонимизированных данных сотрудников (отдел, должность, длительность работы, результаты оценки, история повышений) выявить факторы, влияющие на увольнение. Построить модель для выявления сотрудников группы риска и предложить превентивные HR-мероприятия.

Ключевые задачи: работа с конфиденциальными данными (соблюдение этики и закона); построение прогнозной модели; расчет стоимости замены сотрудника и экономии от снижения текучести.

59. Проект: «Анализ рынка подержанных автомобилей и разработка инструмента для оценки справедливой стоимости»

Суть: собрать данные с основных площадок о продаже подержанных авто (модель, год, пробег, регион, цена). Построить модель, которая учитывает основные параметры и рыночную конъюнктуру для определения реалистичной цены. Проанализировать самые быстро/медленно продающиеся модели.

Ключевые задачи: массовый парсинг данных (Python); регрессионный анализ и feature engineering; создание простого веб-интерфейса или калькулятора для оценки.

8 СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

8.1 Основная литература:

1. Воронова, Л. И. Big Data. Методы и средства анализа Электронный ресурс : Учебное пособие / Л. И. Воронова, В. И. Воронов. - Big Data. Методы и средства анализа, 2022-04-04. - Москва : Московский технический университет связи и информатики, 2016. - 33 с.
2. Воскобойников, Ю. Е. Обработка и анализ экспериментальных данных в пакетах MathCAD и Excel : учебное пособие / Ю. Е. Воскобойников. - Обработка и анализ экспериментальных данных в пакетах MathCAD и Excel, Весь срок охраны авторского права. - Электрон. дан. (1 файл). - Новосибирск : Новосибирский государственный архитектурно-строительный университет (Сибстрин), ЭБС АСВ, 2020. - 161 с.
3. Горелов, В.И. Анализ статистических данных Электронный ресурс : практикум / Т.Н. Ледащева / В.И. Горелов. - Анализ статистических данных, 2020-06-16. - Москва : Российская международная академия туризма, Университетская книга, 2015. - 120 с.
4. Горяинова, Е. Р. Прикладные методы анализа статистических данных : учебное пособие / Е.Р. Горяинова, А.Р. Панков, Е.Н. Платонов. - Москва : Издательский дом Высшей школы экономики, 2012. - 312 с.
5. Кемаева, М. В. Количественные методы анализа экономики Электронный ресурс / Кемаева М. В. : учебно-методическое пособие. - Нижний Новгород : ННГУ им. Н. И. Лобачевского, 2017. - 49 с.
6. Кулаичев А.П. Методы и средства комплексного анализа данных: Учебное пособие / А.П. Кулаичев. — М.: Форум, 2022. — 160 с.
7. Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с.
8. Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л.

Н. Тындыкарь. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 188 с.

9. Селиванова, И.А. Построение и анализ алгоритмов обработки данных Электронный ресурс : учебно-методическое пособие / В.А. Блинов / И.А. Селиванова. - Построение и анализ алгоритмов обработки данных, Екатеринбург : Уральский федеральный университет, ЭБС АСВ, 2015. - 108 с.

10. Шнарева, Г. В. Анализ данных : учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. - Анализ данных, 2024-12-06. - Электрон. дан. (1 файл). - Симферополь : Университет экономики и управления, 2019. - 129 с.

8.2 Дополнительная литература:

1. Криволапов С.Я. Введение в анализ данных. Поиск структуры данных с применение языка Python: учебное пособие / С. Я. Криволапова. – Москва : ИНФРА-М, 2024. – 177 с.

2. Лакшманан В., Тайджани Д. «Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении», Санкт Петербург: Издательский дом «Питер». – 2021. – 496с.

3. Мхитарян В.С. Анализ данных: учебник для академического бакалавриата / В.С. Мхитаряна. – Москва : Юрайт, 2020. – 490 с.

4. Селиванова, И.А. Построение и анализ алгоритмов обработки данных Электронный ресурс : учебно-методическое пособие / В.А. Блинов / И.А. Селиванова. - Построение и анализ алгоритмов обработки данных, Екатеринбург : Уральский федеральный университет, ЭБС АСВ, 2015. - 108 с. - Книга находится в базовой версии ЭБС IPRbooks. - ISBN 978-5-7996-1489-8

5. Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении. – СПб.: Питер, 2021. – 496 с.: ил.

6. Kachur, A., Osin, E., Davydov, D. et al. Assessing the Big Five personality traits using real-life static facial images. Sci Rep 10, 8487 (2020). <https://doi.org/10.1038/s41598-020-65358-6>

7. Обработка данных в SPSS — URL: <https://nafi.ru/academy/prepodavatelayam-spss/>

8.3 Интернет-ресурсы:

1. <https://www.online-python.com/> – онлайн компилятор Python.

2. <https://powerbi.microsoft.com/> - аналитическая система Power BI

3. <https://spark.apache.org/> - Apache Spark

8.4. Программное обеспечение:

1. Операционная система: Microsoft Windows 8: Бессрочная лицензия. Договор № 01-эа/13 от 25.02.2013.

2. Операционная система: Microsoft Windows 10: Бессрочная лицензия. Договор № 544-21 от 08.06.2021.

3. Базовый пакет программ Microsoft Office (Word, Excel, PowerPoint). Microsoft Office Standard 2013: договор № 01-эа/13 от 25.02.2013г., Лицензия Microsoft Office <https://support.microsoft.com/ru-ru/lifecycle/search/16674>

4. Google Colab, онлайн компилятор – требуется доступ в интернет.

5. Интерпретатор и среда разработки (IDE) Python.