Пензенский государственный университет

Факультет вычислительной техники

Кафедра «Вычислительная техника»

**Отчёт**

по лабораторной работе № 4

## на тему «Изучение основных принципов построения обучаемых алгоритмов»

по дисциплине «Программные средства кибертехнических систем»

Вариант № 6

выполнили: ст-ты гр. 22ВВП1

Захаров А. С.

Беляев Д. И.

Демин М. С

Проверили:

Карамышева Н.С.

Зинкин С.А.

2025

**Цель:** изучить основные принципы построения обучаемых алгоритмов

**Задание:**

1.Получить описание заданной предметной области и набор соответствующих ей данных.

2. Провести исследование в соответствии с планом работы в рамках первой фазы исследования.

3. Подготовить отчёт по результатам первой фазы работ.

Отчёт по изучению материала по теме 1.1 должен содержать информацию в соответствии со следующей структурой:

1. Бизнес цель.

2. Ресурсы и риски.

3. Цель анализа данных.

4. План проекта.

Вариант 6

#### Задача кредитного скоринга (15-й столбец набора данных

#### {1 – кредит выдан, 2 – кредит не выдан}) по 14 признакам.

#### Ссылка на репозиторий:

#### https://archive.ics.uci.edu/ml/datasets/Statlog+%28Australian+Credit+

#### Approval%29

**1. Бизнес цель**

Целью исследования является разработка модели для кредитного скоринга, которая предсказывает вероятность получения кредита для клиента на основе 14 признаков. Признаки могут включать возраст, доход, кредитную историю, тип работы, количество задолженностей и другие факторы, которые могут влиять на решение банка о предоставлении кредита. Это поможет банкам и финансовым учреждениям улучшить процесс принятия решений, снижая риски невозврата кредита и повышая эффективность работы с клиентами.

Модель должна быть способна точно предсказать, будет ли клиент одобрен в кредите (1 – кредит выдан, 2 – кредит не выдан), основываясь на доступных данных. Это улучшит процессы оценки и повысит точность кредитных решений, помогая уменьшить риски.

**2. Ресурсы и риски**

**2.1 Доступные ресурсы**

Источники данных: Набор данных с 14 признаками, включая социально-демографические характеристики клиентов и финансовые показатели. Этот набор будет использован для построения модели предсказания.

**2.2 Возможные риски**

* Неполнота или низкое качество данных: В данных могут быть пропуски или ошибки, что потребует дополнительной обработки и очистки.
* Дисбаланс классов в данных: Если количество положительных (кредиты выданы) и отрицательных (кредиты не выданы) случаев сильно отличается, это может привести к искажению результатов модели.
* Переобучение модели: Использование слишком сложных моделей может привести к переобучению на обучающих данных, что снизит ее обобщающую способность на новых данных.
* Сложность интерпретации модели: Некоторые алгоритмы (например, нейронные сети) могут быть сложными для интерпретации, что ограничит их применимость в бизнесе.

**3. Цель анализа данных**

* Определить ключевые факторы, влияющие на решение о предоставлении кредита: Исследовать 14 признаков и их влияние на одобрение кредита.
* Построить модель предсказания: Создать модель, которая сможет предсказать вероятность получения кредита, что обеспечит автоматизацию процесса принятия решений.
* Выбрать метрики для оценки качества модели: Для этой задачи можно использовать такие метрики, как точность, F1-мера, ROC-AUC, чтобы оценить производительность модели.
* Предоставить рекомендации по улучшению процесса принятия решений: На основе анализа данных предоставить рекомендации для повышения качества кредитного скоринга.

**4. План проекта**

**4.1 Изучение данных (Data Understanding)**

1. Сбор данных: Описание и подготовка набора данных, который содержит 14 признаков, включая информацию о клиенте, его финансовом состоянии и истории кредитования.
2. Анализ данных: Изучение статистики по признакам, выявление пропусков и аномальных значений. Анализ распределения классов (кредит выдан/не выдан).
3. Исследование зависимостей: Оценка корреляций между признаками и целевой переменной (выдан/не выдан кредит), визуализация данных с использованием графиков и диаграмм.
4. Документирование: Описание всех этапов работы с данными, выявленные проблемы, принятые решения.

**4.2 Подготовка данных (Data Preparation)**

1. Отбор признаков: Определение наиболее значимых признаков, которые влияют на решение о кредите.
2. Очистка данных: Обработка пропусков, устранение выбросов и аномальных значений.
3. Преобразование данных: Нормализация числовых переменных, кодирование категориальных признаков, создание новых признаков, если необходимо.
4. Генерация новых признаков: Применение методов feature engineering, чтобы улучшить предсказательную способность модели.

**4.3 Моделирование (Modeling)**

1. Выбор алгоритмов: Логистическая регрессия, ближайших соседей, и другие методы для классификации.
2. Разделение данных: Деление данных на обучающую, валидационную и тестовую выборки для предотвращения переобучения.
3. Обучение моделей: Построение и обучение нескольких моделей, оценка их производительности по выбранным метрикам.
4. Анализ важности признаков: Оценка того, какие признаки наиболее влияют на результат модели.

**4.4 Оценка (Evaluation)**

1. Оценка качества модели: Оценка точности предсказания модели и ее соответствие бизнес-целям.
2. Анализ результатов: Анализ сильных и слабых сторон моделей, проверка на переобучение и ошибки.
3. Выбор лучшей модели: Определение модели, которая обеспечит наилучшее сочетание точности и вычислительной эффективности.

**4.5 Внедрение (Deployment)**

1. Разработка прототипа: Разработка приложения или внедрение модели в существующую систему для реальных кредитных решений.
2. Мониторинг качества модели: Регулярное отслеживание точности предсказаний модели и адаптация к изменениям данных.
3. Обновление модели: Модификация модели по мере появления новых данных или изменений в финансовых условиях.

Критерии устаревания модели:

1. Снижение точности предсказаний.
2. Изменения в финансовых характеристиках клиентов или политике кредитования.
3. Появление более эффективных алгоритмов.
4. Обратная связь от пользователей модели (например, от кредитных экспертов).

**Вывод:** изучили основные принципы построения обучаемых алгоритмов, установили цели и задачи, разработали план проекта.