

ТЕХНИЧЕСКОЕ ЗАДАНИЕ
на выполнение опытно-конструкторской работы (ОКР)
по теме: «Разработка программного модуля для прогнозирования оттока
клиентов на основе методов машинного обучения»

ОСНОВАНИЕ ДЛЯ РАЗРАБОТКИ

Основанием для разработки является план выполнения выпускной квалификационной работы (ВКР) студента Спириянова Максима Дмитриевича по специальности/направлению 09.03.01 — Технологии разработки программного обеспечения.

Утвержденный ИТиЭО
РГПУ им. А. И. Герцена
«_» _____ 2025 г.

НАИМЕНОВАНИЕ И ОБОЗНАЧЕНИЕ РАБОТЫ, ПРОДУКЦИИ

Наименование работы: «Разработка программного модуля для прогнозирования оттока клиентов».

Наименование продукции (ПП): «Программный модуль прогнозирования оттока клиентов «ChurnPredictor 1.0» (далее – Модуль).

Условное обозначение продукции: PP.ChurnPredictor-1.0.

НАИМЕНОВАНИЕ ЗАКАЗЧИКА И ИСПОЛНИТЕЛЯ РАБОТЫ

Заказчик: РГПУ им. А. И. Герцена, кафедра ИТиЭО.

Научный руководитель: Зав. кафедрой д.п.н., проф. Е.З.Власова.

Исполнитель работ (Разработчик): Студент Спириянов Максим Дмитриевич, группа 4об_ИВТ-2/22.

ЦЕЛЬ И НАЗНАЧЕНИЕ РАЗРАБОТКИ

Цель работы: Исследование методов машинного обучения для задачи бинарной классификации и создание на их основе работоспособного программного модуля, демонстрирующего возможность прогнозирования оттока клиентов.

Назначение продукции: Разрабатываемый Модуль предназначен для автоматизации процесса выявления клиентов с высокой вероятностью прекращения пользования услугами компании (оттока) на основе исторических данных. Модуль носит демонстрационно-исследовательский характер и может быть использован в учебном процессе, а также как прототип для интеграции в корпоративные аналитические системы.

ТРЕБОВАНИЯ К РАБОТЕ И ПРОГРАММНОМУ ПРОДУКТУ

1. Требования к работе

- Провести анализ предметной области (задача прогнозирования оттока).
- Провести сравнительный анализ алгоритмов машинного обучения для задач классификации: логистическая регрессия, случайный лес, градиентный бустинг (XGBoost/CatBoost/LightGBM), метод опорных векторов.
- Выбрать и обосновать выбор метрик качества (Accuracy, Precision, Recall, F1-score, ROC-AUC).

- Разработать и реализовать pipeline обработки данных и построения моделей.
- Провести эксперименты, оценить и сравнить эффективность выбранных алгоритмов на открытом датасете.
- Разработать программный модуль с веб-интерфейсом для интерактивного использования.
- Оформить техническую и пользовательскую документацию.

2. Требования к программному продукту

2.1. Требования к функциональным характеристикам

Модуль должен обеспечивать выполнение следующих функций:

1. Загрузка данных: Возможность загрузки файла с данными о клиентах в формате CSV.
2. Предварительный анализ данных: Автоматическое отображение основных статистик (количество записей, пропуски, типы данных).
3. Предобработка данных: Автоматическое выполнение этапов:
 - Обработка пропущенных значений.
 - Кодирование категориальных признаков.
 - Масштабирование числовых признаков.
 - Разделение на обучающую и тестовую выборки.
4. Обучение моделей: Реализация обучения не менее 3-х алгоритмов машинного обучения (например, Случайный лес, Логистическая регрессия, Градиентный бустинг) с возможностью настройки основных гиперпараметров через конфигурационный файл.
5. Оценка моделей: Автоматический расчет и визуализация набора метрик качества (Accuracy, Precision, Recall, F1-score, ROC-AUC, матрица ошибок).
6. Прогнозирование: Возможность загрузки нового набора данных (без целевой переменной) и получения прогноза (вероятности оттока и бинарного класса) с использованием лучшей сохраненной модели.
7. Визуализация: Построение графиков: важность признаков, ROC-кривая, распределение прогнозируемых вероятностей.

2.2. Требования к надежности

1. Модуль должен корректно обрабатывать ошибки ввода (некорректный формат файла, несоответствие структуры данных).

- При возникновении внутренних ошибок выполнения должна предоставляться понятная текстовая информация для диагностики.

2.3. Условия эксплуатации

Клиентская часть (веб-интерфейс): Современный веб-браузер (Google Chrome, Mozilla Firefox, Microsoft Edge последних версий).

Серверная часть: Изолированная среда Python 3.8+ на компьютере разработчика/пользователя.

2.4. Требования к составу и параметрам технических средств

- Персональный компьютер (ноутбук) с ОС Windows 10/11, Linux или macOS.
- Процессор с тактовой частотой не менее 2 ГГц.
- Оперативная память не менее 8 ГБ (рекомендуется 16 ГБ для работы с большими датасетами).
- Свободное дисковое пространство не менее 1 ГБ.

2.5. Требования к информационной и программной совместимости

- Основной язык программирования: Python 3.8+.
- Ключевые библиотеки: scikit-learn, pandas, numpy, matplotlib, seaborn, flask/dash (или streamlit для прототипа), xgboost/catboost.
- Формат входных данных: CSV файл с кодировкой UTF-8, разделитель – запятая.
- Выходные данные: Файлы с прогнозами (CSV), графики (PNG), сохраненная модель (pickle или joblib формат).

2.6. Требования к маркировке и упаковке

Не предъявляются (программный продукт распространяется в виде архива с исходным кодом и инструкцией).

2.7. Требования к транспортированию и хранению

Не предъявляются.

3. Требования к документации

Разработать комплект документов в электронном виде:

1. Пояснительная записка к ВКР в соответствии с требованиями учебного заведения.
2. Руководство пользователя: Краткое описание установки зависимостей, запуска Модуля и пошагового выполнения всех его функций.
3. Текст программы (исходный код) с комментариями.

4. Технико-экономические показатели

Работа выполняется в учебных целях. Оценка экономической эффективности не требуется.

5. Стадии и этапы работы

Стадия 1: Исследовательская (1 мес.)

- Этап 1.1. Анализ предметной области и постановка задачи.
- Этап 1.2. Анализ и выбор алгоритмов машинного обучения.
- Этап 1.3. Поиск и подготовка тестового набора данных.

Стадия 2: Проектирование (1 мес.)

- Этап 2.1. Проектирование архитектуры программного модуля.
- Этап 2.2. Проектирование pipeline обработки данных и обучения.
- Этап 2.3. Проектирование простого веб-интерфейса.

Стадия 3: Реализация (2 мес.)

- Этап 3.1. Реализация кода предобработки данных и обучения моделей.
- Этап 3.2. Реализация веб-интерфейса и логики взаимодействия.
- Этап 3.3. Интеграция и отладка компонентов.

Стадия 4: Тестирование и оформление (1 мес.)

- Этап 4.1. Проведение экспериментов, сравнение моделей.
- Этап 4.2. Написание документации.
- Этап 4.3. Подготовка презентации и демонстрации.

6. Порядок контроля и приемки

1. Поэтапный контроль со стороны научного руководителя.
2. Предварительная демонстрация работоспособного прототипа.

3. Приемка работы: Защита выпускной квалификационной работы перед Государственной экзаменаціонной комиссией (ГЭК) с представлением:
 - Пояснительной записки.
 - Действующего программного модуля.
 - Презентации и отчета о результатах экспериментов.
4. Критерий приемки: Полное выполнение требований, изложенных в данном ТЗ, и успешная защита ВКР.

ИСТОЧНИКИ РАЗРАБОТКИ

При разработке следует руководствоваться:

1. ГОСТ 15.016-2016 Система разработки и постановки продукции на производство (СРПП). Техническое задание.
2. ГОСТ 34.602-89 Информационная технология. Комплекс стандартов на автоматизированные системы. Техническое задание на создание автоматизированной системы.
3. Научные и учебные публикации в области машинного обучения, классификации и анализа оттока клиентов.
4. Документация к библиотекам Python (scikit-learn, pandas, flask и т.д.).