

ТЕХНИЧЕСКОЕ ЗАДАНИЕ
на разработку программного сервиса
интеллектуального анализа данных

1. Название	Продолжительность поездки на такси по Нью-Йорку			
2. Исполнение	Групповое			
3. Разработчики	1. Королёв Д.А. 2. Чегодаева Е.Г.			
4. Цель разработки	Основываясь на индивидуальных характеристиках поездки, спрогнозировать продолжительность каждой поездки в тестовом наборе, и провести их анализ.			
5. Решаемые задачи	1. Представить историю поездок, а также данные, содержащие погодные условия 2. Провести анализ и удаление выбросов 3. Создать и обучить модель на основе метода градиентного бустинга (LightGBM) 4. Вычислить ошибку предсказания (RMSLE)			
6. Сроки исполнения	6.1 Начало:	10.11.2022 г.	6.2 Окончание:	31.12.2022 г.
7. Основные требования к разработке				
8. Данные	8.1 Источник данных	Данные, содержащие историю поездок Данные, содержащие погодные условия Данные, содержащие информацию о дорогах NYC		
	8.2 Параметры обучающих данных	Годовая история поездок на такси; история погодных условий; протяжённость Нью-Йоркских дорог		
	8.3 Формат признакового вектора x	$X = \{x_1, x_2, \dots, x_n\}$, x_1 – дата и время включения счетчика x_2 – дата и время выключения счетчика x_3 – количество пассажиров в транспортном средстве x_4 – продолжительность поездки в секундах ...		
	8.4 Формат вектора y (опционально)	$Y = \{y_1\}$, где $y_1 \in (15; 200)$ – продолжительность поездки (в минутах)		

		m = 1
9. Функционал	9.1 Функции предварительной обработки (с трансформацией/без трансформации признакового пространства)	Оптимизация гиперпараметров при помощи библиотеки Optuna
	9.2 Функции содержательной обработки (метод DM, методы DM и принцип интегрирования результата, полученного отдельными методами)	Градиентный бустинг (LightGBM); Классификация статистическая (Наивный байесовский классификатор); Регрессия;
10. Результаты	10.1 Представление результатов/что позволят установить (формулировка полученных неочевидных знаний, способствующих принятию на практике конкретных управленческих решений)	Прогнозы длительности поездки на такси представляются в виде таблицы по требованию заказчика
11. Исследования	11.1 Исследовательская компонента (оценка эффективности и пределов применимости используемых методов, рекомендации по использованию методов в решении аналогичных задач)	Выбор оптимальных гиперпараметров для получения наилучшего прогноза; В результате оценки эффективности использованных методов будут сформированы выводы и даны рекомендации.

Ответственный исполнитель _____ Королёв Д.А.

Заказчик _____ А.В. Замятин

_____ Чегодаева Е.Г.