**ТЕХНИЧЕСКОЕ ЗАДАНИЕ**

**на разработку программного сервиса**

**интеллектуального анализа данных**

| 1. **Название** | Предсказание эффективности лечения препаратом на основе данных пациента | | | | |
| --- | --- | --- | --- | --- | --- |
| 1. **Исполнение** | Групповое | | | | |
| 1. **Разработчики** | 1. Осипова А.Д. (К-ближайших соседей)  2. Локотецкий М.Е. (Градиентный бустинг) | | | | |
| 1. **Цель разработки** | Создать инструмент для предсказания эффективности лечения препаратом на основе данных пациента (возраст, пол, группа крови, заболевание, препарат) | | | | |
| 1. **Решаемые задачи** | 1. Отбор данных. 2. Предварительная обработка (чистка данных). 3. Сделать выводы о наличии зависимости группы крови и частоты выявления заболевания. 4. Построение классификационной модели. | | | | |
| 1. **Сроки исполнения** | 6.1 Начало: | 30.10.2023 г. | 6.2 Окончание: | | 22.12.2023 г. |
| 1. **Основные требования к разработке** | | | | | |
| 1. **Данные** | **8.1 Источник данных** | | | <https://www.kaggle.com/datasets/prasad22/healthcare-dataset> | |
| **8.2 Параметры обучающих данных** | | | 3000 векторов; каждый вектор имеет 15 показателей, значимых – 5 и 1 (прогнозируемый) результат тестирования | |
| **8.3 Формат признакового вектора x** | | | **X =** {*x1*,*x2*,…,*xn*},  где  *x1 – возраст*  *x2 – пол*  *x3 – группа-крови (+резус-фактор)*  *x4 – заболевание*  *x5 – используемый препарат*  n = 5 | |
| **8.4 Формат вектора y (опционально)** | | | **Y =** {*y1*,*y2*,…,*ym*},  где  *y1– результат лечения*  *m* = 1 | |
| 1. **Функционал** | **9.1 Функции предварительной обработки (с трансформацией/без трансформации признакового пространства)** | | | Сокращение объема данных (исключение векторов со схожими признаками) | |
| **9.2 Функции содержательной обработки (метод DM, методы DM и принцип интегрирования результата, полученного отдельными методами)** | | | К-ближайших соседей, градиентный бустинг | |
| 1. **Результаты** | **10.1 Представление результатов/что позволят установить (формулировка полученных неочевидных знаний, способствующих принятию на практике конкретных управленческих решений)** | | | Модель позволяет определить наиболее вероятный результат лечения на основе исходных данных о пациенте | |
| 1. **Исследования** | **11.1 Исследовательская компонента (оценка эффективности и пределов применимости используемых методов, рекомендации по использованию методов в решении аналогичных задач)** | | | Проверка гипотезы о наличии зависимости результата лечения от перечисленных признаков.  В результате проведенной работы будут сделаны выводы и даны рекомендации. | |

Ответственные исполнители: \_\_\_\_\_\_\_\_\_\_\_\_\_\_Осипова А.Д. Заказчик\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_А.В. Замятин

\_\_\_\_\_\_\_\_\_\_\_\_\_\_Локотецкий М.Е.