**Техническое задание**

**на разработку программного сервиса**

**интеллектуального анализа данных**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1. **Название** | Решение задачи классификафии раковой опухоли на следующие группы: доброкачественная или злокачественная | | | | |
| 1. **Исполнение** | Индивидуальное | | | Групповое | |
| 1. **Разработчики** | Первушина Анна Алексеевна | | |  | |
| 1. **Цель разработки** | Создание модели машинного обучения для решения задачи классификации опухли на медицинских снимках. Необходимо предсказать, по набору признаков, является опухоль доброкачественной или злокачественной. | | | | |
| 1. **Решаемые задачи** | 1. Считка и нормализация данных;  2. Выбор алгоритмов машинного обучения, настройка параметров и гиперпараметров;  3. Оценка качества работы моделей. | | | | |
| 1. **Сроки исполнения** | 6.1 Начало: | 01.12.2020 | 6.2 Окончание: | | 25.01.2021 г. |
| 1. **Основные требования к разработке:** | | | | | |
| 1. **Данные** | **8.1 Источник данных** | | | https://www.kaggle.com/uciml/breast-cancer-wisconsin-data | |
| **8.2 Параметры обучающих данных** | | | Размер обучающей выборки составляет 70% от общего количества данных. Размер тестовой выборки составляет 30% от общего количества данных. Данные в обучающей и тестовой выборке должны быть репрезентативны | |
| **8.3 Формат признакового вектора x** | | | **Вектор признаков выглядит следующим образом:**  **X =** {*x1*,*x2*,…,*xn*},  где  n = 569  *x1*ϵ(6.981; 28.11) - radius\_mean  *x2*ϵ(9.71; 39.28) - texture\_mean  *x3*ϵ(43.79; 189) - perimeter\_mean  *x4*ϵ(144; 2500) - area\_mean  *x5*ϵ(0.05; 0.16) - smoothness\_mean  *x6*ϵ(0.02; 0.35) - compactness\_mean  *x7*ϵ(0.0;0.43 ) - concavity\_mean  *x8*ϵ(0.0; 0.2) - concave points\_mean  *x9*ϵ(0.11; 0.3) - symmetry\_mean  *x10*ϵ(0.05; 0.1) - fractal\_dimension\_mean  *x11*ϵ(0.11; 2.87) - radius\_se  *x12*ϵ(0.36; 4.88) - texture\_se  *x13*ϵ(0.76; 22) - perimeter\_se  *x14*ϵ(6.8 ;542) - area\_se  *x15*ϵ(0.01; 0.03) - smoothness\_se  *x16*ϵ(0.0; 0.14) - compactness\_se  *x17*ϵ(0.0; 0.4) - concavity\_se  *x18*ϵ(0; 0.5) - concave points\_se  *x19*ϵ(0.01; 0.08) - symmetry\_se  *x20*ϵ(0; 0.03) - fractal\_dimension\_se  *x21*ϵ(7.93; 36) - radius\_worst  *x22*ϵ(12; 49.5) - texture\_worst  *x23*ϵ(50.4; 251) - perimeter\_worst  *x24*ϵ(185; 4250) - area\_worst  *x25*ϵ(0.07; 0.22) - smoothness\_worst  *x26*ϵ(0.03; 1.06) - compactness\_worst  *x27*ϵ(0; 1.25) - concavity\_worst  *x28*ϵ(0; 0.29) - concave points\_worst  *x29*ϵ(0.16; 0.66) - symmetry\_worst  *x30*ϵ(0.06; 0.21) - fractal\_dimension\_worst | |
| **8.4 Формат вектора y (опционально)** | | | **Y =** {*y1*,*y2*,…,*ym*},  где  *y1*ϵ{B,M} *–* diagnosis  *m* = 569 | |
| 1. **Функционал** | **9.1 Функции предварительной обработки (с трансформацией/без трансформации признакового пространства)** | | | Выбор информативных признаков  Устранение пробелов в данных  Кодирование категориальных признаков | |
| **9.2 Функции содержательной обработки (метод DM, методы DM и принцип интегрирования результата, полученного отдельными методами)** | | | Предсказание переменной Y с помощью алгоритмов машинного обучения | |
| 1. **Результаты** | **10.1 Представление результатов/что позволят установить (формулировка полученных неочевидных знаний, способствующих принятию на практике конкретных управленческих решений)** | | | Предсказание переменной Y с помощью алгоритмов машинного обучения | |
| 1. **Исследования** | **11.1 Исследовательская компонента (оценка эффективности и пределов применимости используемых методов, рекомендации по использованию методов в решении аналогичных задач)** | | | Устранение пропусков/пробелов в данных  Отбор значимых признаков  Сравнение качества работы алгоритмов машинного обучения | |

Ответственный исполнитель\_\_\_\_\_\_\_\_\_\_\_\_\_Первушина А.А. Заказчик\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_А.В. Замятин