**Финальный экзамен по машинному обучению.**

1. Загрузите файл «ДатаСет\_Выдача», в Python и сделайте предобработку данных (preprocessing) – 20/100 баллов:

А) Очистка или заполнение пустых ячеек

Б) Удаление не нужных для модели переменных

В) Преобразование категориальные переменные в фиктивные переменные (переменные с значением 0 и 1) и т.д.

1. Проведите исследовательский анализ данных (Exploratory data analysis) 20/100 - баллов

А) найдите параметры описательной статистики по некоторым значимым переменам;

Б) нарисуйте графики (Box-plot, Scatter-plot и др.) по некоторым значимым переменам;

В) Нарисуйте корреляционную матрицу переменные с непрерывными данными;

1. Используйте все возможные входные параметры (независимые переменные) и зависимую переменную «Просрочки за цикл (кумулятивный)» (меньше 30 дней как 1 и больше, как 0) для решения задачи классификации. Примените такие алгоритмы машинного обучение, как: 60/100 баллов

А) логистическая регрессия;

Б) дерево решений;

В) случайны лес;

Г) К-ближайших соседей;

Д) Наивный Байес;

Е) алгоритмы бустинга