Общее описание задачи:

Необходимо построить предиктивную модель, которая на основании открытых данных о судах и бухгалтерской отчетности позволит спрогнозировать вероятность банкротства компании.

Необходимо сделать:

1. Собрать из предложенных файлов датасет.

2. Провести EDA, отобрать фичи и сгенерировать новые.

3. Обучить модель.

4. Собрать исполняемый код, который можно будет запустить на проде: на вход будет получать файлы аналогичного формата (по одному), возвращать должен прогнозное значение.

Будут оцениваться:

1. Каждый из пунктов задания. Если удалось выполнить не все, а только несколько первых пунктов — не беда.

2. Чистота написания кода. Также мы можем запустить код (и ноутбук для обучения модели, и код для продакшена) самостоятельно и проверить его работоспособность.

3. Наличие необходимых комментариев и пояснений в коде.

4. Разумность и оригинальность идей, используемых при построении модели, а также их обоснование.

5. Качество модели (или моделей), наличие рассчитанных необходимых метрик (список метрик необходимо определить самостоятельно).

6. Наличие и качество выводов, сделанных по результатам решения задачи.

Описание данных:

1. В архиве court\_cases\_sample лежат json-файлы судебных дел, поле case\_sides.type принимает значения {0:'истец', 1:'ответчик', 2:'свидетель'}.

2. В файле df\_bankruptcies\_sample приведен список компаний, претерпевших банкротство. Если компания в списке отсутствует, значит банкротства у нее не было. Факт банкротства — это целевая переменная в задаче.

3. В файле df\_accounts приведена информация об активах и задолженности компаний. Поле okei отвечает за единицу измерения {383:'рубль', 384:'тысяча рублей', 385:'миллион рублей'}.

Желаем успеха!