|  |  |  |
| --- | --- | --- |
| **Модуль** | **Входные данные** | **Выходные данные** |
| **0\_dummy\_generater** – генератор синтетического дата сета. Лучше не запускать, так как словарь для лематизации настроен на дата сеты, выложенные в папку | **HeadHunter\_train.csv** – реальные данные  **HeadHunter\_test.csv** – реальные данные | **HeadHunter\_train\_dummy.csv** – синтетические данные  **HeadHunter\_test\_ dummy.csv** – синтетические данные |
| **1\_preprocessing\_train\_table\_set\_MLP –** предобработка тренировочного дата сета, включая лемматизацию и присоединение числовых переменных к тексту. | **HeadHunter\_train\_dummy.csv** – синтетические данные для тренировки модели  **words\_relation\_for\_lema.csv –** словарь для лемматизации | **file\_df\_set\_train\_MLP\_new.csv –** файл с корпусомтекстов, готовый к векторизации и значениями зависимой переменной |
| **2\_preprocessing\_test\_table\_set\_MLP –** предобработка тренировочного дата сета, включая лемматизацию и присоединение числовых переменных к тексту. | **HeadHunter\_test\_dummy.csv** – синтетические данные для тестирования модели  **words\_relation\_for\_lema.csv –** словарь для лемматизации | **file\_df\_set\_test\_MLP\_new.csv –** файл с корпусомтекстов, готовый к векторизации |
| **3\_Keras\_MLP\_train** – векторизация текста и зависимой переменной, выбор параметров модели по валидационной и стратифицированной тестовой выборках, обучение модели | **file\_df\_set\_train\_MLP\_new.csv –** файл с корпусомтекстов, готовый к векторизации и значениями зависимой переменной | **TfidfVectorizerMLP.pkl –** TF-IDF векторайзер  **KerasModel** – обученная модель |
| **4\_KerasMLP\_test** – векторизация текста тестового дата сета, загрузка данных в модель, получение вектора предсказания зависимой переменной в виде вероятностей, перевод вектора вероятностей в мульти лейбел вид, требуемый для сабмита | **file\_df\_set\_test\_MLP\_new.csv -** файл с корпусомтекстов, готовый к векторизации  **KerasModel** – обученная модель | **submit\_MLP.csv –** файл с предсказанием в требуемом формате |