**Main**

* Modified MLProject file command line
* Modified conda.yml

**mlflow-92125fc05f97e6341aea640d17c242b12f744f49**

(with joblib for hyperparameter optimization)

**get\_data**

* Modified conda.yml

**mlflow-eb425dddc7dd5e5b83d410c9982978d1b2a7530f**

**EDA**

* modified conda.yml
* Modified MLProject file command line to jupyter lab

**mlflow-940c911614ecfe775b04fb8934c8ecf36fc0cb2f**

**basic\_cleaning**

* modified conda.yml
* completed MLProject file
* adjusted run.py

**mlflow-ba3fd2f6fed99a055bd964bb74140cef913ae07a**

mlflow run . -P hydra\_options="main.execute\_steps='basic\_cleaning'"

**data\_check**

* modified conda.yml
* adjusted run.py
* corrected MLProject file command adding -s

**mlflow-4fd29a272f8344adbcf2fa7e7740500cc306677a**

mlflow run . -P steps="data\_check"

**train\_test\_split**

* completed MLProject command line
* modified conda.yml
* added False to temp file in run.py

**mlflow-bcdcef54b6add192be4ac06d0883a50af96f4bc0**

mlflow run . -P steps="data\_split"

**train\_random\_forest**

* adjusted MLproject command line + changed rf\_config type from str to path
* modified conda.yml
* corrected run.py to add in predict

x\_val[processed\_features]

* Added tempfile in wandb.init() to save plots

**mlflow-57806004487849f8bc7e315c33fe00ccf8c1075d**

mlflow run . -P steps="train\_random\_forest"

**Hyperparameters tuning**

**Do**

In MLProject

1/ remove

main.steps={steps}

2/Replace command by

python main.py --multirun modeling.random\_forest.max\_features=0.1,0.33,0.5,0.75,1 modeling.max\_tfidf\_features=10,15,30 hydra/launcher=joblib

**In main.py file:**

**3/ force (uncomment)**

active\_steps = ["train\_random\_forest" ] #$$$$$$$ FOR HYPERPARAM OPTIMIZATION $$$$$$$$$$$$$$$

**then launch optimization with**

mlflow run .

**Regression Test**

* Changed conda.yml
* Added used\_columns in run.py – KEY STEP MISSING

**mlflow-f0411b5e8a6b3b156120da53093f28f806c2aab4**

mlflow run . -P steps=test\_regression\_model