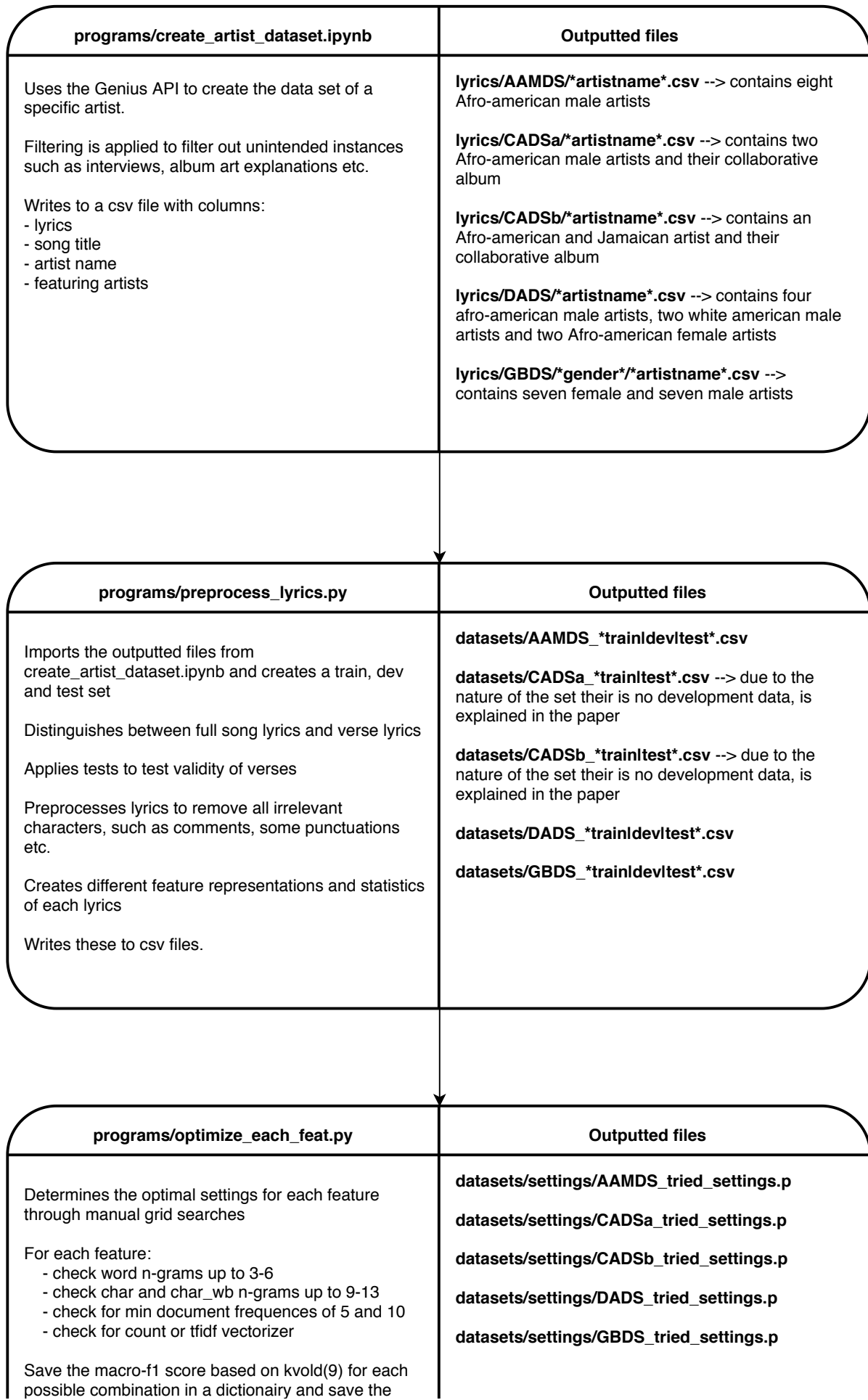


Lyrics-based Author Classification set up



highest word and char(_wb) settings per feature while running.

programs/create_best_feat_set.py

Determines the best combination of features by:

- Starting with the highest macro-f1 scoring feature
- Add the second best feature to the active feature set
- Classify with kfold = 9
- If the macro-f1 score improves add the feature to the feature set

Add the end the best feature set is saved to a file

Outputted files

datasets/settings/AAMDS_best_feat_set.p

datasets/settings/CADSa_best_feat_set.p

datasets/settings/CADSB_best_feat_set.p

datasets/settings/DADS_best_feat_set.p

datasets/settings/GBDS_best_feat_set.p

programs/classify.py

Loads the best settings for a dataset

Classifies using these settings on the dev, test or kfold on the training set and outputs the result

Outputted files