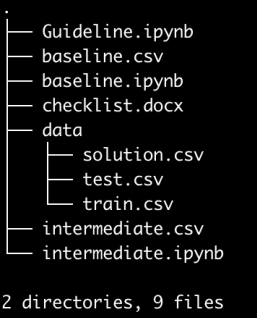


Checklist – topic 2

Checklists	Checkbox / Answer	Comments
Have all 9 files in correct hierarchy? 	yes	
Shape of the train.csv file (output of {dataframe}.shape)	(15000, 2)	
What's the file size of train.csv? (We should avoid big file like bigger than 500mb)	1.6MB	
Guideline.ipynb Do you have overall problem statement and descriptions?	yes	
Guideline.ipynb Do you have all sub-problems and proper problem statements? Sections should include (1)Data analysis, (2) feature engineering, (3) Model development, (4) internal evaluation, (5) external evaluation	yes	
Evaluation metrics Do you have appropriate evaluation metrics? For example, simple accuracy should be avoided for imbalance data. For imbalance data, better choice would be something like balanced accuracy, F1 score, and others	yes	I have used balanced accuracy
Solution csv files Intermediate.csv should perform better with some margin. Does it? Have you perform some sophisticated feature engineering/ model development for intermediate.csv?	yes	## Overall ## Baseline accuracy on test dataset is 80% ## Intermediate accuracy on test dataset is 85%
Solution ipynb files Have you included every requirements mentioned in the instruction? Please read them carefully and try including everything. Fixing this later would be so much harder.	Yes	
Notebook runtime	1hr 30min	

What's your notebook runtime? I think it would be nice to have it shorter than 3 hours.		
Scoring function Have you come up with the scoring function? What's your score for baseline and intermediate? Answer it in {baseline}/{intermediate} format.	## Overall ## Baseline accuracy on test dataset is 80% ## Intermediate accuracy on test dataset is 85%	
Topic-specific question Have you included topic modeling question in your guideline? And solution in your baseline.ipynb, intermediate.ipynb?	yes	