Step	Description	Deliverable	Effort Estimation	Due Date	Owner	Comment
EDA	“Playing” with the data and gaining some understandings	Visualizations (extract relevant directions for features) Feature engineer prioritization Undersand the characteristics of the data, biasing and outliers	2 days	week from data delivery	group	
Learning - pre process methods	Reading relevant articles and tools Brainstorming	Collect ideas and methods to apply	two days	week 2	divide responsibility	
Learning - classification methods	Reading relevant articles	Several Model proposals The chosen model should take into account the data characteristics	two days	week 2	divide responsibility	
Pre-process	Feature extraction Feature engineering	clean, tokeinized and averaged (tf-idf) data, ready for embedding Jupyter notebook which presents the work done	three days	week 3	Each of us independantly	
Training classifiers & Evaluation	Embeddings (fastText/word2vec/GloVe) Implementing classifiers Training Measuring performans (precision/recall for each class)	Trained classifiers: Logistic regression | Random forests | Naive Bayes | KNN | ensemble of those Performans test Jupyter notebook for evaluation stage	two days	week 4	Divide by model	
Active Learn: Framework	Build baseline for evaluation of the active learn Splite the dataset Evaluate performans of re-training with random sampling	Partition of the dataset Random sampling test, for random spliting of the dataset and random sampling from pool All in Jupyter notebook including graghs	two days	week 5	TBD	Splite the data to: train, test, and 2 or more cycles of of re-train (i.e. pools for active lean)
Active Learn: Query methods	Implement samples query methods: Query by uncertainity (using softmax) Query by commity	Running code... Evaluation using the framework we've build	two days	week 6	Divide by method	
Active Learn: Clustering the pool	Clustering the pool (in feature space) in order to divers the sampling of the above query methods.	Implementation Test for improve in performans	two days	week 7	Divide by method	
Optional - model saturation	TBD	TBD	TBD	week 8?	TBD	
Summarize	Write a notebook summarizing the project - the methods learned - the chosen method applied - the model + training - the performence analysis Upload the code to github?	- notebook - Github with the relevant code	2 days	week 9	divide responsibility	