Group 9 / Panic! At The Deadline (Kaggle group name)	Meeting locations: In-person (after Wednesday lecture) for 1 hour	
MEETING 1 (Week 6)	**MEETING 2 (Week 7)**	**MEETING 3 (Week 8)**
Higher-level coordination stuff:	Summarising what we've done so far	Summary of Week 2:
Communication channels - Whatsapp, Zoom chat, Discord.	Start writing parts 1 and 3 of report on Naive Bayes (simple approach).	Neural network:- training is slow despite similar performance to Logistic Regression
Google Colab for code sharing.	Group agreement submission	Imbalanced data classification on domain 2 (hybrid SMOTEENN - gridsearch CV for hyperparameter tuning)
Weekly meetings - Wednesdays 11am after lecture.		domain classification: also imbalanced. need to improve performance
		domain 1 cross validated with hypertuned parameters: ~0.8
Ideas:	**Goals for next week:**	**Goal for final week:**
Feature engineering: Bag of words, TF-IDF.	Changfa: Try stacking classifier (KNN, Decision - Tree, SVM (linear kernel), SVM (rbf kernel), Logistic Regression (linear model) -> Random Forest) on Domain 1 / 2.	Better domain classifier for domain 1 vs domain 2
Index of word in sentence as a feature (to get ordering).	Jed: Basic neural network designs.	Keep working on the model performance (fine tuning)
ML Models: Logistic Regression. SVM. RNN + CNN.	Joey: How to best tackle imbalanced data - coding different methods.	More feature engineering to try: word rarity, lexical diversity
Feature selection: Decision Trees / Random Forests + SVM	Everyone: write a little bit.	**More ideas to try**
Neural networks: dimensionality reduction (PCA?)		Autoencoder/decoder on machine text, then test for human
		Combine all individual model into one big model with one meta-classifier using the output probabilities of each model as features
	Use different evaluation methods, not accuracy for validation.	Unify code - meet in person and run the full model to submit on kaggle
Goals for the next week:		
Code the algorithms	**High-level goal**: Reach 80+ accuracy!	**High-level goal:** Reach above 80+ on kaggle
Jedwin: Feature engineering: bag-of-words, tf-idf, PCA, logistic regression.		
Changfa: Random-forests + SVM.	How to make sure algorithm is generalisable to unseen instances	**More ideas to try:** Autoencoder/decoder on machine text
Joey: SVM.	Keep note things didn't work.	
Other thoughts:		
- Alternative evaluation metric (later down the line) precision, F1, recall, AUC - ROC		