



Assignment 3: Covid-19 Fact Checking

Team Coders For Hire:

Ihsaan Malek, 40024975 & Olivier Racette 40017231

Datasets

- Very dirty/noisy
 - Emojis
 - Twitter Hyperlinks -> Unique
 - Different Languages
 - Grammatical
 - Slang
 - Plurals vs singular, ex: case vs cases
 - Numbers
 - How do you determine if a number is relevant or correct
 - Stopwords
- These all affect the probability

Regular vocab vs Filtered Vocab

- Sample Length:
 - Regular:
 - Yes: 3316
 - No: 1640
 - Filtered:
 - Yes: 898
 - No: 406
 - Takes out unique elements
- Leads to a worse model performance, impact of conditionals are higher

Summary of Model Performances

	<i>Accuracy</i>	F1-SCORE		<i>Recall</i>		<i>Precision</i>	
MODEL		YES	NO	YES	NO	YES	NO
NB - Regular	0.4	0.06	0.56	0.03	0.96	0.5	0.4
NB - Filtered	0.364	0.10	0.51	0.061	0.82	0.33	0.37
LSTM (averages)	0.73	0.81	0.55	0.9	0.45	0.73	0.77

Contributions

Description of
contributions and
responsibilities

Main.py

Olivier, Ihsaan

model_eval.py

Olivier, Ihsaan

PPT Slides

Ihsaan, Olivier