Disaster Tweet Detection

Phase II Demo

Progress

Feb

Mar

Apr
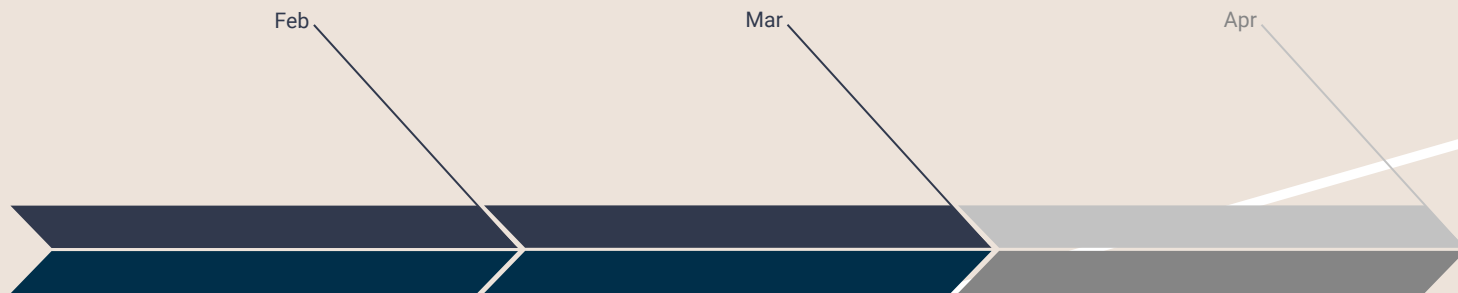
Baseline

Working "simple" model

POC Twitter scraping setup

Improved

NLP / Attention

DistilBERT → Simple Transformer RoBERTa

Final

Interactive dashboard

Clustering mechanism

Performance Evaluation (Phase 1)

	Classifier	AUC	Accuracy	F1 Score
0	Logistic Regression	0.777215	0.782563	0.746634
3	SVM	0.779095	0.796218	0.734247
5	Perceptron	0.754251	0.757878	0.723123
1	Random Forest	0.756242	0.778361	0.696839
2	AdaBoost	0.733616	0.745798	0.686528
4	Gradient Boosting Classifier	0.720549	0.750525	0.630350

Performance Evaluation (Phase 2)

Name	Submitted	Wait time	Execution time	Score
submission.csv	7 days ago	6 seconds	6 seconds	0.81857

Complete

Jump to your position on the leaderboard ▾

Data Preprocessing

- Stop words
- Emojis
- Punctuation

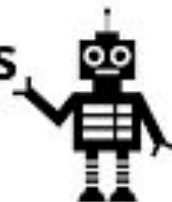|			
ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/lHYXEOHY6C	1
ablaze	Est. September 2012 - Bristol	We always try to bring the heavy. #metal #RT http://t.co/YAo1e0xngw	0
ablaze	AFRICA	#AFRICANBAZE: Breaking news:Nigeria flag set ablaze in Aba. http://t.co/2nndBGwyEi	1
ablaze	Philadelphia, PA	Crying out for more! Set me ablaze	0
ablaze	London, UK	On plus side LOOK AT THE SKY LAST NIGHT IT WAS ABLAZE http://t.co/qqsmshaJ3N	0

Data Augmentation

Updated Model

- XLNet
- RoBERTa
- DistilBERT

Simple Transformers

Hyperparameter Tuning

- Weights and Biases (wandb.ai)
  - https://wandb.ai/brendanmanning/DisasterTweets/sweeps/4xdpwkgc/table?workspace=user-brendanmanning

```python
sweep_config = {
    "method": "bayes",  # grid, random
    "metric": {"name": "train_loss", "goal": "minimize"},
    "parameters": {
        "num_train_epochs": {"values": [1,2,3,4,5]},
        "learning_rate": {"min": 5e-5, "max": 4e-4},
    },
}

sweep_id = wandb.sweep(sweep_config, project="DisasterTweets")
```
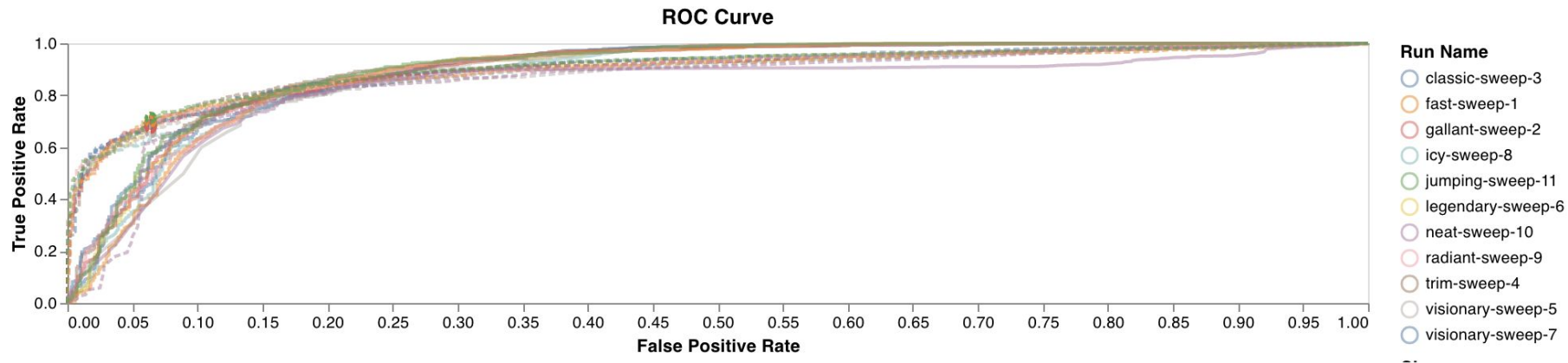
Tuning: Training Loss

Sweep: 4xdpwkgc

	Name (11 visualized)	ID	Created	Runtime	learning_ra	num_train_epochs	Training ▲
·	visionary-sweep-7	ughraasb	10m ago	2m 18s	0.0003118	5	0.1033
·	icy-sweep-8	v0ueapcg	7m ago	2m 17s	0.0001392	4	0.1053
·	gallant-sweep-2	4a9kc8g4	21m ago	2m 12s	0.0001195	5	0.1109
·	visionary-sweep-5	cufz2ou6	14m ago	2m 12s	0.0003242	5 ☰	0.2261
·	fast-sweep-1	2p8k70m	24m ago	2m 36s	0.0002581	5	0.3595
·	classic-sweep-3	x6i610bg	19m ago	2m 13s	0.000381	1	0.376
·	radiant-sweep-9	354k1ggj	5m ago	2m 20s	0.0001488	4	0.4597
·	jumping-sweep-11 ●	jlm8fh33	33s ago	17s	0.0003086	4	0.4683

Tuning: ROC Curve

Tuning: Training Loss

Next Steps

- Predict on scraped Twitter data
- Interactive dashboard