

Rapport d'Expérimentation : Fine-tuning FinBERT sur Tweets Bitcoin

1. Méthodologie et Configuration

Nous avons entraîné un modèle de classification de sentiment (Positif/Négatif) sur des données de réseaux sociaux.

- **Modèle** : ProsusAI/finbert (BERT pré-entraîné sur des textes financiers formels).
- **Données** : Dataset MBSA (Bitcoin Tweets), échantillon de 50 000 tweets.
- **Nettoyage** : "Strict" (Suppression des chiffres, emojis, URLs, mentions, ponctuation).
- **Matériel** : GPU Tesla T4 (Entraînement rapide, ~13 minutes par époque).
- **Architecture** : Entraînement sur 3 époques.

2. Analyse de la Dynamique d'Entraînement

```
Starting training...
--- Epoch 1/3 ---
Training: 100% [██████████] 1250/1250 [13:31<00:00, 1.52it/s, loss=0.608]
Average training loss: 0.6508
Evaluating: 100% [██████████] 313/313 [01:11<00:00, 4.83it/s]
Validation Accuracy: 0.5584 | F1-weighted: 0.5003
--- Epoch 2/3 ---
Training: 100% [██████████] 1250/1250 [13:42<00:00, 1.52it/s, loss=0.595]
Average training loss: 0.6265
Evaluating: 100% [██████████] 313/313 [01:11<00:00, 5.06it/s]
Validation Accuracy: 0.5710 | F1-weighted: 0.5709
--- Epoch 3/3 ---
Training: 100% [██████████] 1250/1250 [13:42<00:00, 1.52it/s, loss=0.557]
Average training loss: 0.5863
Evaluating: 100% [██████████] 313/313 [01:10<00:00, 5.08it/s]
Validation Accuracy: 0.5641 | F1-weighted: 0.5636
Final classification report:
```

Les logs montrent l'évolution du modèle sur 3 cycles (époques) :

- **La Perte (Training Loss) diminue bien :**
 - Époque 1 : 0.6508
 - Époque 2 : 0.6265
 - Époque 3 : 0.5863
 - *Interprétation* : Le modèle "apprend" et arrive à mieux comprendre les données d'entraînement (le *Train set*).
- **La Validation (Accuracy) stagne :**
 - Époque 1 : 55.84%
 - Époque 2 : 57.10% (Pic de performance)

- Époque 3 : 56.41% (Légère baisse)
- *Interprétation* : Bien que le modèle apprenne par cœur les données d'entraînement, il n'arrive pas à généraliser sur des données qu'il n'a jamais vues. Le fait que la précision baisse à l'époque 3 est un signe de début de **surapprentissage (overfitting)**.

3. Analyse des Performances Finales

```
Final classification report:
      precision    recall    f1-score   support
Negative        0.58      0.60      0.59     5279
Positive        0.54      0.52      0.53     4721

accuracy                           0.56     10000
macro avg       0.56      0.56      0.56     10000
weighted avg    0.56      0.56      0.56     10000

Training finished!
```

Classe	Précision (Precision)	Rappel (Recall)	F1-Score	Analyse
Negative (0)	0.58	0.60	0.59	Le modèle détecte un peu mieux les tweets négatifs (60% des vrais négatifs sont trouvés).
Positive (1)	0.54	0.52	0.53	Performance très faible. À peine mieux que le hasard (50%).

Conclusion des chiffres : Un score de 56% sur un problème binaire (pile ou face) est considéré comme **médiocre**. Le modèle n'a pas réussi à capturer la logique du sentiment dans ces tweets.