

Rapport d'Analyse des Performances du Modèle de Sentiment

1. Vue d'ensemble

L'entraînement a été réalisé sur un unique jeu de données pour les deux modèles (positif et négatif) qui fonctionnent en miroir :

- Jeu de données total : 1,229,617 tweets
- Jeu d'entraînement : 1,045,174 échantillons (85%)
- Jeu de test : 184,443 échantillons (15%)

Les deux modèles sont entraînés sur exactement les mêmes données, mais avec des étiquettes différentes :

- Le modèle positif prédit si un tweet est positif (1) ou non-positif (0)
- Le modèle négatif prédit si un tweet est négatif (1) ou non-négatif (0)

Cette architecture en miroir permet une analyse bidirectionnelle du sentiment, où chaque tweet est évalué indépendamment pour sa positivité et sa négativité.

2. Matrices de Confusion et Interprétation

Modèle Positif

```
[[72229 23333]
 [16709 72172]]
```

Interprétation détaillée :

- **Vrais Négatifs (VN) : 72,229**
 - Ces tweets ont été correctement identifiés comme non-positifs
 - Représente 39.16% du jeu de test
 - Montre une bonne capacité à identifier les tweets neutres ou négatifs
- **Faux Positifs (FP) : 23,333**

- o Tweets incorrectement classés comme positifs alors qu'ils ne le sont pas
 - o Représente 12.65% du jeu de test
 - o Indique une tendance du modèle à "sur-classifier" les tweets comme positifs
- **Faux Négatifs (FN) : 16,709**
 - o Tweets positifs manqués par le modèle
 - o Représente 9.06% du jeu de test
 - o Montre une certaine prudence du modèle dans la classification positive
- **Vrais Positifs (VP) : 72,172**
 - o Tweets correctement identifiés comme positifs
 - o Représente 39.13% du jeu de test
 - o Démonstre une bonne détection des sentiments positifs

Indicateurs dérivés :

- **Taux de faux positifs** ($FPR = FP/(FP+VN)$) : 24.41%
 - o Un tweet sur quatre non-positif est mal classé comme positif
- **Taux de faux négatifs** ($FNR = FN/(FN+VP)$) : 18.80%
 - o Moins d'un tweet positif sur cinq est manqué
- **Précision** ($VP/(VP+FP)$) : 75.57%
 - o Quand le modèle prédit "positif", il a raison dans environ 3 cas sur 4
- **Sensibilité/Rappel** ($VP/(VP+FN)$) : 81.20%
 - o Le modèle capture plus de 4 tweets positifs sur 5

Modèle Négatif

```
[ [72172 16709]
  [23333 72229] ]
```

Interprétation détaillée :

- **Vrais Négatifs (VN) : 72,172**
 - o Tweets correctement identifiés comme non-négatifs
 - o Représente 39.13% du jeu de test
 - o Performance miroir du modèle positif
- **Faux Positifs (FP) : 16,709**
 - o Tweets incorrectement classés comme négatifs
 - o Représente 9.06% du jeu de test
 - o Montre une tendance plus prudente dans la classification négative
- **Faux Négatifs (FN) : 23,333**

- o Tweets négatifs manqués par le modèle
 - o Représente 12.65% du jeu de test
 - o Reflète une certaine réticence à classer comme négatif
- **Vrais Positifs (VP) : 72,229**
 - o Tweets correctement identifiés comme négatifs
 - o Représente 39.16% du jeu de test
 - o Performance très similaire à la détection positive

Indicateurs dérivés :

- **Taux de faux positifs** ($FPR = FP/(FP+VN)$) : 18.80%
 - o Moins de tweets sont faussement classés négatifs comparé aux faux positifs
- **Taux de faux négatifs** ($FNR = FN/(FN+VP)$) : 24.41%
 - o Le modèle manque plus de tweets négatifs que le modèle positif ne manque de positifs
- **Précision** ($VP/(VP+FP)$) : 81.20%
 - o Meilleure précision pour les prédictions négatives
- **Sensibilité/Rappel** ($VP/(VP+FN)$) : 75.57%
 - o Rappel plus faible pour les sentiments négatifs

Analyse Croisée des Matrices

1. Symétrie des Performances :

- o La distribution des erreurs est parfaitement symétrique entre les deux modèles
- o Cette symétrie confirme la robustesse de l'approche en miroir
- o Les biais de chaque modèle se compensent mutuellement

2. Biais de Classification :

- o Le modèle positif est plus "optimiste" avec plus de faux positifs
- o Le modèle négatif est plus "conservateur" avec moins de faux positifs
- o Cette complémentarité permet une analyse plus nuancée des sentiments

3. Impact sur l'Application :

- o La combinaison des deux scores permet de détecter des nuances
- o Un tweet avec des scores élevés dans les deux modèles indique un contenu ambigu
- o Un tweet avec des scores faibles dans les deux modèles suggère un contenu neutre

3. Analyse Détaillée des Performances

Les performances miroir des modèles s'expliquent par leur architecture symétrique :

Architecture en Miroir

- Les deux modèles utilisent le même vectorizer avec 2500 features
- Ils partagent la même régularisation ($C=1.5$) et les mêmes hyperparamètres
- La seule différence est l'inversion des labels pour l'entraînement

Modèle Positif

- **Précision :**
 - Classe 0 (Non-positif) : 0.81
 - Classe 1 (Positif) : 0.76
- **Rappel :**
 - Classe 0 (Non-positif) : 0.76
 - Classe 1 (Positif) : 0.81
- **F1-Score :**
 - Classe 0 (Non-positif) : 0.78
 - Classe 1 (Positif) : 0.78
- **Accuracy globale :** 0.78 (78.29%)

Modèle Négatif

- **Précision :**
 - Classe 0 (Non-négatif) : 0.76
 - Classe 1 (Négatif) : 0.81
- **Rappel :**
 - Classe 0 (Non-négatif) : 0.81
 - Classe 1 (Négatif) : 0.76
- **F1-Score :**
 - Classe 0 (Non-négatif) : 0.78
 - Classe 1 (Négatif) : 0.78
- **Accuracy globale :** 0.78 (78.29%)

4. Observations et Analyse des Erreurs

1. Performance Miroir :

- L'accuracy identique (78.29%) pour les deux modèles n'est pas une coïncidence mais le résultat de leur architecture en miroir
- Les erreurs sont symétriques entre les deux modèles, ce qui confirme leur complémentarité

2. Erreurs Fréquentes :

- Les erreurs sont parfaitement inversées entre les deux modèles du fait de leur architecture miroir
- Quand un modèle fait une erreur de classification positive, l'autre fait généralement l'erreur inverse pour la classification négative

3. Biais Observés :

- La symétrie des biais entre les modèles est intentionnelle et permet une analyse plus nuancée des sentiments
- Cette approche permet de capturer des nuances que un seul modèle ne pourrait pas détecter

5. Recommandations d'Amélioration

1. Optimisation du Prétraitement :

- Enrichir la liste des stopwords avec des termes spécifiques aux sentiments
- Implémenter une meilleure gestion des emojis et des expressions idiomatiques

2. Ajustements du Modèle :

- Expérimenter avec différentes valeurs de régularisation (actuellement $C=1.5$) tout en maintenant la symétrie des modèles
- Envisager des architectures plus complexes comme BERT ou XGBoost tout en conservant l'approche en miroir

3. Enrichissement des Données :

- Augmenter la diversité du jeu de données d'entraînement en gardant un équilibre entre sentiments positifs et négatifs
- Ajouter des exemples de cas limites en s'assurant qu'ils bénéficient aux deux modèles

4. Optimisations Techniques :

- Augmenter MODEL_FEATURES (actuellement 2500) pour les deux modèles en parallèle
- Implémenter une validation croisée synchronisée pour les deux modèles

5. Affinement de l'Architecture Miroir :

- Explorer la possibilité d'ajouter une couche de validation croisée entre les deux modèles
- Introduire un mécanisme de calibration pour optimiser la complémentarité des prédictions
- Considérer l'ajout d'un troisième modèle pour la détection de neutralité tout en maintenant la cohérence de l'ensemble