**Création du Fichier Final d’Annotation des Émotions pour un Corpus de 1140 Tweets**

**Contexte**

Dans le cadre de notre TER réalisé en groupe, nous avons annoté les émotions exprimées dans un corpus de 1140 tweets, initialement annotés par trois évaluateurs (Alain, Anton, Tony). Chaque évaluateur a attribué jusqu’à deux émotions par tweet (émotion 1 et émotion 2) parmi sept catégories possibles : 1=peur, 2=colère, 3=joie, 4=surprise, 5=tristesse, 6=dégoût, 7=neutre. Notre objectif était de produire un fichier final d’annotation avec une émotion définitive par tweet, en maximisant l’accord entre les évaluateurs tout en résolvant les cas ambigus de manière rigoureuse, afin de fournir une base solide pour l’entraînement de modèles d’apprentissage automatique.

**Évolution de l’Approche d’Annotation**

Au début du processus d’annotation, nous avions envisagé de conserver au moins une émotion par tweet, avec la possibilité d’en inclure plusieurs (deux ou plus) si les évaluateurs identifiaient des émotions multiples. Cependant, après avoir analysé les premières annotations et mesuré l’accord inter-évaluateurs (Kappa de Fleiss initial de 0.3083), nous avons décidé de modifier notre approche. Nous avons opté pour une seule émotion principale par tweet, en utilisant l’émotion 2 comme un outil pour résoudre les désaccords sur l’émotion 1. Cette décision a été motivée par deux raisons principales : (1) simplifier les annotations pour faciliter l’entraînement des modèles d’apprentissage automatique, qui obtiennent généralement de meilleurs résultats avec des étiquettes uniques et cohérentes, et (2) réduire l’ambiguïté dans les cas où les évaluateurs identifiaient des émotions multiples mais divergentes. Bien que l’émotion 2 n’ait pas été conservée dans le fichier final, elle a joué un rôle crucial pour atteindre un consensus sur l’émotion principale, rendant son annotation utile dans le processus.

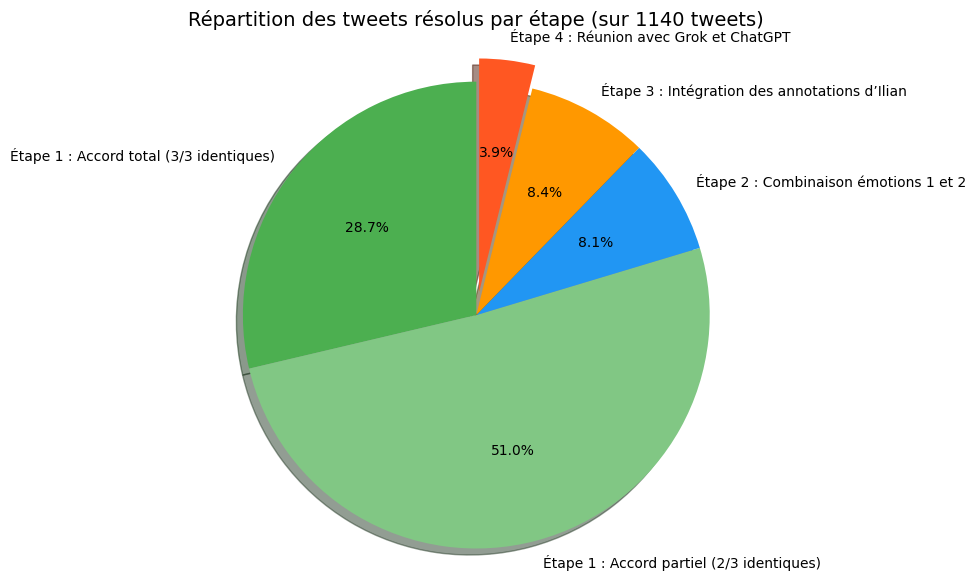
**Méthodologie pour la Création du Fichier Final**

Nous avons suivi une approche en plusieurs étapes pour déterminer une émotion principale définitive pour chacun des 1140 tweets, en combinant des méthodes automatiques et manuelles pour garantir la fiabilité des résultats. Voici les étapes détaillées :

1. **Étape 1 : Identification des tweets avec un accord majoritaire (au moins 2 évaluateurs sur 3 identiques)**
   * Nous avons d’abord examiné les émotions 1 des trois évaluateurs (Alain, Anton, Tony) pour chaque tweet.
   * Si les trois évaluateurs s’accordaient sur la même émotion 1 (par exemple, [3, 3, 3]), nous avons retenu cette émotion comme annotation définitive. Cette condition stricte (accord total) a été satisfaite pour 327 tweets (28,7 % des tweets).
   * Nous avons ensuite élargi notre critère : si au moins deux évaluateurs s’accordaient sur la même émotion 1 (par exemple, [3, 3, 5]), nous avons également retenu cette émotion comme annotation définitive.
   * Résultat : Sur les 1140 lignes, 908 ont atteint un accord majoritaire (au moins 2 évaluateurs sur 3 identiques), dont 327 avec un accord total (3 évaluateurs identiques). Cela représente 79,6 % des tweets. Il restait donc 232 lignes sans accord majoritaire (1140 - 908).
2. **Étape 2 : Combinaison des émotions 1 et 2 pour trouver un consensus**
   * Pour les 232 tweets restants, nous avons combiné les émotions 1 et 2 de chaque évaluateur, soit jusqu’à 6 votes par tweet (par exemple, [3, 5, 2, 2, 7, vide] pour un tweet).
   * Nous avons ensuite identifié l’émotion la plus fréquente parmi ces votes combinés pour déterminer une émotion principale. Par exemple, si les votes étaient [3, 5, 2, 2, 7, vide], l’émotion 2 (colère) apparaissait deux fois et était retenue comme annotation définitive.
   * Résultat : Cette étape a permis de résoudre 92 tweets supplémentaires (232 - 92 = 140 lignes restantes), représentant 39,7 % des tweets sans accord initial.
3. **Étape 3 : Intégration des annotations d’Ilian**
   * Initialement, un quatrième évaluateur, Ilian, avait participé à l’annotation, mais ses annotations avaient été retirées après son départ du groupe. Pour les 140 tweets restants, nous avons décidé de réintégrer sa colonne d’annotation afin d’augmenter les chances de trouver un consensus.
   * Nous avons repris le processus d’accord majoritaire en incluant les annotations d’Ilian (émotion 1 uniquement), soit 4 évaluateurs au total (Alain, Anton, Tony, Ilian). Si au moins deux évaluateurs s’accordaient sur une émotion, celle-ci était retenue comme annotation définitive.
   * Résultat : Cette étape a permis de résoudre 96 tweets supplémentaires (140 - 96 = 44 lignes restantes), soit 68,6 % des tweets restants à cette étape.
4. **Étape 4 : Réunion finale avec aide de Grok et ChatGPT**
   * Pour les 44 tweets restants, aucun accord majoritaire n’a pu être trouvé, même après l’intégration des annotations d’Ilian (par exemple, votes [3, 5, 7, 2]).
   * Nous avons organisé une réunion entre les évaluateurs pour examiner chaque tweet individuellement. Pour nous aider dans cette tâche, nous avons utilisé deux modèles d’intelligence artificielle, Grok et ChatGPT, afin d’obtenir des suggestions d’émotions basées sur le contenu des tweets.
   * Grok et ChatGPT ont fourni une analyse émotionnelle indépendante pour chacun des 44 tweets restants, que nous avons utilisée comme point de départ pour nos discussions. Sur ces 44 tweets, Grok et ChatGPT étaient d’accord sur l’émotion principale pour 42 d’entre eux, ce qui a facilité nos délibérations. Pour les tweets où ils divergeaient, nous avons relu les tweets et discuté pour déterminer l’émotion principale la plus appropriée.
   * Après discussion, nous nous sommes mis d’accord sur une émotion principale définitive pour chacun des 44 tweets restants.
   * Résultat : Cette étape a permis de finaliser les annotations des 44 dernières lignes, complétant ainsi le fichier final pour les 1140 tweets.

**Résultats**

* Fichier final : À l’issue de ce processus, nous avons obtenu un fichier final d’annotation avec une seule émotion principale définitive pour chacun des 1140 tweets.
* Répartition des résolutions :
  + 908 tweets (79,6 %) ont été résolus par un accord majoritaire sur l’émotion 1 (au moins 2 évaluateurs sur 3 identiques), dont 327 avec un accord total (3 évaluateurs identiques, 28,7 %).
  + 92 tweets (8,1 %) ont été résolus en combinant les émotions 1 et 2.
  + 96 tweets (8,4 %) ont été résolus en réintégrant les annotations d’Ilian.
  + 44 tweets (3,9 %) ont été résolus lors de la réunion finale avec l’aide de Grok et ChatGPT.
* Validation croisée : L’utilisation de Grok et ChatGPT lors de la dernière étape a permis de confronter nos interprétations à des analyses externes, renforçant la fiabilité des annotations finales pour les cas les plus ambigus.



**Discussion**

* Impact du passage à une seule émotion principale : Le choix de passer d’une annotation multi-émotion (au moins une émotion, voire plus) à une seule émotion principale a simplifié le processus de mise en commun et a probablement amélioré la qualité du fichier final pour l’entraînement des modèles d’apprentissage automatique. Les modèles fonctionnent généralement mieux avec des étiquettes uniques et cohérentes, car les annotations multi-émotions peuvent introduire du bruit et des ambiguïtés lors de l’entraînement. En utilisant l’émotion 2 pour résoudre les désaccords sur l’émotion 1, nous avons maximisé l’accord tout en garantissant une annotation claire et univoque.
* Efficacité du processus : La première étape (accord majoritaire sur l’émotion 1) a permis de résoudre la majorité des cas (79,6 %), avec 327 tweets ayant un accord total (3 évaluateurs identiques), ce qui reflète une certaine cohérence entre les évaluateurs pour près d’un tiers du corpus. Cela est cohérent avec le Kappa de Fleiss initial de 0.3083 (accord léger). L’approche optimisée combinant émotions 1 et 2 (Kappa de 0.3758) a également contribué à réduire le nombre de cas ambigus.
* Apport des annotations d’Ilian : La réintégration des annotations d’Ilian a été une décision clé, car elle a permis de résoudre 68,6 % des cas restants à l’étape 3, démontrant que l’ajout d’un évaluateur supplémentaire peut améliorer l’accord.
* Utilisation de Grok et ChatGPT : L’aide des modèles d’IA a été particulièrement utile pour les cas ambigus, car elle nous a fourni des perspectives objectives. Cependant, la décision finale reposait sur notre discussion collective, garantissant que les annotations reflètent notre compréhension collective du corpus.
* Limites : Certains tweets restaient très ambigus en raison de leur contenu (par exemple, émotions mixtes ou contexte culturel mal compris), ce qui a nécessité une discussion approfondie.

**Conclusion**

La création du fichier final d’annotation pour les 1140 tweets a été réalisée en quatre étapes successives, combinant des approches automatiques (accord majoritaire, combinaison des émotions 1 et 2, réintégration des annotations d’Ilian) et manuelles (réunion finale avec aide de Grok et ChatGPT). Le passage à une seule émotion principale par tweet, en utilisant l’émotion 2 comme outil pour résoudre les désaccords, nous a permis de produire un fichier d’annotation clair et cohérent, mieux adapté à l’entraînement de modèles d’apprentissage automatique. Ce fichier servira de base pour les analyses ultérieures dans le cadre de notre projet, notamment pour étudier la répartition des émotions dans le corpus et explorer les facteurs influençant les divergences entre évaluateurs.