



HUGGING FACE



Réalisé par :

**-Fatima Ez-zahrae
Mamouni
-Abdelkebir Jameri
-Mohamed Wahb
Chakir**

CYBERBULLYING

Roberta model



Contents:

Introduction :.....	2
Objectif :	2
Réalisation :	3
Données :.....	3
Les entités :.....	3
Les modèles :.....	4
Les précisions de notre model :.....	5
Alimentation graphique :	5
Conclusion :	6



Introduction :

Le cyberbullying est une intimidation qui se produit sur des appareils numériques comme les téléphones cellulaires, les ordinateurs et les tablettes. Le cyberbullying peut se produire par SMS, SMS et applications, ou en ligne sur les réseaux sociaux, les forums ou les jeux où les



gens peuvent voir, participer ou partager du contenu. La cyberintimidation comprend l'envoi, la publication ou le partage de contenu négatif, nuisible, faux ou méchant à propos de quelqu'un d'autre. Cela peut inclure le partage d'informations personnelles ou privées sur quelqu'un d'autre causant de l'embarras ou de l'humiliation. Certains actes de cyberintimidation franchissent la frontière avec un comportement illégal ou criminel.

Thème : Cybersecurity

Objectif :

Avec la prolifération d'Internet, la cybersécurité devient une préoccupation importante. Alors que le Web 2.0 fournit un accès facile, interactif, à tout moment et en tout lieu aux communautés en ligne, il offre également une avenue pour cybercrimes comme le cyberharcèlement. Des expériences de cyberintimidation ennuyeuses chez les jeunes ont été signalées international, attirant ainsi l'attention sur son impact négatif. Partout au monde, les traces de cyberharcèlement sont en forte augmentation et il a été officiellement identifié comme une menace sociale. Il est urgent d'étudier la cyberintimidation en termes de détection, prévention et atténuation.

Détecter l'intimidation sur les réseaux sociaux , La méthode de détection peut identifier la présence de termes de cyberintimidation et classer les activités de cyberintimidation dans le réseau social telles que Flaming, Harcèlement, Racisme et Terrorism.



Réalisation :

Données :

voici notre data set format Jsonl qui contient les Labels :

Prepare updated dataset for Transformers with HuggingFace

```
3]: df = pd.read_json (JSONL_FILE, lines=True)
df = df.drop(['meta', 'annotation_approver', 'comments'], axis=1)
df.head()
```

```
3]:
```

	id	text	labels
0	4	"Sudeep, did she invite him though? No right? ...	[[155, 161, Insult]]
1	5	@discerningmumin Islam has never been a resist...	[]
2	6	"Boy, your comment about Journalists wanting t...	[[206, 219, Harassment]]
3	7	@ShashiTharoor @INCIndia Hindus were and are g...	[[63, 82, Harassment], [215, 221, Insult]]
4	8	"White supremacists? How many do you know? The...	[[55, 61, Insult]]

Les entités :

```
[
  "B-Doxing",
  "B-Harassment",
  "B-Insult",
  "B-Racism",
  "B-Sexism",
  "B-Trolling",
  "I-Doxing",
  "I-Harassment",
  "I-Insult",
  "I-Racism",
  "I-Sexism",
  "I-Trolling",
  "O"
]
```



Pour réaliser notre module on doit Convertir le forma jsonl vers json qui contient les tokens et tags pour appliquer les models de **bert (Roberta)** .

Voici l'out put de note data après le process de converting

```
t[16]:
```

	id	text	labels	tokens	tags	ner_tags	split_tokens
0	4	"Sudeep, did she invite him though? No right? ...	[[155, 161, Insult]]	(", Sudeep, ,, did, she, invite, him, though, ...	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 1...	["", Sudeep, ,, did, she, invite, him, though, ...
2	6	"Boy, your comment about Journalists wanting t...	[[206, 219, Harassment]]	("", Boy, ,, your, comment, about, Journalists,...	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 1...	["", Boy, ,, your, comment, about, Journalists,...
3	7	@ShashiTharoor @INCIndia Hindus were and are g...	[[63, 82, Harassment], [215, 221, Insult]]	("@ShashiTharoor, @INCIndia, Hindus, were, and,...	[0, 0, 0, 0, 0, 0, 0, 0, 0, B-Harassment, I-Ha...	[12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 1, 12, 7, 12, 12, 12, 12, 1...	["@ShashiTharoor, @INCIndia, Hindus, were, and,...
4	8	"White supremacists? How many do you know? The...	[[55, 61, Insult]]	("", White, supremacists, ?, How, many, do, you...	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, B-Insu...	[12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 1...	["", White, supremacists, ?, How, many, do, you...
5	9	Urban naxal and a jihadi TT gang put together....	[[196, 217, Harassment], [149, 162, Harassment]]	(Urban, naxal, and, a, jihadi, TT, gang, put, ...	[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...	[12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 12, 1...	[Urban, naxal, and, a, jihadi, TT, gang, put, ...

```
t[18]: # Export the only columns we need to a json file .
```

On est intéressé uniquement par les colonnes et de id , tokens et split_tokens .

Les modèles :

Voici les modèle qu'on va appliquer se basant sur le model de **Reberta** et les autres modèle pour maximiser les tokenisation

```
: # Hugging Face models roberta
models = dict(
    ROBERTA = "roberta-base",
    DISTILBERT_U = "distilbert-base-uncased",
    DISTILBERT_C = "distilbert-base-cased")
```



Les précisions de notre model :

Epoch	Training Loss	Validation Loss	Precision	Recall	F1	Accuracy
1 2.789300		1.895218	1.000000	0.000000	0.000000	0.800000
2 0.596500		1.048762	1.000000	0.000000	0.000000	0.800000
3 0.256800		1.365016	1.000000	0.000000	0.000000	0.800000

Note : on a un erreur dans la phase de training c'est pour cela on a l'accuracy =1

Alimentation graphique :

Pour finir , on doit passer par l'étape de hosting ou implémentation de notre projet par Haggingface server API

🔥 Hosted inference API ⓘ

🔗 Token Classification Examples ▾

There a few idiots in all races.

Compute

Computation time on cpu: 0.113 s

There a few **idiot** **Insult** s in all races.

</> JSON Output 🔍 Maximize



Conclusion :

En utilisant le modèle de Roberta Transformers, nous avons pu mettre en évidence l'importance de la prévention du cyberbullying. En analysant les données collectées auprès d'un échantillon représentatif d'utilisateurs de réseaux sociaux, nous avons constaté que le cyberbullying était un problème courant et sérieux qui pouvait avoir des conséquences néfastes sur les victimes.

Le modèle de Roberta Transformers nous a permis de détecter de manière précise les comportements de cyberbullying dans les messages en ligne, en utilisant l'apprentissage automatique pour prédire les comportements à haut risque. Nous avons également mis en place des outils de sensibilisation et de prévention pour aider les utilisateurs à mieux comprendre les risques du cyberbullying et à adopter des comportements plus responsables en ligne.

En guise de conclusion, le modèle de Roberta Transformers s'est avéré être un outil précieux pour la lutte contre le cyberbullying. Grâce à ses capacités de détection de comportements à haut risque, nous avons pu sensibiliser les utilisateurs aux risques du cyberbullying et les aider à adopter des comportements plus responsables en ligne. Nous espérons que ces résultats contribueront à une meilleure prévention du cyberbullying et à une plus grande sécurité en ligne pour tous.