

Detecció de depressió a xarxes socials mitjançant varis mètodes de Machine Learning

Martí Caixal i Joaniquet

1 INTRODUCCIÓ

Les xarxes socials són considerades com uns sistemes d'informació en línia que permeten compartir l'estil de vida dels seus usuaris. Cada un té el seu perfil personal on penja actualitzacions del seu dia a dia i la resta d'usuaris poden reaccionar-hi i posar comentaris. De la mateixa manera, també es permet seguir a gent i altres institucions que siguin de l'agrat d'un [1].

L'inici de les xarxes socials es remunta a finals dels anys 90, quan l'internet tot just passava de ser una eina orientada a professionals a ser d'àmbit general. Tot i no ser la primera en aparèixer, la xarxa "MySpace" va ser la que va popularitzar aquest fenomen i va obrir pas a un seguit de noves xarxes socials. La més famosa, i que actualment segueix sent la que té més usuaris actius, és "Facebook", creada per en Mark Zuckerberg. Si bé inicialment l'únic objectiu era estar en contacte amb la gent del teu cercle més proper, avui en dia les xarxes socials són un mitjà per la gent famosa on rebre milers i milers de seguidors i visualitzacions, deixant de banda l'objectiu principal amb el que es van crear. [2]

Les xarxes socials no estan exemptes de problemàtiques. Al cap i a la fi, són un lloc on tothom pot dir la seva sense cap tipus de restricció. Això ha portat fins a un punt on la gent diu allà el que no és capaç o bé no s'atreveix a dir en persona.

Sí bé aquestes xarxes tenen codis de conducta i contenen d'equips de moderadors, no es pot fer front a tots els problemes. Al fet que se li dona més importància i es destinen més recursos de forma activa és a l'anomenat "cyber-bullying", doncs és el que més canta i no deixa de ser un atac des d'un individu cap a un altre. Si més no, també hi ha altres problemes que potser no són tant cridaners, però estan en molta més quantitat. Un d'ells és la depressió que pateixen molts dels seus usuaris. Com bé ja s'ha comentat, les xarxes socials permeten posar comentaris a internet, sent un lloc perfecte per la gent amb problemes d'ànims o de depressió per poder expressar-se i deixar anar tot els que els hi preocupa. [3]

Un estudi demostra que més d'un 20% dels usuaris han penjat comentaris amb indicis que podrien estar patint depressió o similar. No només això, sinó que és una moda que està en augment, havent-hi el doble de casos, proporcionalment, ara que fa 10 anys. Addicionalment, hi ha hagut alguns casos on els usuaris expliquen les "penúries" del seu dia a dia fins al punt on escriuen allà la mateixa la nota de suïcidi. [4]

Clarament, totes aquestes notícies han provocat un seguit de queixes a les empreses propietàries de les xarxes per part de moltes organitzacions i institucions. Els responsables de moderació de les xarxes socials es defensen dient que no hi ha manera de poder veure tots els posts amb indicis de depressió. A diferència dels que contenen "cyber-bullying" o similars, que són reportats per altres usuaris (normalment

les víctimes), els missatges amb continguts depriments passen desapercebuts, o simplement no se'ls hi dona importància, per la resta d'usuaris.

Aquest fet dificulta moltíssim la feina dels equips moderadors, els quals no tenen els mitjans necessaris per avaluar tots els missatges i comentaris. No només això, sinó que, al no estar incomplint cap normativa, tampoc poden prendre cap acció al respecte.

2 OBJECTIUS

Així doncs, hi ha la necessitat d'obtenir un sistema que pugui fer front al problema esmentat anteriorment. Per la seva pròpia naturalesa, s'ha de solucionar no pas actuant un cop passa, sinó de forma preventiva prenent accions abans de que sigui massa tard.

L'objectiu és poder identificar els casos d'usuaris que necessitin ajuda mitjançant models predictius basats en intel·ligència artificial. Més específicament, cal treballar i investigar l'anomenat "Natural Language Processing", traduït a processament de llenguatge natural.

Clarament, ja hi ha molts mètodes i models disponibles que realitzen la tasca desitjada. L'objectiu, per tant, no és crear des de zero un nou mètode, sinó fer un estudi de l'eficiència i èxit que tenen cadascun d'ells. Per tant, s'implementaran un seguit de mètodes diferents i es procedirà a fer les proves adients. Les dades utilitzades són datasets ja classificats correctament. Aquestes dades són extretes directament i sense tractar de les xarxes socials Twitter i Reddit. Aquest fet per una banda permet tenir una representació pràcticament exacte de les dades amb les que s'enfronten els varis models en el moment de la veritat. Per una altra banda, al ser informació sense tractar, també obre la porta a fer un Exploratory Data Analysis (EDA) i treure ja unes estadístiques i característiques preliminars, les quals després es podran comparar amb els resultats arribats un cop executats els models. Addicionalment, els models es posaran a prova tant amb les dades sense tractar, com fent un previ tractament del dataset amb l'objectiu de veure el nou comportament dels models i si hi ha algun indicatiu de millora a les produccions.

Per poder arribar a tots els objectius detallats a continuació, caldrà fer ús d'un conjunt de dades que tinguin la informació amb la que després s'hauran d'enfrontar els models. Des del portal "Kaggle", on es poden trobar molts datasets de qualsevol àmbit, se'n descarreguen uns quants:

- Mental Health Twitter [8]
- Depression social media [9]
- Depression: Reddit Dataset [10]
- Sentimental Analysis for Tweets [11]

Se'n agafa més d'un perquè tots tenen alguna característica especial que els fa únics. El primer és el que es podria considerar més interessant, doncs, a part del text del tweet, té un seguit d'atributs com el número de seguidors o número de tweets diaris. D'aquesta manera permetrà veure realment fins a quin punt el NLP és efectiu per trobar cassos de depressió a xarxes socials. Primer se li s'aplicarà únicament el sentiment analysis, i després també s'utilitzaran la resta d'atributs per veure si hi ha gaire millora. El segon

es caracteritza per estar classificats en una escala de valors que defineixen el nivell de depressió, a diferència de la resta on únicament tenen valors de “depressió” o “no depressió”. El tercer es caracteritza per estar ja netejat d’emojis, links d’internet i similars. Finalment, el quart únicament conté el tweet i la classificació.

Dins del Machine Learning ja s’ha fet un seguit d’investigacions i treballs des dels quals es basaran els objectius.

Per una banda hi ha els mètodes tradicionals de Machine Learning. L’article “Sentiment Analysis of Review Datasets using Naïve Bayes’ and K-NN Classifier” [5] explora tant Naïve Bayes i K-NN, intentant evaluar el rendiment i resultats de cada un. Tant un com l’altre es comporten similarment i es postulen com a bones opcions dins d’aquest tipus de mètodes, però sent el Naïve Bayes el que ho fa millor amb el 90% d’accuracy. Tot i això, l’estudi d’aquests mètodes s’orienta a fer un sentiment analysis d’opinions de pel·lícules, no pas de posts a les xarxes socials. També hi ha més estudis, com el “Sentiment Analysis on Twitter Data using KNN and SVM” [6] que sí que fan el sentiment analysis sobre un cas de xarxes socials. Per aquest últim, el mètode de SVM és el que dona més bons resultats. Tot i això, no s’està buscant específicament depressió, sinó si el tweet és positiu o negatiu.

Un dels objectius als que es vol arribar és comparar varis models tradicionals i posar-los a prova directament en detectar depressió als posts en xarxes socials. D’aquesta manera es podrà veure si els que tenen més èxit en sentiment analysis genèric, també el tenen quan es busca un sentiment en concret. Més específicament, per aquesta part es posaran a prova els classificadors Naïve Bayes, Decision Tree, Random Forest, SVM i KNN.

Adicionalment, també hi ha maneres de millorar les prediccions de varis classificadors, l’article “Comparative Analysis of Bagging and Boosting Algorithms for Sentiment Analysis” [7] n’explora alguns. Els resultats són millors que aplicant un simple classificador, però, de nou, no posa pas èmfasi en la depressió. Per tant, en aquest apartat també es posarà a prova els mètodes d’ensamblament Boosting i Bagging aplicant-los als classificadors prèviament mencionats, amb l’objectiu de veure fins a quin punt es pot arribar a millorar les prediccions de depressió per cada classificador.

Finalment, aquests mètodes més tradicionals tenen un seguit de paràmetres, anomenats Hyper Parameters, amb els quals configurar el seu comportament. Segons l’article “Hyperparameter Search in Machine Learning” [12], aconseguir trobar els valors adequats pot marcar la diferència entre unes bones prediccions i unes males. Caldrà veure fins a quin punt una bona elecció de Hyper Parameters pot afectar a les prediccions quan es busca un sentiment anàlisi de depressió.

Per una altra banda, també hi ha els mètodes basats en el Deep Learning que actualment estan triomfant més. L’estudi “Comparing Machine Learning and Deep Learning Approaches on NLP Tasks for the Italian Language” [13] parla sobre les diferències que hi ha entre el Deep Learning i els mètodes tradicionals de Machine Learning, conclouent que per NLP el Deep Learning no acaba de tenir superioritat en les tasques de classificació que depenen molt en l’anàlisi semàntic.

Un mètode que ja porta un temps utilitzant-se per NLP són les anomenades Recurrent Neural networks (RNN). Es diferencien de les xarxes neuronals convencionals, i són especialment útils en NLP, pel fet que el input és una sola paraula, donant flexibilitat per treballar amb diferents llargades a les frases. Hi ha dos models que fins fa poc eren els més utilitzats, anomenats LSTM i GRU. L'article "LSTM and GRU neural network performance comparison study" [17] explora aquests dos mètodes, on conclou que el fet de retenir informació els hi permet posicionar-se al podi de les RNN. No obstant, està comparant amb tasques de molts camps, sense entrar en detall al NLP i encara menys al Sentiment Analysis. Un dels objectius serà aplicar ambdós mètodes per detectar depressió a les xarxes socials.

L'article "Attention is All you Need" [14], publicat per Google el 2017, presenta un nou model de xarxes neuronals especialment innovador a l'apartat de Natural Language Processing. S'anomena "Transformer" i es basa processar dades d'entrada seqüencials però, a diferència de les Xarxes Neuronals Recurrents, processen tota l'entrada alhora. L'article "Overview of the Transformer-based Models for NLP Tasks"[16] fa especialment atenció en el rendiment dels models Transformer amb els problemes NLP, conclouent que efectivament han millorat molt les prediccions respecte a la resta de mètodes utilitzats fins al moment. Un objectiu serà aplicar varis models Transformers (GPT, BERT i RoBERTa) i veure com es comporten a l'hora de detectar casos de depressió a les xarxes socials, comparant-los tant entre ells com amb la resta de mètodes provats anteriorment.

3 METODOLOGIA

Per poder desenvolupar el projecte de manera adequada i ordenada, és necessari establir una metodologia de treball la qual seguir, permetent així tenir un bon control del flux de treball i compliment de les tasques i entregues. La metodologia escollida és l'anomenada "àgil".

Es basa en iteracions curtes, adaptabilitat als canvis, facilitat de detectar i solucionar errors, i entregues parcials. És un projecte que es pot subdividir en diferents apartats independents entre ells i aquesta metodologia és especialment útil en aquests casos. D'aquesta manera, si una part presenta complicacions, la resta no es veuran pràcticament afectades. També permet fer, al final de cada iteració, un anàlisi de la situació actual, valorant el que s'ha realitzat, el que manca per realitzar i qualsevol aspecte que calgui modificar/adaptar.

Finalment, també es fa us d'un controlador de versions GIT, tenint així un registre i historial de tots els canvis i permetent revertir-los en cas de fer falta.

4 PLANIFICACIÓ

Iteració 1:

L'objectiu és començar amb el TFG, fer un estudi inicial de l'estat de l'art, definir els objectius, metodologia a utilitzar durant tot el desenvolupament, especificar les tasques de la resta d'iteracions.

SETMANA	TASCA	RESULTATS
Setembre, 3	<ul style="list-style-type: none">• Reunió Inicial	<ul style="list-style-type: none">• Fet
Setembre, 4	<ul style="list-style-type: none">• Informe inicial (Objectiu)• instal·lar LaTeX	<ul style="list-style-type: none">• Informe inicial pràcticament finalitzat.

Iteració 2:

L'objectiu d'aquesta iteració és treballar en els mètodes tradicionals de Machine Learning i obtenir resultats de les prediccions de depressió a les xarxes socials. Per una banda s'implementaran els models definits a les tasques, i per una altra banda es mirarà de fer millores mitjançant tècniques com ensambles o hyper-paramter tuning.

Octubre, 1	<ul style="list-style-type: none">• Primera sessió de seguiment• Buscar informació i estat de l'art• Definir objectius• Detallar planificació i tasques	<ul style="list-style-type: none">• Informe inicial finalitzat i entregat.• Reunió Feta.
Octubre, 2	<ul style="list-style-type: none">• Preparar entorn de desenvolupament a l'ordinador<ul style="list-style-type: none">○ Instal·lar Python i IDE.○ Instalar llibreries necessàries per aquesta iteració.• Realitzar un Exploratory Data Analysis dels datasets amb els que es treballarà.• Aplicar els classificadors Naïve Bayes• Aplicar el classificador Decision Tree• Treballar en l'informe de progrés 1.	
Octubre, 3	<ul style="list-style-type: none">• Aplicar el classificador Random Forest• Aplicar el classificador Support Vector Machine	

	<ul style="list-style-type: none"> • Aplicar el classificador K Nearest neighbour • Seguir treballant en l'informe de progrés 1. 	
Octubre, 4	<ul style="list-style-type: none"> • Millorar els resultats trobats fins ara. <ul style="list-style-type: none"> ○ Utilitzar per cada classificador el mètode Boosting ○ Utilitzar per cada classificador el mètode Bagging. • Seguir treballant en l'informe de progrés 1. 	
Novembre, 1	<ul style="list-style-type: none"> • Enllestir informe progrés 1 • Ratificar planificació (si escau) • Proposar canvis (si escau) • Millorar els resultats obtinguts fins al moment utilitzant els mètodes de Hyper Parameter Tuning. <ul style="list-style-type: none"> ○ Aplicar Grid Search ○ Aplicar Random Search ○ Aplicar Bayesian Optimization 	

Iteració 3

L'objectiu d'aquesta iteració és utilitzar els mètodes més recents basats en Deep Learning i obtenir resultats de les prediccions de depressió a les xarxes socials. Per una banda es començarà utilitzant Recurrent Neural Networks, explorant els models que tenen més èxit i com es comporten amb el sentiment analysis.

Per altra banda també s'utilitzarà el mètode més recent i que ara mateix està sent el que té més èxit, les anomenades xarxes Transformer. Es posaran a prova els models que hi estan basats i es veurà quin resultat obtenen, comparant-los tant amb les RNN com amb els mètodes tradicionals.

Novembre, 2	<ul style="list-style-type: none"> • Segona sessió de seguiment • Instal·lar llibreries necessàries per aquesta iteració. • Aplicar la xarxa neuronal LSTM • Seguir treballant en l'informe de progrés 2 	
Novembre, 3	<ul style="list-style-type: none"> • Aplicar la xarxa neuronal GRU • Seguir treballant en l'informe de progrés 2 	

Novembre, 4	<ul style="list-style-type: none"> • Aplicar la xarxa Transformers GPT • Seguir treballant en l'informe de progrés 2 	
Desembre, 1	<ul style="list-style-type: none"> • Aplicar la xarxa Transformers BERT • Seguir treballant en l'informe de progrés 2 	
Desembre, 2	<ul style="list-style-type: none"> • Aplicar la xarxa Transformers RoBERTa • Finalitzar l'informe de progrés 2 	

Iteració 4

L'objectiu d'aquesta iteració ja no és seguir desenvolupant, sinó treure conclusions del que s'ha fet fins al moment i preparar tot el material per la presentació i l'entrega de l'informe final. També caldrà deixar ordenat el dossier del treball.

Desembre, 3	<ul style="list-style-type: none"> • Tercera sessió de seguiment • Treure conclusions del treball 	
Desembre, 4	<ul style="list-style-type: none"> • Vacances 	
Gener, 1	<ul style="list-style-type: none"> • Vacances 	
Gener, 2	<ul style="list-style-type: none"> • Fer versió provisional de l'informe final 	
Gener, 3	<ul style="list-style-type: none"> • Quarta sessió de seguiment • Fer versió provisional de l'informe final 	
Gener, 4	<ul style="list-style-type: none"> • Preparar dossier • preparar presentació provisional 	
Febrer, 1	<ul style="list-style-type: none"> • Cinquena sessió de seguiment • Corregir versions provisionals i deixar-les com a finals 	
Febrer, 2	<ul style="list-style-type: none"> • Preparar defensa TFG 	
Febrer, 3	<ul style="list-style-type: none"> • Defensa TFG 	

BIBLIOGRAFIA

[1] Yasar, K. (2022, 12 abril). *social networking*. WhatIs.com. Recuperado 2 de octubre de 2022, de <https://www.techtarget.com/whatis/definition/social-networking>

[2] boyd, D. M. & Ellison, N. B. (2007, octubre). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>

- [3] Chan, T. K. H., Cheung, C. M. K., & Lee, Z. W. Y. (2020, December 5). *Cyberbullying on social networking sites: A Literature Review and future research directions*. Information & Management. Retrieved October 2, 2022, from <https://www.sciencedirect.com/science/article/pii/S0378720620303499>
- [4] Rosenquist, J. N. (2010, 16 marzo). *Social network determinants of depression*. Nature. Recuperado 2 de octubre de 2022, de <https://www.nature.com/articles/mp201013>
- [5] Dey, L. (2016, 31 octubre). *Sentiment Analysis of Review Datasets Using Naive Bayes and K-NN Classifier*. arXiv.org. Recuperado 5 de octubre de 2022, de <https://arxiv.org/abs/1610.09982>
- [6] *Sentiment analysis on Twitter data using KNN and SVM - semantic scholar*. (n.d.). Retrieved October 5, 2022, from <https://pdfs.semanticscholar.org/05a8/78000170abcd0c6f8208080470858422e17c.pdf>
- [7] *Sentiment analysis on Twitter data using KNN and SVM - semantic scholar*. (n.d.). Retrieved October 5, 2022, from <https://pdfs.semanticscholar.org/05a8/78000170abcd0c6f8208080470858422e17c.pdf>
- [8] <https://www.kaggle.com/code/ardawrld/twitter-sentiment-analysis-about-the-depression/data>
- [9] <https://www.kaggle.com/code/mpwolke/depression-sentiment-analysis-classifiers/data>
- [10] <https://www.kaggle.com/datasets/infamouscoder/depression-reddit-cleaned>
- [11] <https://www.kaggle.com/datasets/gargmanas/sentimental-analysis-for-tweets>
- [12] Claesen, M., & De Moor, B. (2015, April 6). *Hyperparameter search in machine learning*. arXiv.org. Retrieved October 5, 2022, from <https://arxiv.org/abs/1502.02127>
- [13] Bernardo, M., Alberto, L., & Simone, M. (2020, May 18). *Comparing machine learning and deep learning approaches on NLP tasks for the Italian language*. IRIS. Retrieved October 6, 2022, from <https://cris.fbk.eu/handle/11582/322156>
- [14] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017, December 6). *Attention is all you need*. arXiv.org. Retrieved October 6, 2022, from <https://arxiv.org/abs/1706.03762>
- [15] Kokab, S. T., Asghar, S., & Naz, S. (2022, April 10). *Transformer-based deep learning models for the sentiment analysis of social media data*. Array. Retrieved October 6, 2022, from <https://www.sciencedirect.com/science/article/pii/S2590005622000224>
- [16] *Overview of the Transformer-based models for NLP tasks*. IEEE Xplore. (n.d.). Retrieved October 6, 2022, from <https://ieeexplore.ieee.org/abstract/document/9222960/>
- [17] *LSTM and GRU neural network performance comparison study: Taking Yelp Review dataset as an example*. IEEE Xplore. (n.d.). Retrieved October 6, 2022, from <https://ieeexplore.ieee.org/document/9221727>

