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’internat 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 conducte 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 mateix 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 indici 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.

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. Addicionalment, 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.

# 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

|  |  |  |
| --- | --- | --- |
| **SETMANA** | **TASCA** | **RESULTATS** |
| Setembre, 3 | Reunió Inicial | Fet |
| Setembre, 4 | Informe inicial (Introducció, Objectiu, metodologia, ...), instal·lar LaTex | Informe inicial pràcticament finalitzat.  El LaTex es resisteix a funcionar correctament... |
| Octubre, 1 | Primera sessió de seguiment, Buscar informació i estat de l’art | Informe inicial finalitzat i entregat. |
| Octubre, 2 | Investigació i primeres proves |  |
| Octubre, 3 | Investigació i primeres proves |  |
| Octubre, 4 | Comparar primeres proves, donar algunes conclusions preliminars |  |
| Novembre, 1 | Preparar informe progrés 1 |  |
| Novembre, 2 | Segona sessió de seguiment |  |
| Novembre, 3 | Seguir fent més proves amb mètodes diferents |  |
| Novembre, 4 | Proposar millores als mètodes trobats i provar-les |  |
| Desembre, 1 | Proposar millores als mètodes trobats i provar-les |  |
| Desembre, 2 | Preparar informe progrés 2 |  |
| Desembre, 3 | Tercera sessió de seguiment |  |
| Desembre, 4 | Vacances |  |
| Gener, 1 | Vacances |  |
| Gener, 2 | Preparar informe final (deixar-ho tot bonic i ordenat) |  |
| Gener, 3 | Quarta sessió de seguiment |  |
| Gener, 4 | Preparar dossier, preparar presentació |  |
| Febrer, 1 | Cinquena sessió de seguiment |  |
| Febrer, 2 | Preparar defensa TFG |  |
| Febrer, 3 | 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