

APRENTATGE COMPUTACIONAL

SENTIMENTAL ANALYSIS - AMAZON

INFORME

PROJECTE FINAL DE SEMESTRE

CURS 2024 - 2025

UAB | UNIVERSITAT AUTÒNOMA DE BARCELONA

1. INTRODUCCIÓ
 - 1.1. OBJECTIUS I MOTIVACIÓ
 - 1.2. METODOLOGIA
2. ANÀLISIS DE LES DADES
 - 2.1. DESCRIPCIÓ DEL CONJUNT DE DADES
 - 2.1.1.
 - 2.1.2.
 - 2.1.3.
 - 2.1.4.
 - 2.1.5.
 - 2.1.6.
 - 2.2. PREPARACIÓ I NETEJA DE DADES
 - 2.3.

#dense array al model naive bayes

1. INTRODUCCIÓ

Cada dia es generen milions de textos a Internet, des de ressenyes de productes dinas a comentaris en les xarxes socials. Aquesta gran quantitat d'informació no només és aclaparadora, sinó que també planteja un repte fascinant: com podem entendre i analitzar totes aquestes opinions i emocions de manera automatizada? Els models d'aprenentatge automàtic han emergit com una de les solucions poderoses per processar i extreure coneixement útil d'aquestes immens volum de dades, el 90% de les generades són textuals i especialment adequades per aquest tipus d'anàlisis.

L'aprenentatge automàtic s'ha convertit en una eina fonamental per abordar problemes complexos en molts àmbits, incloent-hi el processament del llenguatge natural (NPL). Un dels seus casos d'ús més comuns és l'anàlisi de sentiments, una tècnica que permet determinar les emocions o actituds expressades en text, classificant-les positivament, negativament o de forma neutra segons el context. Aquesta aplicació resulta especialment útil en àrees com el màrqueting, l'opinió pública i la detecció d'opinions en temps real.

Els models de sentiment es basen en algorismes d'aprenentatge supervisat que aprenen patrons a partir de grans conjunts de dades etiquetades. Aquests conjunts de dades contenen textos associats a una classificació específica, cosa que permet als models predir els sentiments en textos nous amb una precisió considerable.

En aquest projecte mateix, ens centrem en aquest tipus d'anàlisis mitjançant un model de classificació que treballa amb un conjunt de dades reals. Explorarem els resultats obtinguts, avaluarem la seva precisió i analitzarem possibles biaixos o limitacions del model. Amb aquest procés, acabarem aprofundint en el funcionament i aplicacions d'aquest tipus de models, a la vegada que identifiquem les oportunitats per optimitzar el seu rendiment.

1.1. OBJECTIUS I MOTIVACIÓ

El projecte té com a objectiu principal desenvolupar i avaluar un model d'anàlisi de sentiment aplicat a un conjunt de dades textuals reals ja donat. Això hi inclou molts reptes específics:

- a) **Desenvolupament tècnic:** implementar un model d'aprenentatge supervisat capaç de classificar sentiments en textos en certs grups, segons la categoria i nivell. A més a més, volem comparar diferents enfocaments/models per identificar quins mètodes i/o agrupacions ofereixen els millors resultats en termes de precisió i robustesa.
- b) **Gestió de dades:** preparar, netejar i transformar el conjunt de dades d'entrada per assegurar que compleix el requisits del model. Seguit d'això, analitzar els patrons subjacents a les dades per identificar possibles biaixos i desafiaments al conjunt de dades utilitzat.

- c) **Avaluació i anàlisi crítica:** mesurament de l'efectivitat del model tenint en compte mètriques relacionades, per exemple, amb la matriu de confusió o altres operacions. També, identificar les limitacions del model segons els resultats donats per tal de proposar millores o ajustos que puguin millorar-ne el rendiment.

La motivació darrere aquest projecte es basa en la rellevància i les aplicacions pràctiques de l'anàlisi de sentiments en múltiples sectors com l'impacte pràctic que té en les estratègies utilitzades en els diferents àmbits d'avui en dia, i el desafiament tecnològic que ens acompanya aquest treball complex en tots els aspectes. Aquest projecte no només serà una eina d'aprenentatge sinó que també una manera d'explorar el potencial transformador de l'aprenentatge automàtic, que està en constant evolució.

1.2. METODOLOGIA