

III Congreso y XVI Jornadas de Usuarios

LobbyBot: Análisis y Clasificación Automatizada de las Estrategias de los Grupos de Interés en las Noticias de los Medios de Comunicación de España

Aritz Gorostiza Cerviño – Universidad de Málaga – olivetti03@uma.es

Encarna Hidalgo – Universidad de Granada - ehidalgo@ugr.es

Antonio Castillo – Universidad de Málaga acastilloe@uma.es

Esta investigación presenta la segunda fase de "LobbyBot", una herramienta desarrollada en R y presentada en el II Congreso de R y XIII Jornadas de Usuarios de R en Barcelona en 2023. En esta segunda fase, nos enfocamos en la construcción de un modelo de aprendizaje automático para clasificar noticias según la estrategia comunicativa empleada por los grupos de interés que aparecen en ellas. Para ello, se desarrollará un modelo de aprendizaje automático utilizando el paquete CARET en R. Este paquete (abreviatura de Classification And REgression Training) es un conjunto de funciones que intenta simplificar el proceso de creación de modelos predictivos. La base de datos está compuesta por 431,797 noticias recolectadas en el último año mediante "LobbyBot", de las cuales, en aproximadamente 5,768, se menciona a algún grupo de interés registrado en la CNMV. El corpus lingüístico obtenido se dividió en un conjunto de entrenamiento (75%), que será etiquetado manualmente por los investigadores para validar el modelo con los datos restantes. Los resultados ofrecerán una visión detallada sobre las dinámicas de las estrategias comunicativas de los grupos de interés en los medios de comunicación españoles. Además, demostrarán la capacidad de las técnicas de procesamiento de lenguaje natural (NLP) y aprendizaje automático (ML) para analizar las estrategias de comunicación predominantes según la naturaleza de cada organización en los medios de comunicación de España.