Text Mining of Veterinary Forums for Epidemiological Surveillance Supplementation

Authors : Samuel Munaf, Kevin Swingler, Franz Brülisauer, Anthony O'Hare, George Gunn, Aaron Reeves

Abstract : Web scraping and text mining are popular computer science methods deployed by public health researchers to augment traditional epidemiological surveillance. However, within veterinary disease surveillance, such techniques are still in the early stages of development and have not yet been fully utilised. This study presents an exploration into the utility of incorporating internet-based data to better understand the smallholder farming communities within Scotland by using online text extraction and the subsequent mining of this data. Web scraping of the livestock fora was conducted in conjunction with text mining of the data in search of common themes, words, and topics found within the text. Results from bi-grams and topic modelling uncover four main topics of interest within the data pertaining to aspects of livestock husbandry: feeding, breeding, slaughter, and disposal. These topics were found amongst both the poultry and pig sub-forums. Topic modeling appears to be a useful method of unsupervised classification regarding this form of data, as it has produced clusters that relate to biosecurity and animal welfare. Internet data can be a very effective tool in aiding traditional veterinary surveillance methods, but the requirement for human validation of said data is crucial. This opens avenues of research via the incorporation of other dynamic social media data, namely Twitter and Facebook/Meta, in addition to time series analysis to highlight temporal patterns.

Keywords : veterinary epidemiology, disease surveillance, infodemiology, infoveillance, smallholding, social media, web scraping, sentiment analysis, geolocation, text mining, NLP

Conference Title : ICSMMSNA 2023 : International Conference on Social Media Mining and Social Network Analysis
Conference Location : Vancouver, Canada
Conference Dates : May 22-23, 2023