

# altinel\_2022\_social\_media\_analysis\_by\_innovative\_hybrid\_algorithms\_with\_label\_propagation

## Year

2022

## Author(s)

Ayse Berna Altinel

## Title

Social media analysis by innovative hybrid algorithms with label propagation

## Venue

Expert Systems With Applications

---

## Topic labeling

Manual

## Focus

Secondary

## Type of contribution

Established approach

## Underlying technique

Manual labeling assisted by associated documents

## Topic labeling parameters

\

## Label generation

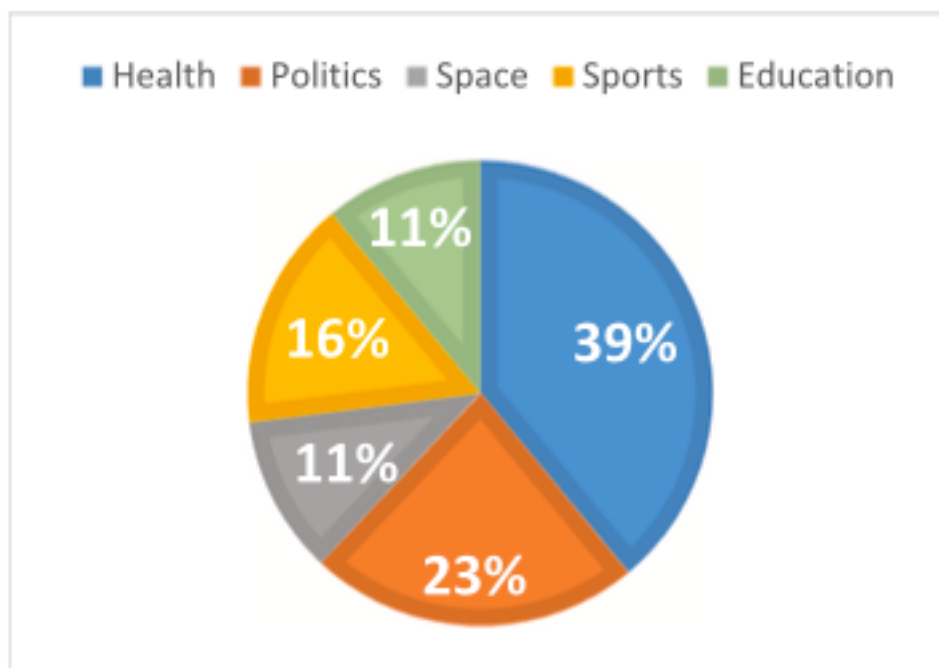
Human experts reviewed the word clusters generated by LDA, and some tweets were extracted from their categories.



**Fig. 4.** Word cloud images for the Health, Politics, and Space classes.

## Motivation

"To complete the dataset, topic labels were added to the pooled tweets. As a result, five different topic categories obtained."



**Fig. 5. Distribution of Categories.**

## Topic modeling

LDA

## Topic modeling parameters

Nr of topics (K): 5

## Nr. of topics

5 (One for each category)

---

## Label

Single word labels belonging to one of five categories (Health, Politics, Space, Sports, Education)

## Label selection

\

## Label quality evaluation

\

## Assessors

\

---

## Domain

Paper: Social media analysis

Dataset: Social media (Twitter)

## Problem statement

Due to the huge size of the data accumulated on microblogging sites, recently, two fundamental questions have become very popular: 1) What percentage of this accumulated data has positive or negative sentiment polarity? 2) How is the distribution of this accumulated data on different topics?

This paper presents several different algorithms which are based on the Label Propagation

Algorithm (LPA) in order to handle previously mentioned two fundamentals tasks: sentiment polarity detection task and topic-based text classification task.

## Corpus

Origin: Twitter

Nr. of documents: 23,214,872

Details:

- tweets within the following categories: Politics, Education, Health, Space, and Sports
- collected between February 1st and March 15th, 2021

## Document

Textual content of a single tweet together with the category to which the tweet belongs to

## Pre-processing

- mentions, hashtags, URLs, emojis, and punctuations were removed
- stemming
- stop word filtering
- pooling of tweets belonging to the same category

---

```
@article{altinel_2022_social_media_analysis_by_innovative_hybrid_algorithms_wit  
h_label_propagation,
```

```
    abstract = {Due to the huge size of the data accumulated on microblogging  
sites, recently, two fundamental questions have become very popular: 1) What  
percentage of this accumulated data has positive or negative sentiment  
polarity? 2) How is the distribution of this accumulated data on different  
topics? Inspired by these motivated necessities, this paper presents several  
different algorithms which are based on the Label Propagation Algorithm (LPA)  
in order to handle previously mentioned two fundamentals tasks: sentiment  
polarity detection task and topic-based text classification task. These  
algorithms are the Label Propagated- Relevance Frequency Classifier (LP-RFC)  
and LP-Abstract Frequency Classifier (LP-AFC). These algorithms can be defined  
as new semantic smoothing classifiers, which take advantage of the semantic  
connections among terms in the label propagation phase of the LPA.  
Additionally, another classifier, namely LP-ComRFC+AFC, was built. LP-
```

ComRFC+AFC is actually a weighted summation classifier of the individual LP-RFC and LP-AFC. Furthermore, considering the shortage of labeled data in real-world scenarios, a semi-supervised version of LP-RFC and LP-AFC, namely ``Merging Unlabeled and Labeled Instances with Semantic Values of Terms'' (MULIS), was designed and implemented. For the experiments of the sentiment polarity detection task, three different datasets were use and for the experiments of topic-based text classification task, a self-collected tweet dataset was use. According to the experimental results, the suggested algorithms, and their composite form, LP-ComRFC+AFC, generated higher F1 scores than all of the baseline algorithms at nearly all of the training splits on the datasets.},

```
author = {Ay{\c s}e Berna Altınel},
date-added = {2023-03-26 19:06:45 +0200},
date-modified = {2023-03-26 19:06:45 +0200},
doi = {https://doi.org/10.1016/j.eswa.2022.118606},
issn = {0957-4174},
journal = {Expert Systems with Applications},
keywords = {Label propagation algorithm, Social media analysis, Topic-based
tweet classification, Sentiment polarity detection},
pages = {118606},
title = {Social media analysis by innovative hybrid algorithms with label
propagation},
url = {https://www.sciencedirect.com/science/article/pii/S095741742201658X},
volume = {210},
year = {2022}}
```