# Fast fake news detection for smart cites

T. Milinda H. Gedara
*Dept. Management & Innovation Systems*
*University of Salerno*
Salerno, Italy
thewadaundagedara@unisa.it

Domenico Santaniello
*Dept. Industrial Engineering*
*University of Salerno,*
Salerno, Italy
dsantaniello@unisa.it

Stefania Tomasiello
*Institute of Computer Science*
*University of Tartu,*
Tartu, Estonia
stefania.tomasiello@ut.ee

## I. Introduction

The current state of life in major cities is characterized by uncertainty and complexity. In recent years, major cities worldwide have faced significant challenges and disruptions caused by the ongoing pandemic, leading to a drastic transformation in people's lives and access to essential services. To effectively address these crisis risks and ensure the well-being of residents, the implementation of a crisis-recovery risk-management plan that incorporates digitalization initiatives has proven to be highly valuable. In addition to the daily risks posed by natural disasters and accidents, significant infrastructure damage and inconveniences have emerged, necessitating swift and decisive action to reduce citizen vulnerabilities and mitigate further harm to infrastructure. As a result, governing bodies are increasingly recognizing the importance of smart city risk management as a critical task.
Wang et al. [1] emphasized the importance of smart city technologies and demonstrated their widespread use for effective risk management during crises. The emergence of the COVID-19 pandemic highlighted the need to implement technologies to mitigate crisis risks and normalize the situation. People lose their faith in information spreading among the communities over time. During the COVID-19 outbreak, there was a lot of fake news about this pandemic on social media, which affected people's health-protective activities [2].

The term ”fake news” is sometimes described as ”misinformation”, ”disinformation”, ”hoax”, and ”rumour”, all of which are variations of misleading information. There are numerous research projects, methods, and applications for fact-checking and fake news identification [3]. However, traditional fact-checking is time-consuming and labour-intensive work. People in smart cities spread the news before it is verified as fake or real. Therefore some automated methods are needed to detect fake news quickly.

## II. Literature review

Fake news spreads quickly in digital space, and detecting it as quickly as possible allows to stop and regulate its spread. Several machine learning models have been used for detecting fake news. Several classifiers, including Nave Bayes, linguistic feature-based, bounded decision tree models, SVM, have been tried, but not leading to high accuracy [4]. Ozbay and Alatas [5] proposed employing supervised ML algorithms such as decision trees, Locally Weighted Learning and many more to detect fake news in social media. The three types of datasets used in this approach were BuzzFeed, random political datasets, and ISOT false news. Sadeghi et al. [6] proposed natural language processing and deep learning to perform experiments. Here the biLSTM model achieved the highest accuracy among the test cases. However, the main issue with deep learning models is that the training time is much higher, and more computational power is needed to develop the model.

## III. Methodology

Armed with such a vision, we propose an LSTM-based model with a data reduction technique to lower the computational time. The pipeline for this approach is:
**pre-processing $\rightarrow$ data reduction $\rightarrow$ classification**

The first stage is pre-processing data, typically removing punctuation and misplaced symbols. Afterwards, the feature reduction step is used to reduce the size of the features. High-dimensionality data is regarded as a key concern during the data pre-processing phase, and it is required to reduce redundant and unrelated features in order to improve accuracy. By minimizing the number of features, the processing pace is reduced, resulting in improved performance. Then a conventional LSTM model is trained for the binary classification problem of fake news detection.

### A. Principal component analysis

Principal Component Analysis (PCA) is a reduction method used in large datasets to reduce their cardinality while retaining significant information. The reduction is achieved by finding directions, known as primary components, and increasing the data variety to its maximum. The resulting dataset is slimmer, yet it keeps the original dataset's features information. The new dataset may have the same or fewer features than the original dataset. The major components are computed using the covariance matrix. These elements are listed in decreasing order of importance.

Let us assume that the original data matrix comprises $a$ attributes and $b$ observations and it is required to reduce its cardinality to $t$ dimensional subspace. Then its transformation can be given by the following equation:

$$Y = (E^Z X), \quad (1)$$

where $E_{a \times t}$ is the projection matrix which contains $t$ eigenvectors corresponding to $t$ highest eigenvalues, and where $X_{a \times b}$ is the mean-centered data matrix.

PCA has several advantages, including the ability to remove associated features, improve algorithm efficiency by reducing the number of features used, reduce overfitting, and improve visualization. Therefore PCA is deemed to be more suitable to achieve higher accuracy while reducing the computation time due to the reduced data cardinality. .

## IV. Conclusion and Future Work

Fast fake information detection in smart cities is critical to enhance risk management during an emergency or crisis. Therefore fake news needs to be identified quickly and with high accuracy. A data reduction technique, such as PCA, has been here investigated for fast detection. The proposed PCA approach was assessed using some publicly available datasets, such as the popular CoVID19-FNIR. Initial experiments showed the satisfactory performance of the approach. In the future, we will dive into reduction techniques to improve accuracy and computational time.

## References


[1] Wang, Hongxin, Wenqing Wu, and Yenchun Jim Wu. 2021. How Does the Citizens' Behavioral Intention to Use Digital Technology in Smart Cities Affect Their Mental Health During the Covid-19 Pandemic? Cham: Springer.

[2] S. van der Linden, J. Roozenbeek, J. Compton, Inoculating against fake news about covid-19, Frontiers in Psychology 11 (2020). URL: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.566790. doi:10.3389/fpsyg.2020.566790.

[3] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018. Detection and Resolution of Rumours in Social Media: A Survey. ACM Comput. Surv. 51, 2, Article 32 (March 2019), 36 pages. https://doi.org/10.1145/3161603.

[4] H. Saleh, A. Alharbi and S. H. Alsamhi, "OPCNN-FAKE: Optimized Convolutional Neural Network for Fake News Detection," in IEEE Access, vol. 9, pp. 129471-129489, 2021, doi: 10.1109/ACCESS.2021.3112806.

[5] Feyza Altunbey Ozbay, Bilal Alatas,Fake news detection within online social media using supervised artificial intelligence algorithms,Physica A: Statistical Mechanics and its Applications,Volume 540,2020,123174,ISSN 0378-4371, https://doi.org/10.1016/j.physa.2019.123174.

[6] Sadeghi, F., Bidgoly, A.J.  Amirkhani, H. Fake news detection on social media using a natural language inference approach. Multimed Tools Appl 81, 33801–33821 (2022). https://doi.org/10.1007/s11042-022-12428-8

[7] D. Hou, J. Zhang, Z. Yang, S. Liu, P. Huang and G. Zhang, "Distribution water quality anomaly detection from UV optical sensor monitoring data by integrating principal component analysis with chi-square distribution", Opt. Express, vol. 23, no. 13, pp. 17487, Jun. 2015.