

# Covid predictions using Google Search Query Data

Nicole Bonilla

<https://github.com/kneecoul/ITCS5156-051.git>

# Problem

- Building a real time framework using online activity data, like Google Search, using Deep Learning and forecasting methods and comparing them with Google's prediction of Covid-19 cases in the United States.

# Paper

- Venkatesh, U, and Periyasamy Aravind Gandhi. "Prediction of COVID-19 Outbreaks Using Google Trends in India: A Retrospective Analysis." *Healthcare informatics research* vol. 26,3 (2020): 175-184.  
doi:10.4258/hir.2020.26.3.175

# Method

- The Google search terms used for the analysis were “coronavirus”, “COVID”, “COVID 19”, “corona”, and “virus”. GTs for these terms in Google Web, News, and YouTube, and the data on COVID-19 case numbers were obtained. Spearman correlation and lag correlation were used to determine the correlation between COVID-19 cases and the Google search terms.

# Summary

“Coronavirus” and “corona” were the terms most commonly used by Internet surfers in India. Correlation for the GTs of the search terms “coronavirus” and “corona” was high ( $r > 0.7$ ) with the daily cumulative and new COVID-19 cases for a lag period ranging from 9 to 21 days. The maximum lag period for predicting COVID-19 cases was found to be with the News search for the term “coronavirus”, with 21 days, i.e., the search volume for “coronavirus” peaked 21 days before the peak number of cases reported by the disease surveillance system.

This study revealed that GTs may predict outbreaks of COVID-19, 2 to 3 weeks earlier than the routine disease surveillance, in India. Google search data may be considered as a supplementary tool in COVID-19 monitoring and planning in India.

Considering the rising menace of coronavirus disease 2019 (COVID-19), it is essential to explore the methods and resources that might predict the case numbers expected and identify the locations of outbreaks. Hence, they have done the following study to explore the potential use of Google Trends (GT) in predicting the COVID-19 outbreak in India.

The Google search terms used for the analysis were “coronavirus”, “COVID”, “COVID 19”, “corona”, and “virus”. GTs for these terms in Google Web, News, and YouTube, and the data on COVID-19 case numbers were obtained. Spearman correlation and lag correlation were used to determine the correlation between COVID-19 cases and the Google search terms.