

## Global Search to Understand the Impact of the scenario of the Keyword

Dr. D Bujji Babu<sup>1</sup>, K. Jayakrishna<sup>2</sup>, Mr.Sk.Anjaneyulu Babu<sup>3</sup>Dr. R. Saran Kumar<sup>4</sup>, Alhaf Malik. K<sup>5</sup>

<sup>123</sup>Department of MCA, <sup>4</sup>Department of ECE<sup>4</sup> QIS College of Engineering&Technology Ongole, AP, India mcahod@qiscet.edu.in, Jayakrishna.k@qiscet.edu.in

<sup>5</sup>Department of Information Technology Skills and Engineering, King Saud University, Riyadh, Saudi Arabia. amohaideen.c@ksu.edu.sa

---

**Abstract.** Due to globalization, the world became a global village, the computer-based search playing a vital role in understanding the scenarios around the globe. Only one solution to explore search engine data is Google Trends, which is at free of cost, and potential to predict the official data immediately. In this paper, we experimented with a keyword-based search to understand various scenarios in different countries. We, analyzed the Google Trends using various techniques like long short-term memory in feed forward networks in Deep Learning, Multi linear regression and Elastic net, SVR and Time Series algorithms for analysis.

**Keywords:** Google Trends, Google search engine, Analysis

---

### 1 INTRODUCTION

Google search engines have major impact on people's daily life. Retrieval of data from search engine is very important. Search engines are act like filters for information available on the internet. They are used to find information as genuine interest without need to walk through irrelevant web pages.

Now a days, only one solution to explore search engine data is Google Trends, which is at free of cost, and potential to predict the official data immediately. It has explained in more indexed papers database.

Search engines provide users search results with relevant information that are available on high quality websites. According to user searches, Search engines deliver results to attain market share in online searches.

Google search is a famous tool for explore the data because it is at free of cost, easy to access and potential to predict the data immediately using google trends data.

The paper was organized with 7 sections. The section-1 describes the Introduction concepts which helps to understand the work. Section-2 focused on Literature Survey, we Sur versed around 26 papers clearly defined the problem statement in section-3 the section methodology web explained. In section 4 the Implementation environment and result are placed, the 5<sup>th</sup> section we presented the conclusion of the work as well as the feature work also mentioned finally in section 6 and 7 expressed the acknowledgment and references.

### 2 LITERATURE SURVEY

In big data research, the data analysis of search engines will be interesting research area.

Google search engine is a tool for searching information in various fields. It deals an immediate response for need of its users. It traffics is voluminous and search engines requests can be tracked [1]. According to this theory [2], As compared to most qualified decision maker, better forecasts can be produced by diverse people.

will produce better and more solid forecasts. As practice shows, the group's guesses mean individual estimation will not be precise than groups guess. The wisdom of the crowd gives us the right answer, it will show, It is good answer than other answers.[3] participant independence, and network devolution. Variety deals with each member's crowd has some personal information and a unique interpretation. Independence means individuals opinions are not decided those around them. Devolution means that crowd member can learn from their sources and their own specializations data as predictive tool for academics and practitioners in various fields. Ginsberg et al [4] noticed these exact features in scientific research. They found that google search engine tracking queries gives the possibility of population's flu established Google Flu Trends, which is epidemic tracking tool.

### **2.1 Google search volume research with data**

Upto 2006, Finally, SVI is an indicator of the flu virus spread launched. An advanced technology i.e., Google Insights launched for providing data of search query is private until google trends latest after two years. GT data of Previous [4][5], have explained that data search in google can help [6] flu spread and rate of unemployment there increase in sales if there is search property of real estate using Google trends in a specific location.

Explored search engine data is Google Trends, which is at free of cost, and potential to predict the official data immediately. It has explained in lot of indexed papers.

Google Trends is a significant tool in housing market, which predict real estate price changes. Due to differences in technology, the reliability cannot be widespread in web and social media.

### **2.2 Google search data**

Google trends filtered data will give users querying volume of aggregated search category, the data collected through google search engine is anonymous and normal. The CSV file of search term index at national and local levels. GT gives 0(zero) index of query if search volume is less for certain amount of time [7].

Previously several studies explored on various online data streams They are articles related to news [8]-[10], websites and blogs of health [11], Wikipedia [13]-[14] search engines [15], and twitter data. Data on twitter provides spation-temporals observations for twitter data. The sets related to infection reflects actual illness accurately [16], false positives and negatives pursuing while doing data collection, preprocessing and classification. Moreover, there is limited usage for twitter [17].

Online tracking of websites, browser attacks and phishing attacks are classified based on substantial literature of the web browsers privacy and security.

Study Of online tracking focused on cookies and related ecosystem, browser fingerprinting used to prevent tracking and interested for long time to preventing tracking using browser IP address. Tor is most prominent technology in this area related to attacks and defences/mitigations.

Timing-based side channels are against browser which leak information such as user browsing history [18]-[20] and website JavaScript is executed against memory-based attacks [21]. [22] focused on user generated data for public health surveillance. [23] suggests twitter is used for public health. Mogo [24] Due to changes in inclinations of technology cultural and economic, we cannot generalize developed countries [25] and [26].

### 3 METHODOLOGY

Algorithms: -

1. Seasonal ARIMA with or without external regressions.
2. MLR means Multiple Linear Regression.
3. EN stands Elastic Net.
4. Support Vector Machine Regression.
5. Feedforward Neural Network (FNN).
6. Long Shot-Term memory (LST).

1. Seasonal ARIMA with or without external regressions.

2. Multiple linear regression

Response variable predictors by multiple regression.

It is an extension of linear regression which used only one explanatory variable.

In the area of econometrics and financial inference, we use MLR

MLR Formula

$$y_i = \beta_0 + \beta_{1x_i} + \beta_{2x_i^2} + \beta_{3x_i^3} + \dots + \beta_{px_i^p}$$

Where, n observation for i value

Dependent variable is  $y_i$

Explanatory variables is  $x_i$

y intercept is  $\beta_0$

Each slope coefficient variable is  $\beta_p$

Models error term is  $\epsilon$

3. Elastic net (EN)

It is a regularized regression method. It combines L1 and L2 penalty functions. Based on the sum of squared coefficient values one penalty is penalized. It is L2 penalty. L2 prevent the coefficient that are removed from the model. L2 minimizes the all-coefficients size.

$$l_2\text{-penalty} = \sum_{j=0}^p \beta_j^2$$

L1 penalty is sum of absolute coefficient values. L1 allow some coefficient to zero. L1 removes predictor from the model.

- $l_1\text{-penalty} = \sum_{j=0}^p \text{abs}(\beta_j)$

It includes L1 and L2 penalties during training and it is penalized linear regression model.

4. Support Vector Regression

It is used to predict discrete values. It works as similar to Support Vector machine principle. Finding the best-fit line is main objective of SVR. SVR is best-fit line that has the maximum number of points.

$$y = wx + b \text{ is hyperplane.}$$

5. Feedforwards Neural Networks

It approximate function.

$$y = f^*(X) \text{ is a formula for classifier assigns X to Y.}$$

$\theta$  FNN will map  $y = f(x; \theta)$ .  $\theta$  will memorize which is appropriate function.

6. Long short-term memory

In AI and deep learning areas, we use this artificial neural network method i.e., LSTM. It is based short-term memory processes to create longer-term memory. LSTMs a complex area of deep learning.

#### 4 IMPLEMENTATION & RESULT

For the implementation we directly used the pandas, pytrends in the Jupiter note book in windows 10 operating system we executed on the machine is the configuration of RAM is 12.0 GB local disk (c) 145 GB, we executed on high end configuration system with 12GB RAM and 145 GB hard disk.

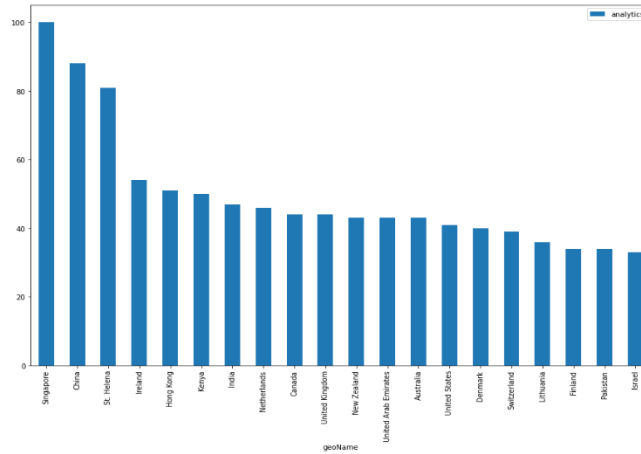


Fig: 1. Geo Name Analytics

Fig 1: Specifies that analytics has been the focus of so many companies and students for the last 3 to 4 years, so let’s have a look at the trend of search queries based on “Analytics” increased or decreased on Google.

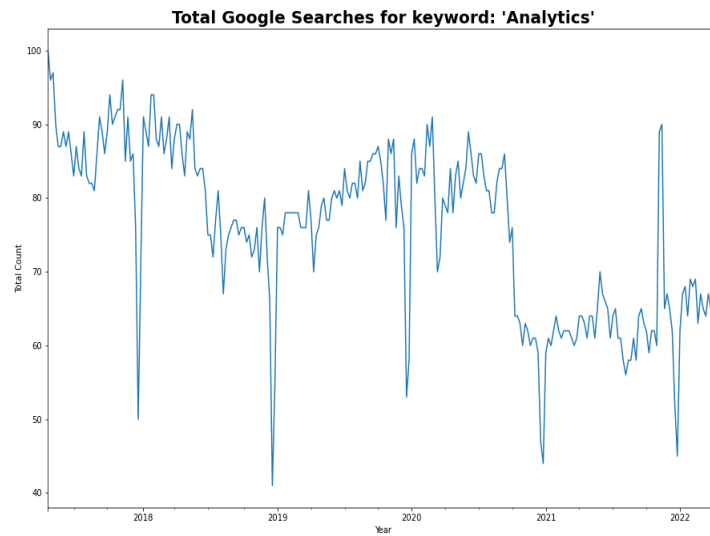


Fig:2. Total Google Searches for Keyword: ‘ Analytics ’

Fig 2: In graph X-axis represented by years and Y-axis represented by total count of Google Searches for Keyword Analytics. So we can see that searches based “ analytics” Google started to decrease in 2018 and highest searches were done in 2022. This is how we can analyze Google searches based on any keyword.

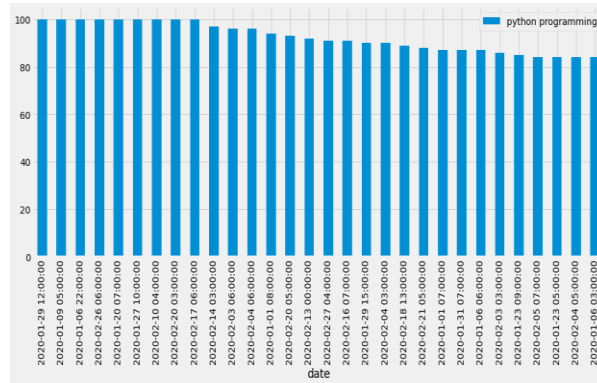


Fig: 3. Python Programming

Fig 3: In this figure shows date by the date dimension and getting the sum of clicks for each of them, it's a type of summrization.

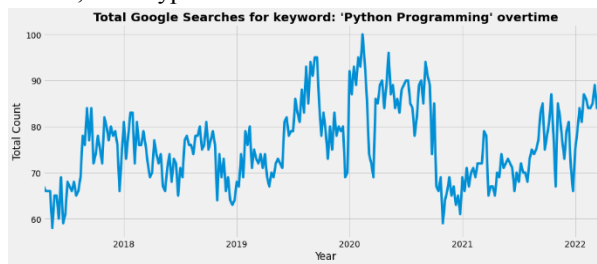


Fig: 4. Total Google Searches for Keyword: 'Python Programming' overtime

Fig 4: In this figure X-axis shows over the period of five years, it has stayed pretty stable with the highest peak around 2020. Y-axis shows over the total count.

## 5 CONCLUSION

The methodology of this research data of was tested on data about user's activity in Twitter and Reddit and also data other than real estate. This will give next stage of development and improvement

Google Trends play vital role in understanding of housing market. Google trends is a significant tool for predict real estate price changes.

Online users of the web predict price trends in housing market. Government reports is crucial for understanding of current housing market and trends of future market forecasting.

Google search data is more reliable for monitoring flu spread and produce comparison of forecasting accuracy to data of real-life ILI. The models forecasting capability enhanced the GT and ILI data is freely available without any delay can be used for addressing problems associated with traditional systems. To achieve sustainable surveillance, It will allow for preparation of better epidemic.

## 6 ACKNOWLEDGMENT

We acknowledge all people who are directly and indirectly motivate us to complete this task. We sincerely thank all the authors and co-authors of the papers which are mentioned in references section.

## 7 REFERENCES

1. I. Bordion, S. Battistion, G. Caldarelli, M. Cristelli, A. Ukkonen, and I. Weber, "Web search queries can predict stock market volume," PLoS ONE, vol. 7, no. 7, Jul. 2012, Art. No. e40014, doi: 10.1371/journal.pone.0040014.
2. J. Surowiecki, *The Wisdom of Crowds*. New York, NY, USA: Anchor Book, 2005.
3. H. Hong, Q. Ye, Q. Du, G. A. Wang, and W. Fan, "Crowd characteristics and crowd wisdom: Evidence from an online investment community," J. Assoc. Inf. Sci. Technol., vol. 71, no. 4, pp. 423-435, Apr. 2020, doi: 10.1002/asi.24255.
4. J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M.S. Somlinski, and L. Brilliant, "Detecting influenza epidemics using search engine Query data," Nature, vol. 457, no. 7232, pp. 1012-1014, Feb. 2009.
5. H. Choi and H. Varian, *Predicting Initial claims for Unemployment Benefits*. Menlo Park, CA, USA: Google, 2012.
6. H. Choi and H. Varian, "Predicting the present with Google trends," Econ. Rec., Vol. 88, no.1, pp. 2-9, 2012.
7. Google Trends: Understanding the Data. Accessed: Aug. 1, 2019. [Online]. Available: <http://storage.googleapis.com/gwebnewsinitiativetraining.appspot.com/upload/GO802NewsInitiativeLessonsFundamentals-L04-GoogleTrends1saYVCP.pdf> B. Liu, *Sentiment Analysis and Opinion Mining*, vol. 5. San Rafael, Ca, USA: Morgan & Claypool, 2012.
8. R. Grishman, S. Huttunen, and R. Yang Arber, "Information extraction for enhanced access to disease outbreak reports," J. Biomed. Information, vol. 35, no. 4, pp. 236-246, 2002, doi: 10.1016/S1532-0464(03)00013-3.
9. M. Abla and M. Blench, "Global public health intelligence network (GPHIN)," in Proc. 7<sup>th</sup> Conf. Assoc. Mach. Transl. Amer., 2006, pp. 8-12. Accessed: Mar. 9, 2018. [Online]. Available: <http://pdfs.semanticscholar.org/7d88/e623aa6ca78510e0093e17e2e00db39bdad5.pdf> A.
10. A. R. Reilly, E.A. Iarocci, C. M. Jung, D. M. Hartley, and N. P. Nelson, "Indications and warning of pandemic influenza compare to seasonal influenza," Inf. Syst., vol. b9, no. 8, p. 2008, 2008.
11. A. Hulth, G. Rydevika, and A. Linde, "Web queries as a source for syndromic surveillance," PLoS ONE, vol. 4, no. 2, p.e4378, Feb.2009, doi: 10.1371/journal.pone.0004378.
12. D. Mciver and J. S. Brownstein. *Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in Near Real-Time*. Accessed: Jun. 21, 2017. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4512362/pdf/ojphi-07-e40.pdf>.
13. B. Bardak and M. Tan, "Prediction of influenza outbreaks by integrating Wikipedia article access logs and Google flu trend data," in proc. IEEE 15<sup>th</sup> BIBE, Nov. 2015, pp. 1-6, and doi: 10.1109/BIBE.2015.7367640.
14. K. S. Hickmann, G. Fairchild, R. Priedhorsky, N. Generous, J. M. Hyman, A. Deshpande, and S. Y. D. Valle, "Forecasting the 2013-2014 influenza season using Wikipedia," PLoS Comput. Biol., vol. 11, no. 5, May 2015, Art. No. e1004239, doi:10.1372/journal.pcbi.1004239.
15. G. Eysenbach, "Infodemiology: Tracking flu-related searches on the web for syndromic surveillance," in proc AMIA annu. Symp., 2006, pp. 244-248.
16. P. M. Polgreen, Y. Chen, D. M. Pennock, F.D. Nelson, and R. A. Weinstein, "Using internet searches for influenza surveillance," Clin. Infectious Diseases, vol. 47, no. 11, pp. 1443-, Dec. 2008, doi: 10.1086/5939098.
17. R. Moss, A. Zarebski, P. Dawson, and J. M. McCaw, "Forecasting influenza outbreak dynamics in Melbourne from internet search query surveillance data," Influenza Other Respiratory Viruses, vol. 10, no. 4, pp. 314-323, Jul. 2016, doi: 10.1111/irv.12376.

18. C. Pelat, C. Turbelin, A. Bar-Hen, A. Flahault, and A. J. Valleron, "More diseases tracked by using Google trends." *Emerg Infectious*, vol. 15, no. 8, ppp. 1327-1328, 2009. doi: 10.320/eid1508.090299.
19. M. Schootman, A. Toor, P. Cavazos-Rehg, D. B. Jeffe, A. McQueen, J. Eberth, and N.O. Davidson, "The utility of Google trends data to examine interest cancer screening," *BMJ Open* vol. 5, no. 6, Jun.2015, Art. No. e006678, doi: 10.1136/bmjopen-2014-006678.
20. Y. Teng, D. Bi, G. Xie, Y. Jin, Y. Huang, B. Lin, X. An, D. Feng, and Y. Tong, "Dynamic forecasting of Zika epidemics using Google trends," *PLoS ONE*, vol. 12, no. 1, Jan. 2017, Art. no. e0165085, doi: 10.1371/journal. Pone. 0165085
21. C. Li, L. J. Chen, X. Chen, M. Zhang, C. P. Pang, and H. Chen, "Retrospective analysis of the possibility of predicting the COVID-19 outbreak from internet searches and social media data, China, 2020," *Euro surveillance*, vol. 25, no. 10, mar. 2020, Art, no. 2000199, doi: 10.2807/1560-7917. ES.2020. 25.10. 2000199.
22. Z. Shakeri Hossein Abab, A. Kline, M. Sultana, M. Noaeen, E. Nurmambetiva, F. Lucini, M. Al-Jefri, and J. Lee, "Digital public health surveillance: A systematic scoping review," *NPJ Digital. Med.*, vol. 4, no.1, pp. 1-13, Dec. 2021, doi: 10. 1038/s41746-021-00407-6.
23. O. edo-Osagie, B. De La Iglesia, I. Lake, and O. edeghere, "A scoping review of the use of Twitter for public health research," *Comput. Biol. Med.*, vol. 122, Jul. 2020, Arr. no. 103770, doi: 10. 1016/j.compbimed.2020. 103770.
24. E. Mogo. Social media as a public Health Surveillance Tool: evidence and Prospects. Accessed: Nov. 2, 2018. [Online]. Available:[http://www.enterprise.sickweather.com/downloads/SWSocialMedia\\_Whitepaper.pdf](http://www.enterprise.sickweather.com/downloads/SWSocialMedia_Whitepaper.pdf)
25. E. O. Nsoesie, O. Olaseni, A. S. A. Abah, and M. L. Ndeffo-Mbah, "Forecasting influenza-like illness trends in Cameroon using Google search data," *sci. rep.*, vol. 11, no. 1, p. 6713, Dec. 2021, doi: 10. 1038/s41598- 021-85987-9.
26. N. L. Bragazzi and N. Mahroum, "Google trends predicts present and future plague cases during the plague outbreak in madagascar: Infodemiological study," *JMIIR Public Health Surveill.*, vol. 5, no. 1, Mar. 2019. Art. no. e13142, doi: 10.2196/13142.
27. S. Jafar Ali Ibrahim and M. Thangamani. 2018. Momentous Innovations in the Prospective Method of Drug Development. In *Proceedings of the 2018 International Conference on Digital Medicine and Image Processing (DMIP '18)*. *Association for Computing Machinery, New York, NY, USA*, 37–41. <https://doi.org/10.1145/3299852.3299854>
28. B. Karthikeyan, K. Alhaf Malik, D. Bujji Babbu, K. Nithya, S. Jafar Ali Ibrahim, N. S. Kalyan Chakravarthy (2021) "Survey of Cooperative Routing Algorithms in Wireless Sensor Networks", *Annals of the Romanian Society for Cell Biology*, pp. 5316–5320. Available at: <https://www.annalsofrscb.ro/index.php/journal/article/view/702>