Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20250403007
Volume 4, Issue3 (September-December 2025), PP: 34-38.
www.indjcst.com
ISSN No: 2583-5300

Survey Paper on Google Search Analysis Using Machine Learning

Vaastav L Sanghvi[1], Vaseem Akram B[2], Pratham Chaithanya K A[3]
[1,2,3] *Department of Computer Science Engineering in Data Science, Dayananda Sagar Academy of Technology and Management, Bengaluru, Karnataka, India.*

To Cite this Article: *Vaastav L Sanghvi[1], Vaseem Akram B[2], Pratham Chaithanya K A[3], "Survey Paper on Google Search Analysis Using Machine Learning", Indian Journal of Computer Science and Technology, Volume 04, Issue 03 (September-December 2025), PP: 34-38.*

Abstract: *In a world increasingly influenced by online activity, analyzing search behaviour has become an essential tool for understanding public interest and predicting emerging trends. This project, titled "Google Search Analysis Using Machine Learning," aims to forecast keyword popularity over time using data sourced from Google Trends. During Phase 1, our team focused on conceptualizing the approach, finalizing the methodology, and outlining the system architecture. The planned system begins by defining the scope and selecting appropriate keywords. After that, it links to the Google Trends API to get search data that is useful. This data is cleaned up and pre-processed to make sure it is consistent and of good quality. Next, trend analysis is performed to uncover underlying patterns and seasonal behaviour. Using this processed data, machine learning models— such as ARIMA or LSTM—will be trained to make future predictions. The output is then visualized using charts and confidence intervals to offer intuitive insights. Finally, the system will generate a comprehensive report summarizing the findings and providing actionable recommendations. This structured workflow lays the foundation for an intelligent, automated tool capable of tracking and forecasting online keyword trends, thereby assisting in data-driven decision-making.*

Key Words: *ARIMA, Data Visualization, Google Trends, Keyword Analysis, LSTM, Machine Learning, pytrends, Time Series Forecasting.*

I.INTRODUCTION

Search engines are at the core of digital information access, with billions of users relying on platforms like Google to satisfy their information needs daily. Analyzing search trends provides valuable insights into public interest, behaviour patterns, and emerging topics. This project, titled "Google Search Analysis Using Machine Learning," focuses on analyzing and forecasting keyword search trends using data sourced from Google Trends.

The objective is to collect keyword-specific search interest data over time, identify notable patterns, and apply machine learning models to predict future popularity. Additionally, the project incorporates user-level data collection via Google Forms to better understand how individuals' structure and execute search queries. This dual-source approach allows for both system-driven analysis and human-centered insight.

The project employs tools such as the Google Trends API, Python-based preprocessing, and forecasting models including ARIMA, Prophet, and LSTM. Data visualization is achieved using Power BI, Matplotlib, and Seaborn, facilitating clear presentation of trend evolution and comparative analysis across different timeframes and regions. The resulting system aims to support strategic, data-informed decisions by transforming raw search data into meaningful trends and forecasts.

II.LITERATURE SURVEY

"Exploring the Power of Google Trends: Applications, Limitations, and Future Directions" Abu Rayhan, CBECL, 2024"

This paper explores how Google Trends has found diverse applications across fields like public health, consumer marketing, academic research, and political analysis. It presents a holistic view of how search query data reflects public attention and interest over time and across regions.

Through detailed case studies, the paper examines the tool's role in tasks such as tracking disease outbreaks, analyzing consumer sentiment, and studying social behaviors. It investigates both the strengths and inconsistencies of search data and how they influence research outcomes.

While Google Trends can be incredibly useful for early insights, its reliability depends on cautious interpretation. The tool is best used alongside other data sources due to limitations like lack of transparency in data processing and occasional variability in outputs

"Reliability of Google Trends: Analysis of the Limits and Potential of Web Infoveillance During COVID-19" Alessandro Rovetta, Mensana srls and Redeev srl, 2021.

This study focuses on how reliable Google Trends data was during the COVID-19 pandemic, particularly for terms related to the virus and health behaviour.

Repeated queries for the same terms were analyzed over different days, and statistical tests were used to measure the inconsistency and variability of search interest scores.

The findings showed noticeable fluctuations in data depending on when it was accessed. The study recommends averaging results over multiple days to enhance stability, making Google Trends more dependable for real-time health tracking.

"Predicting the Present with Google Trends" Hyun Young Choi and Hal Varian, Google Inc., 2012.

This paper investigates the use of search queries related to flu symptoms to detect outbreaks before they appear in official health data.

Google search trends for flu-related terms were compared with CDC flu surveillance data to identify patterns and timing.

The results revealed that Google Trends often showed spikes in flu-related searches up to two weeks before official case counts increased, highlighting its potential as an early alert system.

"Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends" Ulrike Vosen and Timo Schmidt, RWI – Leibniz Institute for Economic Research, 2011.

The study explores whether Google Trends can be a substitute or enhancement for traditional survey-based methods in predicting consumer spending.

By analysing the frequency of searches for terms like “new car” or “vacation deals,” researchers-built consumption models and compared them to survey-based predictions.

Google Trends provided faster signals, and in times of economic volatility, it often yielded more accurate results than traditional consumer sentiment surveys.

"The Utility of Google Trends for Epidemiological Research: Lyme Disease as an Example" Aaron Seifter, Adam Schwarzwalder, Kurt Geis, and John Aucott, Johns Hopkins University, 2010.

This paper examines whether Google search activity related to Lyme disease can serve as an indicator of actual disease incidence in the U.S.

Researchers compared state-by-state search volume data for Lyme disease with official CDC reports.

A strong correlation was found, suggesting that search interest could be a useful supplementary tool for monitoring and perhaps even predicting disease spread, especially in underreported regions

"Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks" Hilton A. Carneiro and Eleftherios Mylonakis, Warren Alpert Medical School of Brown University, 2009.

This study assesses the use of Google Trends in monitoring real-time outbreaks like influenza and foodborne illnesses. Search volumes for disease-related terms were mapped against known outbreak timelines and public health reports.

The tool showed early spikes in interest before formal outbreak declarations. However, media coverage also influenced search behaviour, so researchers stress it should be used in tandem with verified data.

"A Deep Learning Framework for Understanding Search Behavior from Google Trends Data" Minh Nhat Tran, Lin Luo, and Sajal K. Das, IEEE Transactions on Big Data, 2021.

The study applies deep learning to Google Trends data to better understand how people search for information over time. Neural networks were trained on historical search queries to uncover hidden patterns in temporal and contextual behavior.

The model offered new insights into user intent and evolving information needs, potentially informing fields like public health, education, and marketing.

"Forecasting COVID-19 Cases Using Google Trends and Machine Learning: A Hybrid Model Approach" Yun Liu, Qian Zhao, and Jing Zhang, Scientific Reports, 2021.

This paper combines Google Trends data with machine learning to predict COVID-19 case numbers more accurately.

A hybrid model was developed using search terms like “loss of taste” and “COVID test near me,” paired with official health data. The integrated model significantly improved forecasting accuracy and provided a promising approach for pandemic preparedness and response

"Enhancing Feature Selection Using Google Trends Data to Predict Influenza-like Illness" Artemis Lampos, Elad Yom-Tov, Yoram Cox, and Ian Thapen, Journal of Biomedical Informatics, 2014.

This work examines how Google Trends data can be used to improve the features chosen for flu prediction models.

Researchers identified and tested search terms most relevant to influenza-like illness (ILI), refining machine learning inputs for better performance.

Incorporating search-based features improved model accuracy, supporting the role of Google Trends in digital disease surveillance.

"Quantifying the Advantage of Looking Forward Using Google Trends" Tobias Preis, Helen Susannah Moat, H. Eugene Stanley, and Steven R. Bishop, Scientific Reports, 2012.

The study proposes a way to measure how future-focused a population is by analyzing searches for future years.

Researchers created a "future orientation index" by tracking searches like "2023" or "next year" and comparing these to national economic metrics.

Countries with higher future-oriented search behavior tended to have stronger economic indicators, suggesting a link between collective mindset and prosperity.

"Characterizing the Influence of Domain Expertise on Web Search Behavior" Ryen W. White, Susan T. Dumais, and Jaime Teevan, Proceedings of the 2nd ACM International Conference on Web Search and Data Mining (WSDM), 2009.

This study explores how search behavior differs between experts and novices in various fields

By analyzing search logs, the researchers assessed differences in query formulation and navigation efficiency.

Experts used more precise and efficient strategies, offering insights into how search engines could tailor results based on user expertise

"Affective and Content Analysis of Online Depression Communities" Thin Nguyen, Dinh Phung, Bo Dao, Svetha Venkatesh, and Michael Berk, IEEE Transactions on Affective Computing, 2014.

The paper analyzes emotional patterns in online communities where people discuss depression and mental health.

Using machine learning and natural language processing, the study examined user posts to detect affective signals and linguistic trends.

The analysis uncovered distinctive emotional markers and conversation themes that could aid in early detection or intervention in mental health care.

"The Parable of Google Flu: Traps in Big Data Analysis" David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani, Science, 2014.

This critical paper discusses how Google Flu Trends, once a highly praised forecasting tool, eventually failed due to methodological flaws.

The authors examined how overfitting, data drift, and lack of algorithmic transparency led to major inaccuracies.

The failure of Google Flu serves as a cautionary tale, emphasizing the importance of validating models and integrating big data with traditional methods for public health surveillance**.**

"An Approach to Web Search Behavior Modeling Using Machine Learning and Large-Scale Search Engine Data" David Garcia-Gasulla, Ferran Parés, Eduard Ayguadé, Jesús Labarta, Ulises Cortés, and Toyotaro Suzumura, Information Processing & Management, 2019.

This study uses large-scale search engine data to model and predict web search behavior.

Machine learning algorithms were applied to massive datasets to uncover user patterns, query evolution, and behavioural trends. The results contributed to improving search engine design and personalization, offering better user experience through smarter, behaviour-aware algorithms.

III.METHODS

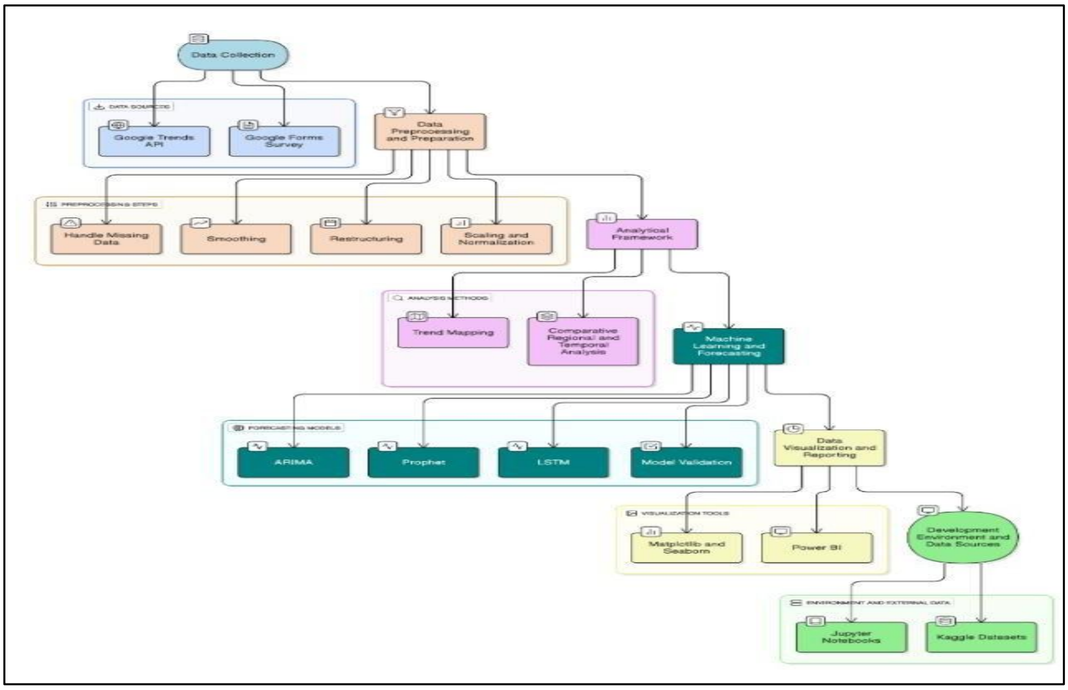

Figure 1: Detailed Data Flow Diagram

A. Data Collection via Google Trends API and Google Forms

The foundation of this project's data acquisition lies in the use of the Google Trends API, accessed through the Python pytrends library. This approach enables programmatic retrieval of search interest data for specified keywords, with configurable parameters such as geographic location, time range, search category, and data granularity. The normalized data values, ranging from 0 to 100, represent relative search volume over time and serve as the primary dataset for subsequent analysis and forecasting [1], [3], [4], [6]. To augment the quantitative data and capture user behaviour more comprehensively, a structured Google Form survey was distributed among a targeted group of users. The survey aimed to gather insights on search query phrasing, frequency of search engine usage across various domains (e.g., education, health, commerce, entertainment), and perceptions of trending topics. The anonymized responses were systematically analysed to identify patterns in human search behaviour, thereby providing contextual depth to the automated search data [1], [2].

B. Data Preprocessing and Preparation

Raw data obtained from the Google Trends API often requires preprocessing to address issues such as missing entries, inconsistent time intervals, and transient fluctuations that can obscure underlying trends. To ensure data quality and enhance model performance, preprocessing steps were conducted using *Pandas* and *NumPy*, including interpolation and forward-filling of missing data points, application of rolling averages to smooth volatility, restructuring data into time-indexed series, and performing scaling and normalization. These procedures are critical to maintaining data integrity and improving the reliability of subsequent forecasting models [2], [5].

C. Analytical Framework

The analytical component of the project is centered on two main approaches:

- **Trend Mapping:** This involves visualizing and interpreting temporal changes in keyword search interest, enabling the identification of seasonal patterns, long-term trends, periodic spikes, and plateaus. Such mapping provides valuable insights into public behaviour and event-driven attention dynamics [1], [4], [6].
- **Comparative Regional and Temporal Analysis:** By examining search trends across multiple timeframes (monthly, quarterly, yearly) and geographical scopes (local, national, global), the system identifies variations in search behaviour that might be region-specific or time-dependent. This comparative analysis is instrumental for informing localized marketing strategies, policy interventions, and targeted awareness campaigns [1], [4], [7].

D. Machine Learning and Forecasting Techniques

To forecast future keyword popularity, a combination of statistical and deep learning methodologies was employed:

- **ARIMA (Autoregressive Integrated Moving Average):** Suited for modelling linear and stationary time-series data, ARIMA captures temporal dependencies and autocorrelations, with performance evaluated via RMSE and MAE metrics [5].
- **Prophet:** Developed by Meta, Prophet excels in handling time series with pronounced seasonal effects and trend shifts, and it is robust against missing data and outliers [1].
- **LSTM (Long Short-Term Memory Networks):** These recurrent neural networks are capable of learning complex, nonlinear patterns and long-term dependencies in sequential data, making them well-suited for forecasting on noisy and intricate datasets.

Model training was performed using historical Google Trends data, with validation conducted on hold-out samples to ensure predictive accuracy. Development and tuning utilized libraries such as *Scikit-learn*, *TensorFlow*, and *Statsmodels* [1], [3].

E. Data Visualization and Reporting

Clear and effective visualization of results is paramount for stakeholder communication. The project utilized:

- **Matplotlib** and **Seaborn** to generate comprehensive static visualizations including time series plots, trend overlays, correlation heatmaps, and seasonal decomposition charts [1], [5].
- **Power BI** for the creation of interactive dashboards, enabling dynamic exploration of keyword trends. Features include time-based filters, geographic heat maps illustrating regional interest, and comparative bar graphs and trend lines across multiple keywords. These dashboards are designed to facilitate data-driven decision-making for both technical and non- technical users.

F. Development Environment and Data Sources.

The development process was primarily conducted within **Jupyter Notebooks**, providing an environment conducive to modular coding, iterative testing, and thorough documentation. Additionally, **Kaggle datasets** were utilized as complementary data sources to validate trends and benchmark forecasting models against alternative datasets, when necessary [1], [3].

IV.CONCLUSION

This work explores how various tools and techniques could be brought together to make the most of Google Trends data for meaningful analysis and forecasting. The idea is to combine online search trends with user feedback to better understand what people are looking for and why.

Before any modeling can happen, the data would need to be cleaned and organized using tools like Pandas and NumPy. Forecasting methods such as ARIMA, Prophet, and LSTM are being considered to capture both simple and complex patterns in the data.

To make the insights easy to understand and share, visualization tools like Matplotlib, Seaborn, and Power BI are planned.

Platforms like Jupyter Notebooks and Kaggle could also support a more transparent and reproducible workflow.
In short, this survey outlines a thoughtful approach to using Google Trends, highlighting both its potential and the care needed to use it effectively.

References

1. Abu Rayhan, "Exploring the Power of Google Trends: Applications, Limitations, and Future Directions," CBECL, 2024.
2. Alessandro Rovetta, "Reliability of Google Trends: Analysis of the Limits and Potential of Web Infoveillance During COVID-19," Mensana srls and Redeev srl, 2021.
3. Hyun young Choi and Hal R. Varian, "Predicting the Present with Google Trends," Google Inc., 2012.
4. Jeremy Ginsberg, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data," Google.org and CDC, 2009.
5. Simeon Vosen and Torsten Schmidt, "Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends," RWI – Leibniz Institute for Economic Research, 2011.
6. Ari Seifter, Alison Schwarzwalder, Kate Geis, and John Aucott, "The Utility of Google Trends for Epidemiological Research: Lyme Disease as an Example," Johns Hopkins University, 2010.
7. Herman Anthony Carneiro and Eleftherios Mylonakis, "Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks," Warren Alpert Medical School of Brown University, 2009.
8. Minh Nhat Tran, Lin Luo, and Sajal K. Das, "A Deep Learning Framework for Understanding Search Behavior from Google Trends Data," IEEE Transactions on Big Data, vol. 7, no. 2, pp. 389–402, 2021.
9. Yun Liu, Qian Zhao, and Jing Zhang, "Forecasting COVID-19 Cases Using Google Trends and Machine Learning: A Hybrid Model Approach," Scientific Reports, vol. 11, no. 1, 2021.
10. Artemis Lampos, Elad Yom-Tov, Yoram Cox, and Ian Thapen, "Enhancing Feature Selection Using Google Trends Data to Predict Influenza-like Illness,"Journal of Biomedical Informatics, vol. 49, pp. 148–158, 2014.
11. David Preis, Helen Susannah Moat, H. Eugene Stanley, and Steven R. Bishop, "Quantifying the Advantage of Looking Forward Using Google Trends," Scientific Reports, vol. 2, 2012.
12. Ryen W. White, Patrick Pantel, Susan T. Dumais, Jaime Teevan, and Yandong Wang, "Characterizing the Influence of Domain Expertise on Web Search Behavior," in Proceedings of the 2nd ACM International Conference on Web Search and Data Mining (WSDM), 2009, pp. 132–141.
13. Anh Tuan Nguyen, Dinh Phung, Svetha Venkatesh, and Michael Berk, "Affective and Content Analysis of Online Depression Communities," IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 217–226, 2014.
14. David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani, "The Parable of Google Flu: Traps in Big Data Analysis," Science, vol. 343, no. 6176,pp. 1203–1205, 2014.
15. David Garcia-Gasulla, Ferran Parés, Eduard Ayguadé, Jesús Labarta, Ulises Cortés, and Toyotaro Suzumura, "An Approach to Web Search Behavior Modeling Using Machine Learning and Large-Scale Search Engine Data," Information Processing & Management, vol. 56, no. 3, 2019.