Department of Computer Science
The University of California, Davis
Davis, CA, USA
☎ *+1 (484) 743 8374*
✉ *yvekaria@ucdavis.edu*
yash-vekaria.github.io
yash-vekaria
vekariayash

Yash Vekaria

Research Interests

Web Security & Privacy; Digital Advertising Ecosystem; Ad-targeting; Ad-fraud; Large-scale Internet Measurements; Web Tracking; Applied Machine Learning.

Education

2021–Present **Ph.D. in Computer Science** *4.00/4.00*
University of California, Davis

2016–20 **B.Tech. in Computer Science** *9.04/10.00*
The LNM Institute of Information Technology (LNMIIT), India.

Experiences

Sep 2021 – Present **University of California, Davis** **Davis, USA**
Research Assistant, *Gunrock Breakerspace Lab*,
Advisor: Prof. Zubair Shafiq
Battling Ad-fraud: Research focus is towards investigating problematic ad practices followed by ad-tech entities that lead to wide-spread ad-fraud. Specifically, I am working to build a system to measure compliance and improve the ad-tech standards, click-fraud practices and brand-safety issues. In the recent work, we built ads.txt and sellers.json crawlers, automated network measurement experiments to collect HAR files upon visits to misinformation websites and analyzed the data to study issues in the ad-tech supply chain.

Summer 2022 **Max Planck Institute für Informatics** **Saarbrücken, Germany**
Visiting Research Fellow, *Internet Architecture Group (INET)*,
Advisor: Prof. Savvas Zannettou
Problematic Ads on Social Media: This is an ongoing research project focused on detecting and studying the problematic ads run on Twitter through an automated audit using NLP techniques to understand the compliance of Twitter ad policies, problematic advertiser practices and effectiveness of Twitter's moderation strategies.

Jan 2020 – Aug 2021 **Innovaccer** **Noida, India**
Data Scientist
- Worked on *healthcare data analytics*, dashboard building, and data insights' generation for 2 USA clientele ACO projects: *Orlando Health* (Florida) and *Dignity Health* (California).
- Performed *sentiment analysis* on Yelp reviews to understand customer pain points.
- Carried out *statistical data analysis* for reducing unnecessary Imaging on patients.
- Built numerous *automations* to reduce manual human efforts for example a tool to convert measure definition JSON files to SQL queries.
- Developed a LinkedIn-based network tool for the marketing team to generate effective leads by leveraging network information.

Summer 2019 **Innovaccer** **Noida, India**
Decision Science Intern
- *Optimized* Innovaccer's data model by improving PostgreSQL queries, thereby reducing the overall execution time from 10 hours to 3 hours.
- Performed patient revisit prediction using *Classification and Regression Tree* (CART) based model.
- Scripted *automated* tree-based visualization of enormous measure definition JSON files using *NetworkX* Python library.

Dec 2017 **National Ilan University** **Yilan, Taiwan**
Student-Exchange Intern
A student exchange program to learn concepts related to ML and brainstorm ideas.

Teaching

Spring 2022 **Computer Networks (ECS 152A)** *Teaching Assistant*
University of California, Davis, USA

Spring 2019 **Data Structures & Algorithms** *Teaching Assistant*
The LNM Institute of Information Technology, India

Fall 2018 **Computer Organization & Architecture** *Teaching Assistant*
The LNM Institute of Information Technology, India

Fall 2017 **Computer Programming** *Teaching Assistant*
The LNM Institute of Information Technology, India

Publications

arXiv 2022 **The Inventory is Dark and Full of Misinformation: Understanding the Abuse of Ad Inventory Pooling in the Ad-Tech Supply Chain**
Yash Vekaria, Rishab Nithyanand, Zubair Shafiq
Under Submission, 2022

WebSci 2021 **Differential Tracking Across Topical Webpages of Indian News Media**
Yash Vekaria*, Vibhor Agarwal*, Pushkal Agarwal, Sangeeta Mahapatra, Sakthi Balan Muthiah, Nishanth Sastry, Nicolas Kourtellis (**first co-authors*)
13th ACM Web Science Conference, 2021

ICWSM 2021 **Under the Spotlight: Web Tracking in Indian Partisan News Websites**
Vibhor Agarwal*, **Yash Vekaria***, Pushkal Agarwal, Sangeeta Mahapatra, Sakthi Balan Muthiah, Nishanth Sastry, Nicolas Kourtellis (**first co-authors*)
15th International AAAI Conference on Web and Social Media (ICWSM), 2021
Acceptance rate: 22.4% (30/134)

RCIS 2021 **A Metadata-based Event Detection Method using Temporal Herding Factor and Social Synchrony on Twitter Data**
Nirmal Kumar Sivaraman, Vibhor Agarwal, **Yash Vekaria**, Sakthi Balan Muthiah
15th International Conference on Research Challenges in Information Science (RCIS), 2021

SocInfo 2019 **Poster Invitation: Event Detection in Twitter**
Nirmal Kumar Sivaraman, Vibhor Agarwal, **Yash Vekaria**, Sakthi Balan Muthiah
11th International Conference on Social Informatics (SocInfo), 2019

Skills

General *Network Measurements, Computational Social Science, Data Science, Data Analytics, Data Visualization, Machine Learning, Data & Text Processing, Natural Language Processing, Image Processing, Web Crawling*

Languages *Python, C, Python Libraries (Pandas, Numpy, Matplotlib, Scipy, OpenCV, Selenium, BeautifulSoup, Multiprocessing), SQL/PostgreSQL*

Tools *Docker, Git/GitHub, LaTeX*

Projects

Oct 2022 – Present **Ad-Inspector**
A Large-scale Vulnerability Notification Study for Ad Ecosystem
This ongoing research project is aimed towards building a system which will perform regular crawls of Top 1 Million websites on the Internet and study prevalence of various issues in the ad-tech ecosystem pertaining to ads.txt, sellers.json, pooling, brand-safety, and malvertisements. It will adopt different notification strategies to inform responsible authorities repetitively as a part of the longitudinal study to mitigate ad-fraud and understand responsiveness of the ad ecosystem.

Jun 2021 – Jul 2021 **Sentiment Analysis on Yelp Reviews**
Understanding Patient Pain Points
Yelp reviews of a physicians Yelp page were scraped using a web crawler to perform sentiment analysis. Next, Topic modeling was performed to extract issue that led patient to put up a negative review. Issues could be like staff was unprofessional, physician was late, etc.

Jan 2020 – Feb 2020 **Reducing Unnecessary Imaging**
Statistical Data Analysis to reduce unnecessary $ expenditure on Imaging
Goal was to identify PCPs within an ACO who were not practising value-based care and suggesting unnecessary imaging to their patients. Skewness, Kurtosis, and Chi-square tests were used to identify statistically deviant PCPs and specific diagnosis codes for which they were suggesting unnecessary imaging.

Aug 2019 – Dec 2019 **NBA Accreditation Report Generator**
Developed a Web App for College Teaching Fraternity
A 4-membered team project towards building an app that automatically generates a report for accreditation authorities for all the offered courses. Its main features included: Automatic CIF and student marks scanning to extract useful information, automated feedback form and question paper generation, etc.

Jan 2018 – Aug 2018 **Suspicious Activity Recognition in ATM**
Predicting an ATM activity as single normal/abnormal or multiple normal/abnormal
Using self-generated dataset covering various usecases of ATM, I trained an ML Model. Different stages of the model involved: Frame generation from a video, background subtraction, MHI generation, Hu Moment computation to generate features, followed by training and prediction to classify an activity into one of the 4 classes using SVM and KNN.

Aug 2017 – Oct 2017 **Text Document Ranking and Summarization**
Applied Information Retrieval & Natural Language Processing Techniques to Text
Various text pre-processing techniques like case-folding, tokenization, stop-word removal, stemming, etc were applied on auto-scrapped text documents. Key-term extraction was done to rank these documents. Finally, text summarization was carried out to produce a brief summary of the text documents using similarity metrics like Cosine similarity, Jaccard coefficient and TF-IDF.

Honors & Awards

2016–2018 **Merit Scholarships** *The LNM Institute of Information Technology, India*
Awarded to Top 5% students with GPA greater than 9.

2019, 2021 **Spot Awards** *Innovaccer, India*
Two Quarterly Spot Awards received: One for efficient optimization of data model from 10 hrs to 3 hrs. Second for automating the process of running measures.

Travel Grants

2022 *22nd Privacy Enhancing Technologies Symposium (PETS)*

2022 *15th ACM Conference on Security & Privacy in Wireless & Mobile Networks (WiSec)*

2021 *42nd IEEE Symposium on Security & Privacy (S&P)*

2021 *34th IEEE Computer Security Foundations Symposium (CSF)*

Invited Talks

2021 **Differential Tracking Across Topical Webpages of Indian News Media**
The LNM Institute of Information Technology, India

Selected Media Coverage

Mar 2021 **Tracking the Trackers (Open The Magazine)** *Ullekh NP*, [Link]

References

Prof. Zubair Shafiq
Associate Professor, University of California, Davis
Email: zshafiq@ucdavis.edu
Webpage: `https://web.cs.ucdavis.edu/~zubair/`

Prof. Savvas Zannettou
Assistant Professor, TU Delft, Netherlands
Email: s.zannettou@tudelft.nl
Webpage: `https://zsavvas.github.io`

Prof. Rishab Nithyanand
Assistant Professor, University of Iowa, Iowa
Email: rishab-nithyanand@uiowa.edu
Webpage: `https://sparta.cs.uiowa.edu/people/rishab/`

Dr. Nicolas Kourtellis
Research Scientist, Telefonica Research Labs, Spain
Email: nicolas.kourtellis@telefonica.com
Webpage: `https://www.concordia-h2020.eu/nicolas-kourtellis/`