SRISHTI GUPTA: srishtig@iiitd.ac.in | +919873816790

PROFILE SUMMARY

Data Scientist with **Ph.D.** in Computer Science and ~3 years of industry experience in building scalable end to end products for consumption by internal and external stakeholders. Overall**, 9+ years** of hands-on experience in leveraging machine learning models and data mining techniques

EDUCATION

Ph.D. in Computer Science	**2014 – 2019**
Indraprastha Institute of Information Technology, Delhi (IIIT-Delhi)	**CGPA**: 9.2 / 10
M. Tech in Computer Science (CS) with specialization in Information Security	**2011 – 2013**
Indraprastha Institute of Information Technology, Delhi (IIIT-Delhi).	**CGPA**: 9.2 / 10
B. Tech in Computer Science and Engineering (CSE)	**2007 – 2011**
Northern India Engineering College, GGSIPU	**Aggregate**: 83.25 %

TECHNICAL SKILLS: Python, MySQL, PySpark, Machine Learning, NLP, Data cleansing, Hive, MongoDB

INTERNATIONAL EXPOSURE

Visiting Researcher, **Georgia Institute of Technology (GaTech), Atlanta, USA**   **June 2016 – August 2016**
- Quantitative study of emerging threats related to phone numbers abuse on online social networks

Visiting Researcher, **New York University (NYU), Abu Dhabi, UAE**   **October 2014 – June 2015**
- Built a framework to understand the feasibility and effectiveness of targeted attacks on OTT applications

Research Intern, **University at Suny Buffalo (UB), NY, USA**   **June 2013 – August 2013**
- Qualitative analysis to understand online privacy perceptions of elderly citizens living in India

JOB EXPERIENCE

Data Scientist, American Express (India)   **August 2018 – Current**

Knowledge Graph platform for Data Discovery
- Goal was to automate the manual tedious task by data analysts to discover insights from tables
- Built an **ingenious** algorithm to identify latent features from sample values, classified columns into data types
- **Achieved a 200x gain in run-time** with Spark technology for computing similarity b/w **330M column pairs**
- Designed wireframes with networkX to showcase millions of relationships which now has **100+ monthly users**
- Led the partnership with technology team to convert the wireframes into a full-fledged product
- **Recipient of "GSG Tribute award for innovation"**; research paper was submitted to **PAKDD, 2021**

Merchant Outlier Detection
- Built an outlier detection package based on statistical **z-score** to identify deviant merchants in dispute calls
- Formulated the method of **bucketization** to identify merchants in a homogenized pool of data
- Identified **subject matter** of these merchants calls to gather pain themes across merchants for rectification

Topic Categorization
- Designed and developed topic categorization algorithm to predict topics from call transcripts using ML models
- **The generated data source is consumed by Amex** to gauge the intention of the call coming to customer care
- The features used in bag of words were augmented using **word2vec** created based on the domain knowledge

Recorded Line Message Identification
- Identified IVR call segments by tagging recorded line messages using **fuzzy matching techniques**
- Helps in eradicating the call snippets that are irrelevant for further text processing to gather insights
- The **package is being used as pre-processing step** in several text analytics techniques across several projects in the team

Word Trend Analysis
- Developed word trend statistical algorithm to predict trending words in calls coming to the organization
- Leveraged **z-scores** and moving window historical data prediction for implementation
- **Product used widely by the organization** to understand trending spikes in the customer calls

PUBLICATIONS
- **Gupta, S.,** Bhatia, G., Suri, S., Kuchhal, D., Gupta, P., Ahamad, M., Gupta, M., and Kumaraguru, P. Angel or Demon? Characterizing Variations Across Twitter Timeline of Technical Support Campaigners. Journal of Web Science. 2019.
- **Gupta, S.,** Khattar A., Gogia, A., Kumaraguru, P., and Chakraborty T. Collective Classification of Spam Campaigners on Twitter: A Hierarchical Meta-Path Based Approach. 27th International World Wide Web Conference (WWW), Lyon, April 2018.
- **Gupta, S.,** Kuchhal, D., Gupta, P., Ahamad, M., Gupta, M. and Kumaraguru, P. Under the Shadow of Sunshine: Characterizing Spam Campaigns Abusing Phone Numbers Across Online Social Networks. Accepted in the 10th ACM Conference on Web Science, Amsterdam, 27-30 May 2018.
- **Gupta, S**., Gupta, P., Ahamad, M., and Kumaraguru, P. Exploiting Phone Numbers and Cross-Application Features in Targeted Mobile Attacks. 6th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices. ACM, 2016.
- **Gupta, S.** Emerging Threats Abusing Phone Numbers Exploiting Cross-Platform Features. 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
- **Gupta, S**., Gupta, P., Ahamad, M., and Kumaraguru, P. Know your targets: Privacy and Security Implications in Instant Messaging Applications. Poster at 2nd NYUAD Annual Research Conference, Abu Dhabi, 2015.
- **Gupta, S**., and Kumaraguru, P. Emerging Phishing Trends and Effectiveness of the Anti-Phishing Landing Page. Accepted at 9th APWG eCrime Research Symposium (eCRS), 2014.
- **Gupta, S**., Gupta, M., Kumaraguru, P. OCEAN: Open-source Collation of eGovernment data and Networks. **Best Poster award** at IIT-Kanpur Security and Privacy Symposium, 2013.
- **Gupta, S**., Kumaraguru, P. OCEAN: Open-source Collation of eGovernment data And Networks. Understanding Privacy Leaks in Open Government Data. Submitted in partial fulfilment of the requirements for the Degree of M.Tech in Computer Science, with specialization in Information Security. IIIT-Delhi. 2013.
- **Gupta, S**., Gupta, M., Kumaraguru, P. Is Government a Friend or Foe? Privacy in Open Government Data. Poster at IBM-ICARE, IISc Bangalore, 2012.

ACHIEVEMENTS
- Received the prestigious **GSG award** for **technical innovation in knowledge graph project**, Q3, 2020
- Invited by VP head to Phoenix and Florida to train **500+** colleagues in Python, ML, and NLP, 2020
- Won **2nd prize** for developing technical **skill matrix** of Amex employees, 2020
- Quarterly **Recognition of performance winner** for exceptional performance in Q4, 2019

LEADERSHIP RESPONSIBILITIES
- Trained **500+** Amex employees in India, Florida, and Phoenix on ML, NLP, and Python
- **Partnered** with Technology team to build variable search component in knowledge graph portal
- Partnered with EDA team for metadata curation for the critical & standardized data elements in Cornerstone
- **Developed learning resource** for data analytics team across organization on **newly launched platform, Spacy**
- Mentored 10+ B.Tech. and M.Tech. students during PhD; responsible for converting **4 M.Tech. thesis**
- Served as **Teaching assistant** for 5+ B.Tech. and M.Tech. courses
- Gave several **talks** on online social media analytics to **teach** police officers, teachers, and college students
- **Organized** summer school for social media analytics in IIT Kanpur, IIIT Hyderabad, IIIT-Delhi

ACADEMIC ACHIEVEMENTS
- **9+** publications during PhD in highly reputed conferences and journals; **70+** citations received
- Invited to Vienna, Amsterdam, San Francisco, and Lyon conferences for paper presentations
- Amongst 20 Ph.D. students from India for Doctoral Symposium at NIIT University, 2017
- Invited **Panelist** at Messaging Malware Mobile Anti-Abuse Working Group (M3AAWG), Philadelphia, 2016