Gurpreet Singh Bhatia

gsbhatia@umich.edu
+1-734-882-9498
Linkedin

EDUCATION

- **University of Michigan** — Ann Arbor, MI
  Master of Science in Information, Specializing in Data Science — *May 2021*

- **Guru Gobind Singh Indraprastha University** — Delhi, India
  Bachelor of Technology, Information Technology.; GPA: 3.76 — *July 2015*

EXPERIENCE

- **Indraprastha Institute of Information Technology** — Delhi, India
  Research Associate — *Sep 2017 - Aug 2019*
  - **Advanced Analytics for Social Media Analysis (AASMA)**:
    - Key contributor to designing, developing, and maintaining large-scale data mining tool.
    - Scaled the project from 60 to 100+ organizations leading a team 5 developers. [Python, Rails, Apache Storm]
  - **Spam Campaign Characterization**:
    - Developed framework to detect and suspend spammers by analyzing over 23 million tweets using Random Forest.
  - **Digital Signals on Stack Overflow**:
    - Identified evidence of positive correlation between reputation scores and badges, and social attributes *Popularity* and *Impact* of Stack Overflow (SO) users.
    - Applied Gradient Tree Boosting Regression to model dataset of 3.8 Million users; $R^2(Popularity) = .956$ and $R^2(Impact) = .858$.

- **Sapient Consulting** — Delhi, India
  Associate Software Engineer L1 — *April 2016 - Jul 2017*
  - **Data Integration Middlware Platform**:
    - Built Data Integration framework with Publisher / Subscriber format that provides a middleware platform for communication between proprietary systems.
    - Steered development of component for file format conversion using business logic.[IIB, Java Swing, PL/SQL]

PROJECTS

- **Honda CV Hackathon**: Developed application to detect parking spot on the road using Yolov3 and Darknet.
- **Headline influence on Stock Value of a company**: Emperical analysis showed that there is 7.4% chance for value change of more than 5% in Indian Market for year=2008. [Hadoop, PySpark, Spark SQL]
- **DeepTraffic**: Utilized Deep Reinforcement Learning to model a car that drives through virtual highway traffic.
- **Video Analysis**: Added support for Non-GPU servers to poll and execute video functionality on GPU servers in the network using RPyC.
- **Track On Twitter**: NodeJS application to track keywords on Twitter using Chrome notifications. Achieves Async I/O using Callbacks and Promises.

PUBLICATIONS

- "Signals Matter: Understanding Popularity and Impact of Users on Stack Overflow", Short Paper, **The Web Conference 2019 (WWW)**, 20% acceptance. [Link]

- "Under the Shadow of Sunshine: Characterizing Spam Campaigns Abusing Phone Numbers Across Online Social Networks", **Journal of Web Science 2019**.

SKILLS

- **Languages**: Python, Java, SQL, R, Ruby, Javascript **Tools**: Hadoop, Spark, Git, Docker
  Machine Learning: Keras, TensorFlow, Pandas **Frameworks**: NodeJs, Spring MVC, MongoDB

COURSES

- Computational Data Science and Machine Learning, Data Manipulation, Algorithms, Computer Networks, Operating Systems