# Ashutosh Tiwari

812-606-5974 | ashutiwa@iu.edu | checkashu@gmail.com | Linkedin | Homepage | GitHub

## EDUCATION

**Indiana University (Graduating May 2023)** — Bloomington, IN
*Master of Science in Data Science (Computational and Analytical Track) 3.83/4.0* — *Aug. 2021 – May. 2023*

**National Institute of Technology** — Patna, India
*Bachelor of Science in Computer Science 8.32/10.0* — *Aug. 2011 – Jun. 2015*

## RESEARCH EXPERIENCE / INDEPENDENT STUDY

- Part of project, **"User Intent as a Network"** with Prof. YY Ahn, Prof. P Kantak and Prof. FB Yara. Project is funded by Kelly Business School. Goal is to be able to quantify and thus act upon that "intent" network. Summer 2022 - Present

- Implementing novel training frameworks that result in unbiased (or fair) models from "biased" datasets at IUNI with Prof. YY Ahn and Prof. S Kojaku as part of my Independent Study. Summer 2022 - Present

- Was part of NLP Lab@IUB with Prof. D Cavar in Fall 2021. Contributed extensively to design of TieML and Events' Timeline modelling. Fall 2021

## WORK EXPERIENCE

**Senior Software Dev Engineer** — Jan. 2019 – Jul. 2021
*SWIGGY* — *DATA SCIENCE PLATFORM, TIME SERIES FORECASTING*

- Was part of team that worked on Feature Store and pipeline which feeds on-demand features to deployed ML models at production scale(4Bn rows, 1Mn QPS / day). Pipeline supported multichannel ingestion, i.e. Spark, Flink and user files etc.
- Founding member of Forecasting and Correlation Platform which was considered by many teams to forecast concerned time series. These forecasts power critical scaling decisions across organizations in real time.
- Led DAQ, a tool used to scrape APIs at scale. Used to collect data for analysis/ model training at a scale of 15 M rows daily.

**Software Development Engineer(ML)** — Sep. 2017 – Jan. 2019
*FLIPKART (a Walmart company)* — *SEARCH RELEVANCE, QUERY INTENT, NLP*

- Was responsible for improvements/inception of search intent models(CRF/Neural Network based), identifying error classes, coming up with solutions, and fixing them. These models power user search and discovery for millions every day.
- Implemented a FastText based query store classifier, which predicts the category of a tail query.
- Implemented the first workflow to automate training and auto-deployment of various search models in Flipkart. First was written using Luigi and later migrated to Airflow.
- Wrote a generic framework using Airflow which at runtime creates generic dags for different ML models and orchestrates their training to deployment flow, including data and model validations.
- Implemented large scale (4Bn+ datapoints) pipelines using Cascading/HDFS to extract data from user events and then transform it to be used for training these models.

**Software Development Engineer** — Sep. 2016 – Sep. 2017
*GROUPON* — *BACKEND ENGINEERING*

- Worked on a component called Cyclops, an interface between Customer representatives and internal services

**Software Engineer** — Sep. 2015 – Aug. 2016
*NETSPEED SYSTEMS (Acquired by Intel)* — *GRAPH ALGORITHMS, NETWORK ON CHIP*

- Led engineering efforts on modules like Polarity based Arbitration, Multi-Cast Filtering, Structural Latency Breakdown etc.

## TEACHING EXPERIENCE

Teaching Assistant for Machine Learning (CSCI-B 555) with Prof. R Khardon — Fall 2022
Teaching Assistant for Network Science (INFO-I 606) with Prof. YY Ahn — Spring 2022

## SELECTED PROJECTS

**BiasNet** | *Deep Reinforcement Learning, PyTorch, Actor Critic Algorithm, On-policy Model Free* — Code/Report
- Learning to fight in Street Fighter II with induced relational bias from differential game scenes

**DeepFoodie** | *Python, Tensorflow, Self Supervised Deep Clustering, Deep Learning, Transfer Learning* — Code/Report
- Clustering dishes on basis of their ingredient embeddings. These ingredient embeddings are generated by a NN.

**Continuous Dominant Set Repair** | *C++, Graph Algorithms, Guha and Khuller's Algo., Greedy* — Code
- Repairs a broken link in Continuous Dominant Set in $O(\Delta^2)$, where $\Delta$ being the avg cardinality of connected graph

**Humana Mays Healthcare Analytics Case Competition** | *Boosted Trees, Feature Engineering* — Code
- 11th rank on leaderboard. Hosted by TAMU and Humana Mays, 2021.

**AnalyticsVidhya Solutions** | *Python, Tensorflow, Torch, Time Series, Feature Engineering, Catboost* — Code

## Technical Skills

Graph Neural Networks, Computational ML, NLP, Computer Vision, DL, Akka, Distributed Systems, Spark, HDFS, Python, Scala, Java, C++

**Frameworks/Libraries/Tools**: Pytorch Geometric, Pytorch, Tensorflow, Sklearn, Numpy, Pandas, Matplotlib, AWS, Neo4j

## Relevant Courses so far

**Computational ML(B555), Deep Learning(E533), Reinforcement Learning(B659), Computer Vision(B657), Statistics(S520), Independent Study (Working on Fairness Aware AI and Bias manifolds)**