





# SHUBHAM LOHIYA

 shubhlohiya.github.io |  shubhlohiya@gatech.edu |  shubhlohiya |  lohiya-shubham

## EDUCATION

### Georgia Institute of Technology

Master of Science in Computer Science (Machine Learning Specialization)

Atlanta, GA, USA

Aug 2022 – May 2024

- Cumulative GPA: **4.0/4.0**

### Indian Institute of Technology Bombay

Bachelor of Technology (with Honors) in Mechanical Engineering with Minors in CS and AI

Mumbai, India

July 2018 – May 2022

- Cumulative GPA: **9.59/10.0**; **Rank 1 – Institute Silver Medal**
- Other Awards: **Institute Academic Prize** [2020 & 2021], **Academic Outstanding Student** [2022]

## KEY PROFESSIONAL AND RESEARCH EXPERIENCE

### Machine Learning Engineer Intern | Adobe

Incoming ML intern for **NLP** projects with Adobe's Digital Experience Team

Summer 2023 (Incoming)

San Jose, CA

### Graduate Research Assistant | NLP | Pathology Dynamics Lab

Guide: *Prof. Cassie S. Mitchell, Department of Biomedical Engineering*

Aug 2022 - Present

Georgia Tech

#### Document Filtering for Drug Repurposing and Clinical Meta-Analysis

- Involved in development of **BioSift**, a new human-annotated document classification dataset for drug repurposing
- Implemented a suite of **NLI-based Zero-shot Multi-label Classifiers** for a document classification benchmark

#### Biomedical Entity Linking Survey

- Introduction:** Biomedical entity linking (BioEL) is the identification of biomedical concepts in text using **NLP**
- Conducted a comprehensive survey of BioEL models on a **new benchmark** of uniformly formatted datasets

### Joint Completion and Alignment of Multilingual Knowledge Graphs

Bachelor's Thesis, Guide: *Prof. Soumen Chakrabarti, Department of Computer Science*

Aug 2021 - July 2022

IIT Bombay

- Introduction:** For multilingual KGs, AlignKGC attempts to study whether Knowledge Graph Completion or fact prediction adds valuable information for the Relation Alignment and Entity Alignment tasks, and vice versa.
- Developed AlignKGC, a new **SOTA framework** for joint completion and alignment of Multilingual KGs
- Created novel datasets from DBP-5L, DBP15k, and OpenEA benchmarks for joint analysis of KGC, EA, and RA

### Data Scientist Intern | Anheuser-Busch InBev | Growth Analytics Center

Engineered a ML Framework for Accounts Receivable **Forecasting**, delivering **\$40 million+** in working capital benefits

May 2021 – July 2021

- Developed trend and seasonal features from historical data, and data from Sales, Collections and the Economy
- Trained ensembles of models like **ARIMAX**, **XGBoost**, **Random Forest**, and **MLP** achieving **< 5%** dispersion

## KEY TECHNICAL PROJECTS

### Mastering Atari Games using Deep Reinforcement Learning

Spring 2021

- Trained and compared Deep Reinforcement Learning agents capable of surpassing human performance on classic Atari games like Pong, Breakout and Boxing using high-level sensory information in the form of game pixels

### Automatic Raga Recognition in Indian Classical Music

Fall 2020

- Trained **LSTMs** with **attention** on random subsequences from the Carnatic Music Dataset for raga classification
- Leveraged techniques like **tonic normalization** and **pitch tracking**, achieving a test-set accuracy of **96.67%**

### Shortest Path in a Maze

Fall 2020

- Modeled 2D mazes as **MDPs** with appropriate states, actions, rewards and transitions to find the shortest path

## TECHNICAL SKILLS

### Programming Languages Technologies/Frameworks Concepts/Courses

Python, C++, SQL, HTML, CSS, JavaScript, Bash, R, MATLAB, Java  
PyTorch, TensorFlow, Keras, OpenCV, Numpy, Pandas, Docker, Git  
Limited Supervision in ML, Natural Language Processing, Deep Learning, Intelligent Agents, Digital Image Processing, Operating Systems, Computational Data Analytics, Web Search and Text Mining, DL for Robotics, ML in Computational Biology

## PUBLICATIONS

- Joint Completion and Alignment of Multilingual Knowledge Graphs**, *accepted at EMNLP 2022*
- BioSift: A Dataset for Filtering Biomedical Abstracts for Drug Repurposing**, *accepted at SIGIR 2023*
- A Comprehensive Evaluation of Biomedical Entity Linking Models**, *Working Paper*

## TEACHING EXPERIENCE

- Graduate Teaching Assistant** | Georgia Tech | *Graduate Artificial Intelligence course*, Fall 2022
- Teaching Assistant** | IIT Bombay | *Intro ML, Deep Learning, Computer Programming, Linear Algebra*
- Python is Cool, Kids** | Student-run summer course for Practical Python Programming | *1000+ enrollments*