Siddhi Potdar

siddhi.potdar@utdallas.edu | +1 469-998-4184 | siddhipotdar | siddhi-potdar | siddhi-potdar.github.io

Education

University of Texas at Dallas, Richardson, USA *Aug 2022 - May 2024 (Expected)*
Masters of Science - Computer Science **GPA: 3.7/4.0**

Savitribai Phule Pune University, Pune, India *Aug 2017 - Jul 2021*
Bachelor of Engineering - Computer Science **GPA: 3.8/4.0**

Experience

Hewlett Packard Enterprise, Houston, USA *May 2023 - Present*
AI & Machine Learning Engineer Intern

- Contributed as an ML Engineer Intern at HPE Cray by engineering three innovative features using machine learning and data analytics within a three-month timeframe to enhance the efficiency of HPC systems.
- Created a prototype for visualizing HPC network data, facilitating the identification and localization of network bottlenecks.
- Analyzed diverse data sources for HPC system logs and constructed an intuitive dashboard for HPC system administrators to depict real-time job execution trends within the network.
- Trained a robust classifier to predict the likelihood of job failure in HPC systems, to reduce job failures, and promote efficient utilization of resources. Prototype showcased an accuracy of 80% on test data.
- Contributed to presentations to internal stakeholders and leadership that focused on communicating project progress, results, and potential impacts.
- Continuing my summer internship through the fall semester as a part-time position.

PubMatic, Inc., Pune, India *Apr 2021 - July 2022*
Machine Learning Engineer

- Contributed to the ML team at the programmatic advertising company, serving over 500 global clients on its platform.
- Constructed and enhanced Spark and Hadoop pipelines to execute automation test suites for machine learning models involved in header bidding.
- Generated comprehensive reports for 3 machine learning projects focusing on header bidding and ad-monetization.
- Conducted thorough unit testing and regression testing, achieving an impressive code coverage of over 80%, across various ad-bidding projects. These efforts helped identify data drift issues that led to model decay in production.

Persistent Systems Pvt. Ltd., Pune, India *Jan 2021 - Mar 2021*
Software Developer Intern

- Underwent training and assesments in industry technologies like Java, RDBMS, Maven, and Git at the multi-national software product company.

myMO IT Solutions, Remote *Jan 2020 - Mar 2020*
Software Developer Intern

- Developed a web app that extracts the table of contents of a documentation website, whilst preserving its structure.
- The deployment showed a 90% accuracy on standard documentation websites and created a browsable index per site.

Publications & Patents

Artificial Vision for Visually Impaired (Paper: Aug 2020, Patent: July 2021)

- Developed an end-to-end voice-controlled android app that assists the visually impaired by answering questions about their surroundings. It give near-instant and accurate responses to users by using web APIs for speech-to-text and text-to-speech and lightweight machine learning libraries for computer vision.

Selected Projects

Generating Counter Narrative for Hate Speech *(Tech: Python, Tensorflow)* *Apr 2023*

- Utilized BERT model to perform hate speech classification on different narratives. Fine-tuned GPT-2 and XLNet using multi-conan dataset for NLG task. Evaluation resulted in XLNet achieving a BLEU score of 0.598.

arXiv Paper Tagger *(Tech: Python, scikit-learn)* *Sep 2022*

- Used the scikit-learn library to solve a multi-label classification problem to tag arXiv papers with relevant categories. Performed preprocessing of the data using word embeddings and sentence embeddings. Assessed the performance of different supervised machine learning algorithms, performed hyper-pameter tuning and achieved 96% accuracy.

Revenue Forecasting *(Tech: Python, Prophet)* *Jul 2022*

- Used the Prophet library from Facebook to forecast the daily and quarterly revenue of PubMatic using real-time data. Insights and trends obtained helped the company maximize their profits by scaling efforts in peak holiday season.

Facial Recognition Based Attendance Marking System *(Tech: Python, TensorFlow)* *Sep 2019*

- Developed a desktop-based efficient attendance marking system and data analysis tool using Python. Front-end built using PyQt5, and database management implemented using MongoDB.

Skills

Computer Languages: Python, C++, HTML, Java, LaTeX

Frameworks: scikit-learn, TensorFlow, Flask, Scrapy, HTML, Bootstrap, Docker, Git, Spark, Hadoop, Jupyter

Databases: MySQL, MongoDB

Honors and Awards

First place at the Revenue Forecast Competition held over 2 quarters at PubMatic with more than 10 teams.
Awarded **#TeamWork individual award** at PubMatic for significantly reducing automation backlogs and contributing to framework enhancements.
First place at the IET-Hackathon at PCCOE, among more than 50 participating teams.