Sonal Yadav

Seattle, WA | sonaly@uw.edu | LinkedIn | GitHub

EDUCATION

University of Washington| *Master of Science, Computer Science and Software Engineering* September 2022 – August 2024
Coursework: Machine Learning, Algorithm Design, Parallel Programming, Software Architecture, Distributed Computing **[GPA: 3.6/4]**

Dr. A.P.J Abdul Kalam University | *Bachelor of Technology, Computer Science* August 2017 – July 2021
Coursework: Database Management, Data Structure and Algorithms, Object Oriented Programming, Operating Systems **[GPA: 8.0/10]**

WORK EXPERIENCE

University of Washington September 2023 – August 2024
Research Assistant Bothell, WA

[PROJECT: Developed "Inflorescence," a model-agnostic Python library utilizing Clustered Federated Learning (CFL) to securely train data and visualize results.]

- **Developed and Deployed**: Created the **Python library** "Inflorescence" using **PyTorch**, enabling researchers to experiment with any dataset by leveraging built-in **CFL** strategies for privacy-preserving machine learning.
- **Implemented Metrics**: Integrated key performance metrics such as accuracy, recall, and precision to evaluate model performance, leading to initial results demonstrating a 15% improvement in model effectiveness compared to traditional methods.
- **Contributed to Research**: Supported the Ubicomp Lab's research in **privacy-preserving machine learning**, providing a versatile tool that advances the study of secure and federated learning frameworks.

Infosys July 2021 – September 2022
Software Engineer – Data Platform India

[PROJECT: Financial data management for a bank, enhancing data accessibility, transaction processing, and customer service management]

- **Designed and Managed**: Led the development of an end-to-end **ETL pipeline** to migrate sensitive financial data into a **Redshift data lake** using **Python, Apache Airflow,** and **AWS**. This enhanced data accessibility and increased analytic processing speed by 30%
- **Collaborated Across Teams**: Worked closely with cross-functional teams to define data requirements and implement data models using **SQL,** ensuring efficient data organization and retrieval
- **Developed Automation:** Created **automated scripts** to monitor and maintain pipeline health, reducing downtime by 20% and ensuring seamless data operations
- **Provided Technical Support:** Trained team members on best practices for data management, contributing to more efficient team.

PROJECTS

- **Diabetes Data Analyzation (XGBoost, SVM, Random Forest, Decision Tree Regressor) -** Predicted diabetes risk with high precision and developed a recommendation system for personalized diabetes management
- **Python Library FairMobil (Python, PyTorch):** Developed and deployed the Python Library “fair-mobil” as an open source tool to assess the accuracy and fairness of generative mobility models, furthering fairness research at UbiComp labs
- **Distributed Web Services Optimization (Python, Apache Kafka, Zookeeper, Web Workers, Prometheus, Grafana)** - Developed a custom scheduling algorithm, reducing data processing times by 30% and boosting throughput by 30%
- **Parallel Programming for Algorithm Optimization (OpenMP, MPI, Apache Spark, MapReduce)** - Applied parallel computing techniques to optimize the grasshopper optimization algorithm, significantly reducing computation time

SKILLS

Programming Languages: Java, Python, Go, C++, JavaScript, HTML, CSS, and SQL
Cloud: AWS Elastic Beanstalk, AWS Lambda, AWS ELB, Amazon EC2, Amazon SQS, Amazon RDS and Amazon S3
Database & Tools: MongoDB, DynamoDB, MySQL, PostgreSQL, MS SQL, Git, GitHub, Visual Studio Code, IntelliJ
Frameworks: Spring-Boot, React, Node.js, Flask, Apache Kafka, Redux, JUnit, Apache Airflow, and Postman
Work Areas: Distributed Systems, Machine Learning and Full Stack Development