Priya Jha

6824083705 | pj35134@gmail.com | PriyaJha@my.unt.edu | Denton,Texas
Github: https://github.com/priya-zha | **LinkedIn**: https://www.linkedin.com/in/priya-jha-105866167/

SKILLS

Languages: Python | Java | Solidity | Javascript | HTML/CSS
Data Science & Engineering : Pandas | Pyspark | numpy | Matplotlib | Plotly | Streamlit | Kafka | Hadoop | Spark | Scikit Learn | Tensorflow/Keras | OpenCV | Colab | Jupyter | Excel automation
Database: MySQL | Cassandra
Cloud Services: Azure IoT Hub | Azure Blob Storage | Azure Data explorer | Azure data factory | Azure stream analytics | Azure Key Vault | AWS EC2 | AWS S3 Bucket | AWS Lex
Project Management Tools : Git | Trello | Powerpoint | Word | Software Diagramming tools | Slack
Soft skills: Team work | Adaptability | Leadership | Problem-solving. Good command over English, Hindi and Nepali language.

PROFESSIONAL EXPERIENCE

University of North Texas, United States | Graduate Instructional Assistant
Denton, Texas | (current)
- *Assist in grading, teaching and preparing course materials for the 'Fundamentals of AI' class with a total of 55 students*

Expensify Inc | Data QA Analyst - Contributor in the open source project
Oregon, USA | January 2023 - July 2023
- Tested website functionality and features, specifically emphasizing the quality assurance of data inputs to ensure the integrity and reliability of the software product before release. Played a key role in communicating and reporting these issues to the development team.
- Crafted comprehensive bug reports with well-structured components such as titles, descriptions, step-by-step reproduction instructions, expected and actual outcomes, the specific platform and the environment where bugs were encountered, and software version details. These detailed reports expedited the debugging process and facilitated efficient resolution of issues.

Coursera Inc | Project Developer
Mountain View, California USA | August 2020 - August 2022
- Designed and implemented end to end practical projects in the field of Data Science & Engineering and instructed learners on the same. Some of them include :
- Building ETL pipelines with Azure data factory that includes data ingestion from Azure storage and transforming data with mapping data flows in ADF.
- Analyzing and visualizing the UK accidents datasets using pandas and plotly express to get the rate of road accidents in UK between a time period, based on weekdays, based on accident severity distribution, and speed limits.
- Analyzing and querying real-streaming data generated from the Raspberry Pi and sensors with Azure stream analytics.
- Use of Azure data explorer and Kusto Query Language (KQL) to analyze and get insights from the Storms Data (weather) in the US. Created adx cluster, database and tables within the cluster and querying the data using Kusto Query Language(KQL)
- Creation of database instance on Azure SQL database and then utilize SQL Server Management Studio (SSMS) for tasks such as database users' role assignments and querying student data, ensuring seamless management and efficient data retrieval.
- Performed correlation of the independent and dependent variables and feature selection to get the most important independent variables before feeding the data into the machine learning algorithms.
- Implemented a system to detect covid patients using their X-ray reports using deep learning. Also, a system to predict diabetes and breast cancer using classical machine learning algorithms.

LIS Pvt Ltd | Data Engineer Intern
Nepal | January 2020 - August 2020
- Conducted comprehensive analysis of ecommerce data using python and pandas, extracting valuable insights to inform business strategies and decision-making processes that includes identifying the top selling products, items that are frequently bought together, customers spending time, frequency of a repeated customer etc.
- Developed and implemented a recommendation system that dynamically showcased similar items purchased by other customers, thereby increasing cross-selling opportunities and enhancing user experience.
- Got hands-on training on data engineering tools and techniques including pandas, pyspark, matplotlib and AWS.

EDUCATION

- **University of North Texas** | Denton, TX, USA | Masters in Computer Science | *(August 2023 - May 2025)* | 4.0 GPA
- **Tribhuvan University** | Kathmandu, Nepal | Bachelors in Computer Engineering | *(November 2016 - March 2021)*

AWARDS | ACHIEVEMENTS | ACTIVITIES

- Currently working as a Secretary Officer at ACM - W (Association for Computing Machinery for Women)
- Finalist in the biggest tech award of Nepal- ICT Award Nepal, 2020 for my project "COVID19 Pneumonia detection using deep learning", among the pool of 300+ applicants.
- Winner at the National Locus Exhibition, Nepal for the project "Smart Street Monitoring system" in Thematic Electrical that optimize energy consumption in urban areas by intelligently managing street lighting based on real-time traffic conditions.
- Top 10 finalist at Hult Prize, Nepal - a competition that crowdsources ideas from university-level students to solve a pressing social issue around the globe.
- Event Organizer at Girl To Code in association with Girls in Tech - an event that teaches cloud computing skills to young girls in order to motivate them to pursue their careers in IT. Organizing this event showed my leadership skills and volunteer activities.
- Organizer at Yomari Code Camp - a comprehensive platform for students and young researchers to unleash their potential to develop their ideas into useful applications
- Participant at UNT annual hackathon - Building a food alert app that helps university students to get the food availability alerts in a real time
- Participant at Quantum hackathon in Nepal
- Participant at Everest hackathon in Nepal
- Member of Women in Computing at UNT

PROJECTS

- Health data mining using Big data Tools in distributed environment   **Python | AWS EC2 | Kafka | Spark | Hadoop | Cassandra | Linux**
Implemented lambda architecture with a data pipeline built on a distributed cluster of 4 nodes created with Aws Ec2 instances for processing real time as well as batch data to predict diabetes in real time and locating diabetic people based on their regions.

- Processing IoT data using Azure Stream Analytics   **Azure IoT Hub | Azure Stream Analytics**
 Processing and querying the real-time streaming data that includes temperature, humidity, device and message id coming from the IoT device using Azure Stream Analytics and storing the result to the Azure storage.

- Creating ETL pipelines using Azure Data factory   **Azure Storage | Azure Data Factory**
 Created an ETL pipeline that focused on data ingestion from Azure Blob containers, and Mapping Data Flows were utilized to transform data, utilizing Select, Derived Column, Join, Conditional Split, Filter, and Sort transformations. The transformed data was then saved to Azure Storage, ensuring a comprehensive and efficient data processing workflow.

- Azure Data explorer to analyze storms data in US   **Azure Data Explorer | Kusto Query Language**
 Utilized Azure Data Explorer to analyze storm data in the United States by creating an ADX cluster, and database, and tables within that cluster. Employed Kusto Query Language (KQL) to execute queries and analyze the storm data, generating valuable insights.

- AI based Android app for blind users (https://bit.ly/3HoerJk)   **Android | Java | Python | YOLO | Text-to-Speech | Speech-To-Text | Flask API | Image Captioning -** Link
An app that operates through voice commands, providing real-time object recognition and scene descriptions to assist blind users in understanding and navigating their surroundings.

- Graduate Admission Prediction with Pyspark MLlib   **Python | Pyspark MLLIB | Colab | Kaggle**
A system that predicts the chances of getting admission in a university using evaluation factors like TOEFL, GRE, Papers, GPA, etc

- Skin Cancer detection using Deep Learning   **Python | Kaggle | Colab | Tensorflow/Keras | Matplotlib | CNN | Android | Java**
An android app that detects skin cancer based on the given input image

- Covid-19 Pneumonia Classification using Chest X-rays   **Python | Streamlit | Kaggle | Colab | Tensorflow/Keras | Matplotlib | CNN**
 A web application that takes patients X-ray reports and classifies the image of being covid infected or not

- Face Mask Detection in a real-time video to prevent COVID 19 using deep learning **Python | Tensorflow/keras | OpenCV | Kaggle | Colab**
To challenges posed by the COVID-19, a real-time video system that detects whether people are wearing face masks or not.

- Human Activity Recognition with Video Classification (CNN+LSTM)   **Python | Tensorflow/Keras | CNN | RNN (LSTM) | Plotly**
A system that classifies the activity of human beings in a video based input

- Breast Cancer Prediction Using Logistic Regression in ML   **Python | Kaggle | Scikit-Learn**
A system that predicts the breast cancer using machine learning

- Diabetes Prediction Using Pyspark- ML   **Python | Kaggle | Pyspark | Pyspark MLLIB | Colab**
A system that predicts diabetes based on the medical report of the patient like insulin, glucose, BMI, age , gender etc.