

SHATABDI PAL

16995 NW Anita St., Portland, OR, 97229

Phone: 443-467-7218, **Email:** sumi.shat@gmail.com, **GitHub:** <https://github.com/shatabdi-pal>

Immigration status: U.S. legal permanent resident

EDUCATION

Predoctoral Student | Oregon Health and Science University **2025 - Present**
Major: Health & Clinical Informatics
GPA: 3.85

Master's | Portland State University, Portland, OR, USA **2022 –2024**
Major: Computer Science
GPA: 3.7

Bachelor's | Heritage Institute of Technology, Kolkata, India **2008 –2011**
Major: Electronics & Communication Eng.
GPA:7.9/10.0

Diploma | Women's' Polytechnic, Kolkata, India **2005 – 2008**
Major: Electronics & Telecommunication Engineering
Score: 81.6 %

TECHNICAL SKILLS

Languages Proficient: Python, PostgreSQL, C++, HTML5, CSS3, JavaScript

Experience: Machine Learning, NLP, ETL, AI, Statistical Analysis, Relational Database

Technical Stack: Git, Power BI, Tableau, MS Excel, Snowflake, GCP, Confluent Kafka, Azure, Jupyter Notebook

PROFESSIONAL EXPERIENCE

Data Science Intern | Data Glacier India **2024**

- Performed exploratory data analysis and applied statistical methods for data analysis to make business decisions.
- Led on projects, including market insight for a cab investment firm and a bank marketing campaign.
- Utilized Bit Bucket for software version control, Kanban for project management, deployed Flask applications on Azure and Heroku
- Handled file ingestion with Modin, Ray, and Dask and performed schema validation using YAML.

Software Engineer | Tech Mahindra Limited, India **2012 - 2015**

- Leveraged critical thinking and problem-solving abilities to supervise jobs using the Citrix Control-M tool within the Operational Data Management project for British Telecom, UK.
- Utilized analytical skills to perform ETL processes and data analysis, facilitating efficient data management within the project.
- Collaborated effectively with designer, developer, and testing teams to resolve complex issues.
- Showcased strong written and verbal communication skills while working in a cross-functional environment, including data visualization to communicate insights.

PROJECT EXPERIENCES

Design and Evaluation Method for Clinical Research

2026

- Conceptual project on AI-driven clinical decision support systems (CDSS) for diabetes management
- Project aims to improve the design and evaluation of AI-driven CDSS in diabetes management
- It assesses effectiveness across clinical outcomes, patient-centered measures, health-system impact, and ethical considerations, using prospective, real-world, and equity-focused evaluation methods.

Graduate Data Analyst / Prediction of readmission for diabetes patients

2024

- Led a comprehensive Exploratory Data Analysis (EDA) to uncover patterns and address data quality issues in the dataset, including outliers and missing values.
- Conducted in-depth analysis to identify critical factors influencing hospital readmission rates for diabetes patients, providing actionable insights.
- Developed recommendations based on EDA findings to enhance healthcare outcomes and improve predictive accuracy in the Machine Learning phase.

Graduate NLP Researcher / Diverse Model Approach to Fake News Detection

2023

- Explored finetuned Pretrained Language Models (PLMs), Large Language Models (LLMs), and traditional machine learning models to classify real and fake news in a project focused on fake news detection.
- Evaluated the performance of various language models using three different datasets to determine their effectiveness in detecting fake news.
- Employed the Liar and WELFake datasets for model training and evaluation.

Graduate Data Engineer / Build TriMet Data Pipeline

2023

- Collaborated in a group project to extract, clean, validate, integrate, and visualize TriMet bus data for the Portland metropolitan area.
- Configured a GCP Linux VM to gather and parse TriMet breadcrumb data, sending JSON records to a Kafka topic.

- Developed a Confluent Kafka consumer to validate, transform, and store data in a database, ensuring daily ingestion.
- Built a secondary pipeline to integrate transit vehicle stops with breadcrumb data for comprehensive visualization.

Graduate ML Engineer / Binary classification for diabetes detection **2022**

- Performed data cleaning and handled missing values to prepare the dataset for ML models.
- Prepared the data for machine learning by ensuring its quality and suitability for model training.
- Employed various machine learning techniques for diabetic prediction based on diagnostic measures.
- Evaluated and compared the performance of a multilayer perceptron and a Bayesian classifier in predicting diabetes.

Graduate Database Designer / Implementation of Health Database using Census Data **2022**

- Built a database for the health domain using data from the United States Census Bureau.
- Enabled users to analyze how different groups of people are affected by the COVID-19 pandemic.
- Utilized census data to provide insights into the pandemic's impact on various demographics.

Graduate NLP Researcher / Coherency-based Text Summarization Metrics **2022**

- This project is based on the research article SNaC, which aims to characterize the coherence errors generated by state-of-the-art summarizing models.
- Compared and reported the performance metrics of various models on SNaC for the same dataset.

CERTIFICATION

- | | |
|---|-------------|
| • Snowflake for Data Science: Intro to Snowpark ML for Python by Coursera | 2024 |
| • AWS S3 Basic by Coursera | 2024 |
| • Microsoft Learn AI Skills Challenge | 2023 |
| • HTML5 and CSS Fundamentals Certification by W3C | 2018 |
| • Introduction to Python: Fundamentals Certification by Microsoft | 2018 |

PRESENTATION AND TERM PAPERS

- **Submitted research abstract** for OHSU Research Week 2026
- **BMI 638 Presentation:** Decision-making for a suspected sepsis patient in the emergency department
- **BMI 605F:** Overview of HIV infection and Women's Health
- **BMI 610 Term paper:** Securing Privacy through Federated Learning and Blockchain in Healthcare
- **BMI 630 Presentation:** Overview of Hepatitis B infection
- **Sentiment Analysis:** A Comprehensive Exploration Using Traditional and Advanced Models in Medium Platform

- **CS554 Term paper:** How can LLMs help improve software testing and validation
- **CS545 Term paper:** Write a research paper analysis based on the article “Automatic Speech Classifier for Mild Cognitive Impairment and Early Dementia,” which is an application of NLP
- NLP course presentation based on the research paper “Continual Training of Language Models for Few-Shot Learning.”
- Reinforcement Learning course presentation based on the research paper “Playing Atari with Deep Reinforcement Learning.”

LANGUAGES AND HOBBIES

- **Bengali:** Native Language
- **English:** Speak fluently and read/write with proficiency (TOEFL score: 84/120)
- **Hindi:** Speak with basic competence
- **Interests/Hobbies:** Cooking and Travel