

PRANTOSH NEUPANE

Oklahoma City, OK

☎ (405) 349-6469 | ✉ aprantosh@gmail.com | 🔗 Profile: [prantoshneupane.site](https://www.linkedin.com/in/prantosh-n-55318b294/)

🔗 LinkedIn: <https://www.linkedin.com/in/prantosh-n-55318b294/>

PROFESSIONAL SUMMARY

Ph.D. candidate in Statistics and Machine Learning specializing in large-scale data analytics, machine learning, and healthcare applications. Experienced in building and deploying data pipelines and ML solutions using Databricks, Apache Spark, SQL, and modern BI tools. Deep expertise in NLP, LLMs, and MLOps, with demonstrated impact in claims analytics, academic research, and enterprise-level system integration.

SKILLS

Programming & Development

Java | Python | J2EE | Spring Boot | Spring MVC | JSP | C/ C++ | ASP.NET | JavaScript | Azure | Dart

Data Engineering & Analytics

SQL | Databricks | Apache Spark | Informatica | SAS | R | Tableau | Power BI | PyCharm | PyTorch | TensorFlow

Machine Learning & AI

Natural Language Processing (NLP) | Predictive Modeling | MLOps | Large sampling | OpenCV | Large Language Models (LLMs) | Generative AI

Front-End Technologies

HTML5 | CSS3 | Bootstrap | Angular | JSON | UI/UX Design

Databases & Modeling, Cloud

Oracle SQL | Lucid chart | Erwin | Data Lakes | AWS (EC2, S3) | Docker | Kubernetes | Ant Maven | Firebase | JIRA | Postman | Swagger | MATLAB | Slack | Bigdata | Databricks

EDUCATION

Ph.D. in Statistics & Machine Learning

(In Progress)

Oklahoma State University, Stillwater, OK, USA

M.S. in Computer Science & Applied Mathematics

Aug 2021 – May 2023

University of Central Oklahoma, Edmond, OK, USA

B.S. in Computer Engineering

Mar 2016 – Sep 2019

Tribhuvan University, Kathmandu, Nepal

WORK HISTORY

Graduate Teaching & Research Assistant

Oklahoma State University | Oklahoma, USA

Aug 2024 – Present

- Conduct data collection, statistical analysis, and visualization using Tableau, Power BI, and R Studio.
- Designed and managed Oracle SQL databases and data models using Lucid chart.
- Built scalable big data pipelines using Databricks and Apache Spark for research projects.
- Developed machine learning and NLP models for academic research using Python and R.
- Assisted in preparing research publications, reports, and literature reviews.

Business System Analyst III

Oklahoma Health Care Authority (OHCA) | Oklahoma, USA

Jan 2024 – Aug 2024

- Led analysis of healthcare claims and Electronic Visit Verification (EVV) systems.
- Designed predictive machine learning models to enhance claims processing performance.
- Integrated data lakes to centralize healthcare datasets and enable scalable analytics.
- Built Databricks and Apache Spark pipelines to optimize large-scale healthcare data processing.
- developed Tableau dashboards for stakeholder reporting and operational decision-making.

- Applied MLOps methodologies for scalable deployment of analytics models.

Web Development & Data Analyst Intern

Intern Protégé Inc. | California, USA

Nov 2023 – Dec 2023

- Designed responsive UI/UX interfaces using HTML5, CSS3, Bootstrap, and JavaScript.
- Developed backend logic and Oracle SQL database schemas integrated with Spring Boot.
- Built user authentication modules using JSP and MVC architecture.
- Deployed applications on AWS EC2 and S3, using Git for version control.
- Conducted API testing using Postman and Swagger UI.

Graduate Teaching & Research Assistant

University of Central Oklahoma | Oklahoma, USA

Sep 2021 – May 2023

- Supported research in big data analytics and machine learning.
- Designed data preprocessing and visualization workflows using Spark, Databricks, Tableau, and R Studio.
- Assisted in model-driven development (MDD) for reusable research components.
- Prepared academic reports and research publications.

System Data Analyst & Associate Project Manager

Deer Walks Inc. Services | Kathmandu, Nepal

Aug 2019 – Nov 2019

- Developed Tableau dashboards and SQL-based ETL workflows, web applications using Java, JavaScript, JSP, Spring MVC.
- Built predictive analytics models to forecast healthcare trends.
- Managed project deliverables using JIRA, Dynamics 365, and HubSpot CRM.
- Created RESTful APIs and validated services using Postman and Swagger.

CERTIFICATIONS

- Google Data Analytics Capstone, **Google Inc** (Jan 2025)
- SQL for Data Science, **University of California, Davis** (Jan 2025)
- SQL Problem Solving, **University of California, Davis** (Jan 2025)
- SAS Programming & Essentials, **Oklahoma State University** (Sep 2024)
- Flutter Mobile Application Development, **University of Central Oklahoma** (May 2023)
- Azure REST API with Node.js, **Coursera Project** (Oct 2023)
- Business Analysis & Process Management, **Coursera Project** (Nov 2023)
- WordPress & JavaScript Website Development, **Coursera Project** (Nov 2023)

PROJECTS

❖ Soccer Data Mining & Clustering Analysis

R Studio | Machine Learning | Statistical Analysis | Python | PyCharm | PyTorch | LLM | NLP

- Implemented unsupervised machine learning techniques (K-Means, Hierarchical Clustering) to identify player performance and team dynamics patterns.
- Performed data cleaning, feature selection, and normalization to improve clustering accuracy.
- Evaluated cluster quality using statistical metrics and visualizations.
- Delivered a research report translating analytical findings into actionable insights for sports analytics.

❖ Healthcare Full-Stack Application with Microservices Architecture

Java | Spring Boot | REST APIs | Oracle SQL | AWS

- Developed a microservices-based healthcare web application using Spring Boot and RESTful APIs.
- Implemented service discovery and load balancing using Eureka Server.
- Designed and optimized Oracle SQL database schemas for healthcare data management.
- Deployed applications on AWS EC2 and S3, ensuring scalability and reliability.

❖ Real-Time Distributed System for Data Streaming (MLOps-Oriented Infrastructure)

Core Java | Networking | Distributed Systems | MLOps

- Integrated JavaScript-based client components for interactive user experience.
- Demonstrated foundational concepts applicable to distributed systems and real-time AI pipelines.
- Optimized resource utilization to support scalable, real-time data flows, ML inference pipelines.
- Enabled low-latency, bidirectional data exchange between multiple clients and servers.