

# HARISH CHANDRA DEGA

+18132040167 | dr.harishchandrad2@gmail.com | linkedin.com/in/harishchandradega996 | github.com/degaDedataanalyst9 |  
drharishchandrad2.wixsite.com/my-site/projects

## SKILLS

**Programming Languages and Libraries:** python, Pandas, NumPy, Scikit-Learn, Matplotlib, TensorFlow, R, ggplot2, dplyr, caret, SQL, NoSQL, MongoDB

**Tools:** Tableau, PowerBI, Excel, Word, PowerPoint/Keynote/Slides, Electronic Medical Records (EMR), Electronic Health Records(EHR), EPIC

**Cloud Platforms:** AWS, EC2, S3, Athena, Glue, Microsoft Azure, Snowflake

**Methods:** Statistical Analysis, Predictive Modelling, Linear/Logistic Regression, Random Forest, Data Mining, Natural Language Processing (NLP), Neural Networks

**Languages:** Hindi, Telugu

## PROFESSIONAL EXPERIENCE

### University of North Carolina - Chapel Hill

Chapel Hill, NC, USA

Data Analyst/AI-ML developer

February 2023 - Present

- Spearheaded implementation of NLP pipelines in TensorFlow & BERT analyzing thousands of patient safety reports at UNC-HEALTH enabling real-time harm prevention alerts that reduced events 19%
- Developed CNN and BiLSTM (ML) models in PySpark on AWS EMR clusters for context-aware analysis of for adverse event report analysis, decreasing turnaround time from 14 days to 2 hours
- Created ML analytics Tableau dashboards improving workflow efficiency 46% and physician productivity

### Kathmandu University

Bharatpur 44200, Nepal

Physician (Equivalent to MD in the USA)

May 2018 - May 2022

- Experienced Medical Doctor with a track record of providing exceptional care to over 10,000 patients, showcasing a deep commitment to patient health and well-being

## PROJECTS & OUTSIDE EXPERIENCE

### Orchestrating a Scalable Data Platform for COVID-19 Insights

- Built end-to-end cloud data pipeline on AWS ingesting multi-modal COVID-19 data into data lake architecture leveraging Lambda, Glue, Athena
- Designed schema-on-read data processing layers applying PySpark/Pandas for efficient data preparation and feature engineering
- Developed ML models using SageMaker for epidemiology analysis predicting outbreak trajectories based on mobility patterns

### Designing a Web-based Intelligent Medical Diagnosis Application

- Architected a web application leveraging TensorFlow Keras models accurately classifying 7 disease types from patient symptoms/profiles
- Curated dataset of 50,000 medical records applying NLP techniques for feature extraction - tokenization, vectorization
- Developed deep neural network architectures fine-tuned using transfer learning from BERT clinical models pretrained on hundreds of notes
- Performance matched expert clinicians reducing misdiagnoses by 29% demonstrated using rigorously documented testing methodology

### Designing a SQL Database and Dashboards for a Pizza Restaurant

- Designed end-to-end SQL database solution on PostgreSQL capturing customer, order, payment and menu data across multiple locations with appropriate constraints and relationships.
- Developed 100+ SQL (Basic to advanced) queries for aggregate analytics combined with Tableau dashboards creating visibility into sales patterns, peak hours, customer loyalty and product affinity analysis

### Building an Operational Dashboard for Emergency Room Performance

- Developed Tableau dashboard with statistical charts on multiple hospital ER performance - patient volume trends, waiting times, outcome KPIs over 5 years of granular data
- Combined staffing data-modelling with narrative visualizations on root causes behind patient experience score fluctuations

## EDUCATION

### University of North Carolina - Chapel Hill

January 2023 - May 2024

Master's, Biomedical & Health Informatics (MPS BMHI)

GPA: 4

### Kathmandu University

Bachelor's, MBBS