

# PARV THAKKAR

parvthakkar212@gmail.com | +1 (617) 749-6011 | LinkedIn: @parv2 | GitHub: @parv212 | Boston, MA

## EDUCATION

**Master of Science, Data Science** Sep 2021 - Dec 2023(Expected)  
**Northeastern University**, Boston, MA GPA: 3.91 / 4.00

Relevant Coursework: Machine Learning, Data Management and Processing, Database Management Systems

**Bachelor of Technology, Information Technology** Jul 2017 - Jun 2021  
**Pandit Deendayal Energy University**, Gujarat, India GPA: 3.81 / 4.00

Relevant Coursework: Natural Language Processing (NLP), Artificial Intelligence (AI), Data Structures and Algorithms

## TECHNICAL SKILLS

**Programming Languages** : Python, R, SQL, NoSQL, C, C++, Java, Shell Script  
**Cloud-based Technologies** : Amazon Web Services (AWS), Microsoft Azure, MySQL, PostgreSQL, SQL Server  
**Data Visualization Tools** : Tableau, Power BI, Looker, Matplotlib, Seaborn, Plotly, Ggplot  
**Data Science Techniques** : Data Governance, Data Mining, Data Integration, Statistics, Cross-validation, Exploratory Data Analysis, Hyperparameter Tuning, Predictive Analytics  
**Data Engineering Techniques** : Data Warehousing, ETL, PySpark, Hadoop, Hive  
**Tools/Applications** : Scikit-Learn, TensorFlow, PyTorch, SciPy, Advanced Excel, RStudio

## PROFESSIONAL EXPERIENCE

**Data Analytics Intern, Point Focal** Jan 2023 - Present

- Spearheaded development in API code by **98%** by implementing **asynchronous programming**.
- Collected, cleaned, and converted **1,000,000 rows** of raw data into database models, and analyzed in **TIBCO Spotfire**.
- Integrated **natural language narratives** into analytic content in 3 projects using **Arria Studio**.
- Implemented scripts in **AWS Lambdas**, and built **ETL pipelines**, to consume new data sources having **3,000,000 rows**.

**Teaching Assistant, Northeastern University** Sep 2022 - Dec 2022

- **Guided 450+** students in Python Programming and Pandas, resulting in a **35%** increase in performance and improved practical knowledge of programming.

*Used Technologies: Piazza, Gradescope*

**Machine Learning Intern, TomTom** Jan 2021 - Jun 2021

- Deployed a **Machine Learning (ML) Model** with Streamlit and leveraged **Microsoft Azure** to extract, clean and geotag text from images and videos with an accuracy of **90%**. Also, updated the database utilizing **PostgreSQL**.
- **Developed an algorithm** utilizing proprietary probing technology in a span of 2 months to assign road directions for autonomous vehicles and driving assistance.

*Used Technologies: GDAL, Rasterio, GeoPandas and OpenCV*

## ACADEMIC PROJECTS

**T20 Cricket Dashboard** (Pandas, NumPy, PowerBI, DAX, Power Query, JSON) GitHub

- Designed an **interactive dashboard** showcasing the top 5 cricket performers for each role, **integrating 20 new measures** and a customized tooltip for each report, with the help of **Power BI**.

**PubMed Data Analysis** (R, SQL, ETL Pipeline, Data Manipulation, MySQL, Tidyverse, Dplyr) GitHub

- Extracted, transformed, and loaded XML data from a file to an **AWS data warehouse** utilizing a **star schema**, resulting in a **6x faster** execution and retrieval time.

**Predict Patient Outcomes** (Feature Engineering, Model Selection, TensorFlow, LSTM, RNN, Python) GitHub

- Teamed with three students to build a **Deep Learning Model** employing **Bio-Signals and Electronic Health Records** to predict patient conditions and needs, with an average accuracy of **86%** for 3 out of 4 predictions.

**Boston Crime Analysis** (Time Series Modeling, Model Evaluation, Statistics, Python) GitHub

- Demonstrated a **Time Series Model** to forecast crime occurrences with a **91% accuracy** using **ARIMA** and **SARIMA**.

**AI Face Recognition** (CNN, OpenCV, Keras, Data Validation, Data Gathering, Python) GitHub

- Presented a **Convolutional Neural Network Model** identifying classmates' faces with **96%** accuracy and returning relevant images associated with the matched student, with the help of Computer Vision and Neural Networks.