

ASHWIN JOHN CHEMPOLIL

+1 (857) 763-7980 | Boston, MA | chempolil.a@northeastern.edu | [GitHub](#) | [LinkedIn](#) | [ashwinchempolil.me](#)

EDUCATION

Master of Science in Data Analytics Engineering

Sep 2019 - Dec 2021

Northeastern University, Boston, MA

GPA: 3.86/4.00

Coursework: Machine Learning in Engineering, Neural Networks and Deep Learning, Data Mining in Engineering, Data Management and Database Design, Engineering Probability and Statistics, Fundamentals of Cloud Computing

Bachelor of Technology in Production Engineering

Aug 2014 - May 2018

University of Kerala, India

TECHNICAL SKILLS

Programming Languages: Python, SQL, R

Tools: Tableau, MySQL Workbench, MongoDB, Git, R Studio, Jupyter Notebook, R Shiny

Libraries: pandas, scikit-learn, TensorFlow, NumPy, seaborn, matplotlib, NLTK, PyTorch, transformers, ggplot2, dplyr

Certifications: AWS Academy Graduate – AWS Academy Cloud Foundations (AWS), Tableau 10 A-Z (Udemy)

EXPERIENCE

Graduate Course Assistant | Northeastern University, Boston

Sep 2021 - Dec 2021

- Mentor a graduate class of 49 students by conducting weekly office hours, tutoring sessions, and grading coursework
- Assist professor in timely assessment and provided relevant feedback to students to support continuous learning

Data Scientist Intern | Active.ai, Edison, NJ

Jan 2021 - Jul 2021

- Created an EDA tool to explore redundant data and implemented various data preprocessing methods and effectively reduced the size by **70%** on a financial query dataset of **500,000** rows and **42** different labels
 - Analyzed 4 different Question-Answering models (**GPT-3**, **distilBERT**, **BERT** and **T5**) for extractive summarization task on different aspects of a product on Amazon Product Review dataset and interpreted the results to upper management
 - Fine-tuned **distilBERT** for **NER** tasks and achieved an F1 score of **99.57%** on a financial dataset of more than **35,000** queries for classifying **25** different entities
-

ACADEMIC PROJECTS

[Cloud/AWS] Serverless Object Detection App

Jul 2021 – Aug 2021

- Coordinated and designed a serverless architecture to automate the inspection process in a widget manufacturing process using various AWS services based on best practices
- Created a lambda function that gets triggered when an image is uploaded in the S3 bucket, to analyze the image using Amazon Rekognition and send the results to the Quality Assurance team using Amazon SNS

[ML] Prediction and Analysis of Customer Behavior in a Telecommunications Company

Jul 2020 – Aug 2020

- Applied advanced statistical methods (**LassoCV**, **Random Forest**) to understand top **18** features that contribute towards customer dissatisfaction to ensure fact-based decision making
- Deployed **6** machine learning classification algorithms to predict the customer's willingness to churn and achieved **F1 score of 0.67**, **AUC of 0.86** and **log loss error of 0.40** for the optimal model (**Logistic Regression**)

[ML] Analysis and Prediction of NYC Taxi Rides

Jan 2020 – Apr 2020

- Conducted and illustrated various statistical analysis exploring different seasonal trends on Taxi fare and ridership density in 5 different boroughs
- Trained and tested 3 ML models (**Multi Linear Regression**, **Random Forest Regressor**, **Gradient Boosting Regressor**) to predict the taxi fare and obtained an optimal model (Gradient Boosting Regressor) that has a **RMSE of 2.67** and variance of **0.258**

[Database] Medical Insurance Database

Mar 2020 – Apr 2020

- Conceptualized a **Relational Database** and an **ER** (Entity Relationship) Diagram to map out the relationship between **8** different tables of the Medical Insurance Database and achieved normalization to ensure data quality
- Developed an **ETL** pipeline to extract, transform and load live tweets of Insurance Providers to **AWS RDS** using sqlalchemy and tweepy library
- Conducted various **statistical analysis** using SQL queries and optimized the database by partitioning and indexing