

Udip Bohara, Data Scientist

Links : [Email](#) | [LinkedIn](#) | [Portfolio](#) | [Github](#)

EDUCATION

Mercyhurst University

Erie, PA

Master of Science in Data Science | GPA 4.0 | Awarded Full Tuition Waiver

Jan. 2019 – Dec. 2020

Mercyhurst University

Erie, PA

Bachelor of Science in Biostatistics/Public Health

Aug. 2013 – Dec 2017

RELEVANT COURSES

Probability and Statistics, Algorithm Development, Data Structures and Algorithms, Data Wrangling, Relational and Non-Relational Databases, Healthcare Analytics, Big Data Analytics, Machine Learning, Text Mining, Research Methods, Research Project, Data Visualization, Geospatial Data Analytics, Biostatistics I and II, Principles of Epidemiology I and II

EXPERIENCE

Graduate Research Assistant

Aug. 2019 – Present

Department of Computer Science, Mercyhurst University

Erie, PA

- Regularly researched and implemented Machine Learning (ML) methods for research in multiple domains.
- Researched and improved existing ML Intrusion Detection Systems (cyber-attack prevention) by reducing run-time complexity using dimensionality reduction (PCA) and n-gram modeling.
- Mined Twitter data and applied Natural Language Processing (NLP) and Neural Networks to identify key factors that influences cognitive decision-making.
- Wrangled, analyzed and modeled history data from web-browsers and deployed an interactive GUI application for inter-department students to use with Django, D3, Python and SQL.
- Boosted prospect students conversion by utilizing Decision Trees, Clustering and Random Forest models.
- **Teaching Assistant:** Conducted lectures and graded exams for CIS-200 Linear Data Structures (70+ students)

Student Research Analyst

Feb. 2020 – May 2020

Johns Hopkins Applied Physics Lab

Remote

- Worked in John Hopkins program of Forecasting Counterfactuals in Uncontrolled Settings (FOCUS).
- Provided continuous analysis and recommendations for intelligence/conflict strategies using hypothesis-testing.

Data Scientist

Jan. 2019 — May 2020

Department of Institutional Effectiveness, Mercyhurst University

Erie, PA

- Cleaned and migrated data pipelines (ETL) from Ellucian Colleague to Google Cloud Platform (BigQuery).
- Continuously utilized PySpark and SQL for effective querying, ad-hoc analysis, KPI identification and modeling.
- Developed and optimized highly interpretable institutional reports to tackle issues such as risk-modeling, cost minimization, optimized scheduling, etc using Python and Statistical Modeling.
- Deployed a classification Retention/Churn model using Python and Flask by identifying at-risk students.
- Analyzed and identified trend of Grade-Inflation with hypothesis testing and statistical techniques.
- Regularly presented and communicated key findings using Matplotlib, Seaborn, Google Data Studio, Plotly, etc to Provosts and Deans of the University to successfully build intervention and recommendation models.

HIGHLIGHTED PROJECTS | [LINK TO MORE PROJECTS](#)

ArXiv Papers Recommendation System | *PySpark, Neo4j, Gephi, Graphframes, Google Cloud Platform, sigma.js*

Built end-to-end Natural Language Processing (topic modeling and semantic/cosine similarities) pipeline with Graph Database Modeling (Neo4j) to model recommendation systems from arXiv papers.

Electricity Demand Forecasting | *Python, Dash, Keras, Heroku, MongoDB*

Modeled advanced econometric forecasting models such as SARIMAX and Prophet and deep learning methods such as Dilated-CNN and LSTM for Electricity Demand in the USA. Deployed Dash App to display interactive and live results.

Optical Character Recognition (OCR) | *Python, PyTorch, Tesseract*

Developed scalable end-to-end extraction of information from receipts using OCR and semi-supervised deep learning with Graph Convolutional Networks.

TECHNICAL SKILLS

Languages: Python, PySpark, R, MySQL, Mongo Query Language, Cypher/GraphQL, JavaScript, HTML/CSS

Platforms/Frameworks: Flask, Django, Dash, d3.js, jQuery, Spark, Hadoop, MongoDB Atlas, Neo4j, Microsoft Office

Tools: Git, Databricks, Google Cloud Platform, Azure, Databricks, Jupyter, SPSS, Weka, Rapidminer, ArcMap, Tableau

Libraries: Pandas, NumPy, Matplotlib, Plotly, Scikit-learn, Statsmodels, PyTorch, spaCy, gensim, Tensorflow, Tidyverse