# Udip Bohara

Email | LinkedIn | Portfolio | Github

## Education

**Mercyhurst University** — Erie, PA
*Master of Science in Data Science | GPA 4.0* — *Jan. 2019 – Dec. 2020*

**Mercyhurst University** — Erie, PA
*Bachelor of Science in Biostatistics/Public Health* — *Aug. 2013 – Dec 2017*

## Experience

**Graduate Research Assistant** — Aug. 2019 – Present
*Department of Computer Science, Mercyhurst University* — *Erie, PA*

- Researched and built novel methods of Intrusion Detection System in cybersecurity using deep learning and Natural Language Processing technologies (n-gram modeling with PCA).
- Mined and analyzed Twitter data to apply machine learning techniques (Natural Language Processing, Convolutional Neural Network and other classification algorithms) to identify key factors that affect cognitive decision-making.
- Analyzed and wrangled history data from web-browsers to visualize it and built an interactive GUI application for inter-department students to use with Django, D3, Python and SQL.
- Built classification models to analyze and boost prospects conversion for the department.
- **Teaching Assistant**: making lectures, teaching and grading for CIS-200 Linear Data Structures (70+ students)

**Student Research Analyst** — Feb. 2020 – May 2020
*Johns Hopkins Applied Physics Lab* — *Remote*

- John Hopkins program of Forecasting Counterfactuals in Uncontrolled Settings (FOCUS)
- Worked on Simulation-based hypothesis-analysis to develop counter-factual predictions.

**Data Scientist** — Jan. 2019 — May 2020
*Department of Institutional Effectiveness, Mercyhurst University* — *Erie, PA*

- Cleaned, migrated and maintained data from Ellucian Colleague to Google Cloud Platform (BigQuery) for effective ad-hoc analysis and modeling.
- Developed highly interpretable ad-hoc institutional reports using Python to be presented to key stakeholders (Provosts and Deans of the University)
- Applied wrangling, visualization and machine learning techniques such as SMOTE, Decision Trees, Regression models and other ML technologies.
- Produced pragmatic solutions and intervention models to institutional problems such as Grade Inflation and Retention using Statistical Inference and ML.

## Personal Projects

**ArXiv Recommendation System** | *PySpark, Neo4j, Gephi, Graphframes, sigma.js*

- Utilized Natural Language Processing (topic modeling and semantic/cosine similarities) along with Graph theory (eg.centrality) to model recommendation systems based on open data from ArXiv utilizing Google Cloud Platform
- Visualized Results utilizing Gephi and sigma.js for usability and interpretability.

**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 Network.s

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

- Compared advanced forecasting models such as SARIMAX with deep learning methods such as Dilated-CNN and LSTM for Electricity Demand in the USA.
- Utilized multiple APIs to pull live data from multiple sources and developed an interactive Dash Application to be deployed in Heroku with live updates and visualizations with a backend database in MongoDB Atlas.

## Technical Skills

**Languages**: Python, PySpark, R, MySQL, Cypher, JavaScript, HTML/CSS
**Platforms/Frameworks**: Flask, Django, Dash, d3.js, jQuery, Apache Spark, MySQL, MongoDB Atlas, Neo4j
**Tools**: Git, Databricks, Google Cloud Platform, Azure, Databricks, Jupyter, SPSS, Weka, Rapidminer, ArcMap, Tableau
**Libraries**: Pandas, NumPy, Plotly, Matplotlib, Scikit-learn, Statsmodels, PyTorch, Tensorflow, Tidyverse, ArcPy