Udip Bohara

Ph: 814-881-0523 | Email | LinkedIn | Portfolio | Github

Education

Mercyhurst University — Erie, PA
Master of Science in Data Science | GPA 4.0 — *Jan. 2019 – Dec. 2020*
Awarded Graduate Assistantship with full tuition waiver and stipend for academic excellence.

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

Experience

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

- Researched 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

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

- Assist professor in teaching/grading/proctoring and holding office/mentoring hours for the classes CIS-200 Linear Data Structures and CIS-100 Intro to Computer Science that consists of 70+ students

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

- Cleaned and migrated 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 (eg. Provosts and Deans of the University)
- Applied wrangling and machine learning techniques such as SMOTE, classification algorithms to produce pragmatic solution to insitutional problems such as Grade Inflation and Student Retention.
- Important projects include detailed analysis/modeling of student retention and grade inflation.

Projects

ArXiv Recommendation System: On-Going | *PySpark, Neo4j, Gephi, Graphframes, sigma.js*

- Utilize NLP methods along with Graph theory to model recommendation systems based on open data from ArXiv utilizing Google Cloud Platform
- Visualize Results utlizing 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 Networks

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 APIs to pull live data from multiple sources and developed an interactive Dash Application to be deployed in Heroku with live updates and visualizations.
- Built a backend database in MongoDB Atlas for effective storage and analysis.

Technical Skills

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