Udip Bohara

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*

- 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, KPI identification 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/predictive modeling techniques such as SMOTE, Decision Trees, Regression Models, Risk-Modeling, Churn Analysis and other ML technologies.
- Produced pragmatic solutions and intervention models to institutional problems such as Grade Inflation, Financial Forecasting and Student Retention using Statistical Inference and ML.

Personal Projects - Links

ArXiv Papers Recommendation System | *PySpark, Neo4j, Gephi, Graphframes, Google Cloud Platform, sigma.js*
Utilized end-to-end Natural Language Processing (topic modeling and semantic/cosine similarities) along 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. Built a 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, JavaScript, HTML/CSS
Platforms/Frameworks: Flask, Django, Dash, d3.js, jQuery, Spark, 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