ANAM IQBAL

Pittsburgh, PA | +1 (646) 207-4431 | anami@andrew.cmu.edu | LinkedIn | Github

EDUCATION

CARNEGIE MELLON UNIVERSITY (CMU) Pittsburgh, PA

Master of Information Systems Management: Business Intelligence & Data Analytics (Focus in Data Science) December 2022

Relevant Coursework**:** Introduction to Deep Learning, Big Data & Large Scale Computing, Unstructured Data Analytics, Machine Learning for Problem Solving, Interactive Data Science, Distributed Systems

LAHORE UNIVERSITY OF MANAGEMENT SCIENCES (LUMS) Lahore, Pakistan

Bachelor of Science (Honors) - Management Science June 2018

SKILLS

Functional: Machine Learning, Deep Learning, Natural Language Processing (NLP), Artificial Intelligence, A/B Testing, ETL, Data Visualization, Data Engineering, Statistics, Regression, Clustering, Statistical Modeling, Computer Science

Languages: Python, SQL, Java, R, Stata

Tools: Pandas, NumPy, Scikit-learn, PyTorch, SciPy, TensorFlow, Matplotlib, Git, Tableau, Streamlit, Seaborn, AWS, Spark, Hadoop, Altair, R Shiny, R Markdown, ggplot2, leaflet, Power BI, MS Excel, Jupyter Notebook

Business: Collaborative, Attention to Detail, Teamwork, Problem Solving, Organizational Skills, Communication

WORK EXPERIENCE

HITACHI ENERGY Raleigh, USA

Data Science Intern May – August 2022

- Drove increased service revenue by creating a supervised + unsupervised machine learning text and numeric matching pipeline in Python that identifies matching features for generator and bus mapping; The pipeline feeds in rough mappings and SME-defined indicators to a Linear Regression model to identify custom weights, which are then used for final calculation score; Improved mapping by 7% as compared to legacy processes and saved hours of tedious manual effort in matching text
- Experimented with multiple text matching algorithms like Jaccard Similarity, Levenshtein distance, fuzzy matching and Jaro distances; and text-based features like Parts-of-Speech (POS) tagging and Named Entity Recognition (NER) for text matching
- Developed and deployed a first-of-its-kind python-based algorithm design by collating large volumes of data from clients' database/ERP systems, cleaning, transforming it, normalizing it, and calculating complex variables contributing to capacity price forecasting attributes of over 25 years, which helped in making better revenue and investment decisions for power plants
- Presented project findings to technical and non-technical stakeholders in a succinct manner and incorporated feedback

ENERGY INFORMATICS GROUP, NATIONAL CENTRE IN BIG DATA & CLOUD COMPUTING Lahore, Pakistan

Data Analysis Research Associate November 2019 – April 2021

- Experimented and prototyped automation of operational dashboard on users' smart meter activity in 6 different areas of Lahore; eliminated human error and saved 60+ hours of manual effort per month in collecting meter readings via testing
- Designed and used A/B testing to optimize the widgets and user experience (UX) of the dashboard
- Collaborated with government client to visualize live status updates of neural network based short term load predictive model for Pakistan's electricity grid and enhanced business performance through online issue reporting efficiency
- Created real-time alerting time series classification reporting system, using R Markdown, of electricity consumption at a software company and enabled cost optimization by highlighting unusual energy consumption activity

PROJECT EXPERIENCE

Disaster Tweet Prediction | *project link* July – August 2022

- Developed a Natural Language Processing (NLP) pipeline which involves text cleaning (removing stop words and punctuations, tokenization, lemmatization), converting text to features, and training a classifier to predict whether a tweet is actually talking about a disaster or not to achieve an F1 score of 79.7% on the Kaggle leaderboard for a novel approach.
- Experimented with different techniques like TF-IDF, word2vec, POS tagging, BERT embedding to featurize text.

Predict Churning Customers | *project link* January 2022

- Conducted data preprocessing, exploratory analysis, and feature selection for multiple features
- Experimented with various classification models and selected the best model using cross-validation based on F1 Score
- Evaluated customer churn by using AI model algorithms like XGBoost, Logistic Regression and SVM; finalized on XGBoost ml algorithm with an f1-score of 93% to communicate business strategy and prevent churn

Climate Change Tracker | *application link* January – May 2022

- Designed and deployed a novel data science interactive application on visualization software i.e. Streamlit that enables the exploration of multiple climate datasets with interactive visuals and enables users to track the impact of certain hyper parameters on the ml models in real time
- The app helps track climate change trends and patterns for countries, enabling preemptive action and improved policy making

Calorie and Workout Plan Recommender Application | *demo link* August – December 2021

- Engineered a recommender application in python which helps users determine their calorie intake through API & JSON files, and automatically generates customized weekly meal and workout plans PDF based on metrics using their physical parameters