SYED DANISH AHMED

+1 (412) 708 5459 | syeddanish@cmu.edu | linkedin.com/in/syeddanishahmed/ | github.com/SyedDanishAhmed

Education

CARNEGIE MELLON UNIVERSITY — Pittsburgh, PA
Master of Information Systems Management, BIDA – Data Science, **QPA: 3.89 (High Distinction)** — Dec 2020
Relevant Courses: Machine Learning, Artificial Intelligence, Unstructured Data Analytics, Statistics, Econometrics, Business Analytics

S.G.S.I.T.S — Indore, India
Bachelor of Engineering in Information Technology – Computer Science, **GPA: 3.81 (First Class with Distinction)** — Jun 2015
Relevant Courses: Data Structures, Algorithms, Database Management, Data Mining, Linear Algebra, Distributed Systems

Skills

- **Machine Learning:** Regression, Classification, Segmentation, Dimensionality Reduction, Ensemble, Deep Learning - DNN, CNN, RNN
- **Functional:** NLP, Anomaly Detection, Time Series Forecasting, Hypothesis Testing, Power Analysis, A/B Testing, Agile, ETL, Web Scraping
- **Languages:** Python, Java, R, SQL, HTML, CSS, PowerShell, HiveQL
- **Business Intelligence and Data Visualization:** R Shiny, Tableau, Power BI, Excel VBA, Matplotlib
- **Data Engineering:** Message Queue, PySpark, Docker, Flask, Jenkins, Unix, Hadoop, MongoDB, Redis, Azure - ML, ADF, Blob, AWS - EC2, S3

Academic Experience

CARNEGIE MELLON UNIVERSITY — Aug 2019 – Present

- Developed a model for Musculoskeletal Disorder physiotherapy exercise recognition using sensor data from accelerometers
  - Generated features by windowing with overlap and used Discrete Cosine Transform (DCT) for feature transformation and selection
  - Implemented a **1D Time Distributed CNN – LSTM** architecture using **Keras and Tensorflow** & achieved AUC of 97% over seven exercises
- Implemented a **Multilayer Perceptron (MLP)** classifier for text-based Emotion Classification using **Word2vec word embedding**
- Developed a **scalable Logistic Regression model** for Bank Marketing using **PySpark SQL & ML** to predict client subscription to term deposit
- Created a **Tableau Dashboard** for tracking COVID-19 cases across the world
- Built a Landmark Classification model by **Transfer Learning** using a pre-trained **ResNet18** CNN in **PyTorch** with 85% accuracy over ten classes
- Implemented ML algorithms - **Decision Trees, K-means, Logistic Regression and Neural Network with SGD** from scratch in **Python**

Work Experience

PricewaterhouseCoopers (PwC) DT&I Center — Pittsburgh, PA
Graduate Applied Researcher – Data Science — Jun 2020 – Dec 2020

- Developing Machine Learning models for **detecting fraud or corruption** on large volumes of General Ledger transactional dataset
- Exploring **unsupervised feature selection** – Statistical Tests (ANOVA, Kruskal-Wallis), Dimensionality Reduction, Cardinality Reduction
- Conducting research on **unsupervised anomaly detection** - **RBM**, **Autoencoder, Clustering** and anomaly score-based **ensemble** techniques

MU SIGMA: Data Science provider for 140 Fortune 500 Companies — Bengaluru, India
Data Scientist — Oct 2015 – May 2019

- Implemented **Named Entity Recognition (NER)** model using **Bidirectional LSTM** and **ELMo embeddings** for a Conversational AI system
  - Identified and extracted entities from conversation transcripts with an accuracy of 81% and deployed as a **REST API** service using **Flask**
- Developed & deployed a **Shiny** app on **RStudio Server** to visualize high-dimensional data using **Voronoi Tessellations** & **Sammon Projection**
- Created a **real-time anomaly detection** system for **predictive maintenance** of A/C units – **density-based clustering & polynomial regression**
  - $60M+ cash flow generated for the client through subscription fees and reduction of truck rolls for Contractors
- Surveyed ML models - **SVM**, **Naive Bayes**, **Decision Trees**, **Random Forest**, **XGBoost** with **PCA** on **TF-IDF** features for Spam detection model
  - Resampled the dataset using **under-sampling** and **over-sampling** techniques like **SMOTE** for countering class imbalance
  - Created a **Docker** container for **Flask API**. Hosted **Jenkins** on **AWS EC2** instance & added Jenkins webhook to GitHub for CI/CD pipeline
- Automated the annotation process for a Customer Support Chat-Bot using **Topic Modeling** algorithms - **Latent Dirichlet Allocation (LDA)**
  - 27% more unanswered texts mapped to their correct intents and manual effort reduced by 40 man-hours per month
- Developed and deployed a **real-time microservice** for auto-releasing incorrectly flagged financial transactions (SWIFT messages)
  - Built a distributed system using **Message Queues** and **Two-Phase Locking** for reviewing transactions using **Cosine Similarity** metric
  - 20% reduction in false positives which led to a reduction of 1.8 man-hours of manual effort per day
- Implemented **ARIMA time-series forecasting** to predict the baseline Spend dollar value for **Search Advertisers** with MAPE less than 9%
- Developed a **time-series anomaly detection** model for detecting Bot traffic to prevent **Click Fraud** in **Search Syndication** with 78% accuracy

Awards & Recognition

- Won the 3rd prize in Deloitte Case Challenge 2019 organized at Carnegie Mellon University — September 2019
- Dean's List for outstanding academic performance at Carnegie Mellon University — June 2020
- Working as a Graduate Teaching Assistant for Intermediate Statistics course under Professor Daniel Nagin — Fall 2020
- Received multiple Mu Sigma Spot Awards for showcasing quality and timeliness in project execution — 2016, 2017, 2018