Vishva Shah

vshah69@uic.edu | 773-790-7379 | GitHub | Portfolio | LinkedIn | Medium | Chicago, IL, USA

Education

University of Illinois at Chicago — Major GPA: 4.0/4.0
Master of Science in Business Analytics (STEM), Expected May 2021 — GPA: 3.87/4.0
Data Science Courses: Deep Learning, Data Mining, Advanced Statistics, Machine Learning, Time Series, Big Data, Network Analysis, ML Deployment

Institute and Faculty of Actuaries (UK) — Apr 2014 - Apr 2018
Professional degree in Actuarial Science: 9 exams
Major Courses: Probability distributions, Statistical and forecasting models, Financial mathematics

University of Mumbai (India) — Apr 2018 – Sept 2019
Master of Economics & Commerce

University of Mumbai (India) — Apr 2013 – Apr 2016
Bachelor of Economics & Commerce

Skills

Certifications	Udacity AWS Machine Learning, June 2020
Technical	Python, R, AWS, Pyspark, Hive, Tableau, VBA, MySQL, HTML, Git, Microsoft Office
Machine Learning	Data Mining, Decision trees, Cluster analysis, Deep Neural Networks, Probability networks, NLP, Ensemble methods, Random forest, Gradient Boosting, SVM, KNN, Association rules
Statistics	Hypothesis testing, Naive Bayes, Regression, Markov-chain, Confidence-intervals, Time Series

Projects

Image Classification on FER 2013 [GitHub repo] — Spring 2020
Built an image classifying model with Pytorch, Scikit-Learn, NumPy, SciPy, Pandas, Pickle, MLlib, OpenCV

- Performed data augmentation by transforming images including rotating, mirroring, cropping, and padding which increased training data by ~35% , reducing overfitting by ~25%.
- Customized VGG16 architecture with required outputs achieving ~85% accuracy over baseline CNN Model with ~60%.
- Enhanced the accuracy to ~89% after appropriate tuning of hyper-parameters and error analysis with confusion matrix and F1-score metrics.

Generative Adversarial Networks [GitHub repo] — Spring 2020
Built GANs on the Pokémon data set using Pytorch, Scikit-Learn, NumPy, SciPy, Pandas downloaded from Kaggle

- Experimented Amazon's deepcomposer AI frameworks and implemented GANs on the unique image data set of over 1000 images.
- Augmented data by normalizing, centre-cropping, flipping, mirroring which increased training set and reduced overfitting by ~15%.
- Custom built robust end-to-end deep learning GAN architecture using CNN and CNN-transpose for generator and discriminator.
- Used binary cross entropy as a loss function, hyper-parameters were tuned by trial and error and evaluating using recall of ~90%.

Manipal Hospital Harvard business review Case study [GitHub repo] — Fall 2019
Developed supervised machine learning models for NPS score by ggplot2, rpart, randomforest, GBM, ROCR, Caret

- Created data exploration, cleaning to include removal of redundant columns and imputation of missing values for 40K rows; and SMOTE to balance data.
- Designed stacked ensemble models architecture with Random forest and Gradient boosting, after reduction in dimension from 200 to 60 using PCA.
- Evaluation - Tuned hyper-parameters using K-fold validation, confusion matrix, ROC curve; test accuracy was ~88.5%.

Work Experience

University of Illinois at Chicago (CAA research) | USA — May 2020 – Present
Graduate Research Assistant

- Parallelized extraction using map-reduce algorithm and speed up the process of modelling and training by ~65%.
- Conducted data pre-processing for massive data, reduced the dimensions using PCA, RIDIT transform, and extracted meaningful metrics in PostgreSQL & Python.
- Streamlined a bi-directional LSTM for anomaly detection, and pattern recognition which resulted in ~60% better accuracy than previously assessed models.
- Deployed the ML model using AWS and speed up the pipeline by 10X through vectorization and GPU/CUDA processing.

ICICI Prudential Life Insurance Company Ltd. | India — Jan 2018 - July 2018
Actuarial Data Analyst

- Leveraged a hypothesis-driven approach to extract statistical insights using Python, giving ~95% accurate distribution of the demographics of customers for targeting.
- Optimized visual data and created real-time dashboards in Tableau that enhanced sales by ~25%.
- Built a propensity model for the sales team to predict the likelihood of an opportunity to convert from early sales using sentiment of customer and history with client.

Reliance Nippon Life Insurance Company Ltd. | India — Nov 2016 - Jan 2018
Actuarial Data Analytics Consultant

- Developed mathematical models for forecasting uncertainty that identified patterns in the data using time series models to include ARIMA, exponential smoothening resulting in ~87% accuracy.
- Conducted cluster analysis using K-means to generate segmented profiles of customers using 350K rows of claims data.
- Automated the procedure of extracting massive amounts of results using VBA and sped up the process by ~70%.

Leadership Activities

Business Analytics Organization	Corporate Relations Manager, Organizing involvement fairs	May 2020-Present
Volunteer of National Service Scheme	Head of the Golden Jubilee fest, Organized Blood Donation Drives	2014-2016
Core Member of Rotaract Club	Visiting Orphanages, Teaching kids, Plantation drive	2011 – 2012