Vishva Shah

shah.vishva5@gmail.com 773-790-7379 GitHub Portfolio LinkedIn Medium Chicago, IL, USA

Education

Master of Science in Business Analytics (STEM), University of Illinois at Chicago, GPA: 3.87/4.0 — Aug 2019 – May 2021
Courses: Machine Learning, Data Mining, Advanced Statistics, Time Series, Big Data, Network Analysis, ML Deployment, Data Visualization, Advanced Analytics

Master of Commerce & Economics, University of Mumbai (India) — Apr 2018 – Dec 2020

Professional Degree in Actuarial Science, Society of Actuaries (USA) and IFOA (UK) — Apr 2014 - Apr 2018
Courses: Statistics for risk modelling, Mathematical courses, Probability, Actuarial mathematics, Predictive analytics, Investment and financial markets, Economics

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

Skills

Certifications	Udacity AWS Machine Learning, June 2020
Language, Software	Python, R, PostgreSQL, Hadoop, Tableau, AWS, PySpark, MapReduce, Git, PyTorch, TensorFlow, Keras
Data Science	Pandas, NumPy, SciPy, Seaborn, Matplotlib, Plotly, Scikit-Learn, web-scraping (BeautifulSoup), Dplyr, Rstudio
Machine Learning	Natural Language Processing (NLTK, Spacy, Flair, Embeddings), Neural Networks (CNN, RNN-LSTM, GAN), Reinforcement Learning, Unsupervised Learning (Clustering, PCA), Regression (Linear, Logistics, Lasso, Ridge), Classifiers (SVM, K-NN, Decision Trees), Generative (Naïve Bayes, LDA)

Work Experience

Graduate Assistant – Data Science | University of Illinois at Chicago, CAA Research — May 2020 – Present
- Constructed data pipelines for pharmacy data by feature engineering, reduced the dimensions from 150 to 25 using PCA.
- Detected fraud in opioids oversupply in real-time pharmacy data, applied RIDIT transform, and extracted over 25 meaningful metrics using PostgreSQL & Python.
- Streamlined a machine learning pipeline using supervised and unsupervised models to predict the oversupply of opioids resulting in ~89% F1-score.

Actuarial Data Analyst | ICICI Prudential Life Insurance Company Ltd — Jan 2018 - July 2018
- Leveraged a hypothesis-driven approach to extract statistical insights using Python, giving ~95% accurate distribution of the demographics of customers, consumer sentiment or perception, for targeting.
- Optimized visual data and created real-time dashboards in Tableau for claims data, informing about the segmentation of customers and increasing sales in the next quarter.
- 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.

Actuarial Data Analytics Consultant | Reliance Nippon Life Insurance Company Ltd — Nov 2016 - Jan 2018
- Performed market analysis for feasibility, cost, and product placement against competing products and created functional A/B testing to ascertain remedies.
- 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%.

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% accuracy.
- Enhanced the accuracy to ~89% after tuning of hyper-parameters and error analysis with confusion matrix and F1-score metrics.
- Deployed the model on Heroku using Gunicorn, flask for real-time user experience with the model.

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 with Random forest and Gradient boosting, after reduction in dimension using PCA.
- Evaluation - Tuned hyper-parameters using K-fold validation, confusion matrix, ROC curve; test accuracy was ~88.5%.

Publications

Statistical concepts	Numerical measures, Data types, Probability Distributions
Effects of Social Media	Summary of Social Dilemma, Data Privacy

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