Vishva Shah
vshah69@uic.edu | **773-790-7379** | **Chicago, IL**
https://vishva-shahh.github.io/ | https://github.com/Vishva-Shahh | https://www.linkedin.com/in/vishva-shah-0201/ | https://medium.com/@Vishva_Shah

Summary

Analytics professional with 2+ years of experience in building **advanced statistical** and **predictive models**. Constructed data-driven insights after processing large-scale data using **SQL** and implementing **machine learning** algorithms in **Python, R and Tableau.**

Work Experience

Data Science Graduate Assistant | UIC – *Client: Claris Health* | Chicago   Aug 2020 - Present
- Detected fraud in opioids oversupply with Auto-encoders and PCA in Python, eventually saving over *~$10 million* in claims.
- Architected end-to-end **LSTM** algorithm for anomaly detection in *9 billion rows* of claims data to mitigate claim risk by *~25%*.
- Crafted fraud detection system using machine learning in healthcare data in **PostgreSQL** and presented mitigating measures.

Data Science Research Assistant | UIC | Chicago   May 2020 - July 2020
- Applying **NLP** for information retrieval (**NER**) in Python to learn the intent of the input sentence and predict sense of sentence.
- Implemented **NLTK** Flair model after pre-processing the text data using tokenization, stemming and lemmatization.

Actuarial Data Analyst | ICICI Prudential Life Insurance Company Ltd | India   Jan 2018 - July 2018
- Built **time series** propensity models to predict likelihood of an opportunity to convert from early sales, giving *~89%* accuracy.
- Optimized dashboards for the segmentation of customers to target marketing strategies thereby increasing sales by *$8* million.
- **Led** a team of 5 in computing projections, writing, and presenting technical reports outlining risk preventive strategies.

Actuarial Data Analytics Consultant | Reliance Nippon Life Insurance Company Ltd | India   Nov 2016 - Jan 2018
- Devised **cluster analysis** with K-means segmenting customers, conducted monte-carlo simulations, increasing sales by *~20%.*
- Automated the procedure of extracting massive amounts of results using VBA thereby speeding up the process by *~70%*.
- Performed **statistical modelling** for financial predictions, cost-benefit analyses, and risk projections for hedging strategies.

Projects

Image classification and captioning | *Python, ResNet50, LSTM, Flask* [GitHub repo]
- Performed data augmentation (Rotating, Flipping, Mirroring) increasing training data by *~35%*, reduced overfitting by *~25%.*
- Designed the pretrained Resnet50 as encoder, used LSTM as decoder, cross-entropy loss function, mini-batch gradient descent.

Manipal Hospital Harvard Case study | *R, K-means, xgboost, Randomforest, Stepwise regression* [GitHub repo]
- Imputed missing values using KNN, computed key features using t-tests & stepwise regression, outlier detection.
- Coded random forest and xgboost model, tuned parameters with K-fold, validated results with F1 score of *~87%* and AUC/ROC.

Movie Recommendation system | *PySpark, Alternating Least Squares, K-Means clustering*
- Created a big-data processing pipeline with Apache Spark to load *~27 million* user ratings and Implemented collaborative filtering with Alternating Least Squares method and optimized the model to achieve an **RMSE** of *~0.81*.
- Developed TF-IDF vectorizer of the movie descriptions text and created content-based recommendations with K-Means clustering.

Chicago Crime Predictive Modelling | *Python, L2-regularization, Hypothesis testing*
- Devised **L2-regularized** Linear regression with 10-fold Cross Validation, trained using stochastic gradient descent to predict arrests.
- Tuned modelling pipeline and improved performance to achieve *89%* accuracy and minimized False positive with *85%* precision.

Leadership & Publications

Business Analytics Organization [Board of Director] - Leading over 20 grads, providing resources for enhancing their skill set.
Medium publications [Numerical measures, Data types, Probability Distributions], [Summary of Social Dilemma, Data Privacy]

Competencies

Language & Tools	Python, R, PostgreSQL, Hadoop, PySpark, HTML, Tableau, AWS [EC2], MapReduce, Git, Docker, Kubernetes.
Data Science Libraries	Scikit-Learn, NumPy, SciPy, Pandas, MLlib, OpenCV, ggplot2, rpart, randomforests, ROCR, Caret.
Statistics	Inferential Statistics, Predictive Modelling, Hypothesis Testing, Sampling, Regression, Confidence Intervals
Machine Learning	Neural Networks (CNN, GAN), Unsupervised Learning (Clustering, PCA), Regression, SVM, K-NN, Decision Trees, Generative Algorithms (Naïve Bayes, LDA), NLP(BERT, Transformers)

Education

Master of Science in Business Analytics (STEM), University of Illinois at Chicago, **GPA**: 3.84/4.0   Expected May 2021
Courses: Machine Learning, Data Mining, Advanced Statistics and analytics, Time Series, Big Data, Network Analysis, ML Deployment, Data Visualization, Multivariate Analytics | **Awarded Full tuition Scholarship.**

Master of Commerce, University of Mumbai (India)   Dec 2020

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

Bachelor of Commerce, University of Mumbai (India)   Apr 2016

Certifications

Python for Data Science *LinkedIn*, **Advanced Predictive modelling** *LinkedIn*, **AWS Machine learning foundations** *Udacity*