ANKIT SINHA

ankit9sinha@gmail.com | 773-729-9685 | linkedin.com/in/ankit19sinha/ | ankit19sinha.github.io/

EDUCATION

Master of Science, Business Analytics   *Aug 2019 - Present*
University of Illinois at Chicago   GPA: 3.70
Bachelor of Engineering, Computer Engineering   *Aug 2013 - May 2017*
University of Mumbai

EXPERIENCE

Graduate Assistant (Data Science & Analytics), ClarisHealth   *Sep 2020 - Present*
- Developed a fraud detection model using TensorFlow, scikit-learn and SQL, on approximately 1.2 million observations, for health insurance claims that incorporates longitudinal scoring using LSTM autoencoder, and cross-sectional scoring using PCA and RIDIT transform for anomaly detection on a sliding window of 3 months
- Developed features using medical claims data for peer-to-peer analysis of providers categorized by their specialties
- Developed a classification model based on the transition matrix of the temporal graph data structure of medical procedures of various patients to determine the suspiciousness of a sequence of transition of nodes across the graph

Summer Research Intern, University of Illinois at Chicago   *Jul 2020 - Aug 2020*
- Researched on community detection algorithms for directed graphs using networkx and igraph for capturing temporal relations between medical procedures to develop a probabilistic model to identify episodes of symptoms for an individual
- Compared different scoring methods to determine the best metrics for evaluation of communities in a directed network using Pythons cdlib (a niche package for community detection), and scikit-learn

Assistant Systems Engineer / Software Engineer, Tata Consultancy Services   *Jul 2017 - May 2019*
- Developed 100+ stored procedures and triggers using T-SQL for manipulation, aggregation, and analysis of clients HRMS data from across 20 different countries, thereby increasing efficiency of data retrieval and querying by 30%
- Developed 200+ web forms using ASP.NET, C#, HTML, CSS, and JavaScript for data collection and analysis of information of 50,000+ users of clients HRMS portal, thereby improving stakeholder/employee engagement by 150%
- Collaborated with team of developers, business analysts, stakeholders, and technical support from offices across the globe to determine optimal specifications and solutions as per the business requirements

PROJECTS

Sentiment Analysis of Hotel Reviews using Recurrent Neural Networks (RNN)
- Implemented and tuned 3 variants of RNN - namely, vanilla RNN, GRU, LSTM, and attention models for sentiment analysis of 515,000+ reviews of 1,400+ hotels across Europe, and 100-dimensional GloVe representation for embedding
- Models with attention work best with an accuracy of 94% for additive attention and 96% for multiplicative attention, thereby increasing accuracy by 66.67% as compared to models without attention

Target Marketing for Paralyzed Veterans of America
- Developed two different models using glmnet and gbm for calculating the likelihood of response of a user and estimating their respective donation amount from a highly imbalanced dataset with 487 variables and over 95,000 observations
- Algorithms like Random Forest, LASSO regression, and Ridge regression were used on an under sampled training set of 50/50 proportion and their scores were calibrated to account for different baseline rates of the minority class
- Principal Component Analysis was also performed to reduce the number of variables to 80 principal components

Human Activity Recognition
- Developed and compared performance of Decision Trees, SVM, and K-NN algorithms for a multi-class classification on a dataset consisting of 561 attributes and 3,500+ observations
- Used PCA to reduce the the number of attributes to 17 principal components and compared model performance using classification accuracy, with SVM achieving the highest accuracy of 91.97%

LEADERSHIP & HONORS

- Received "The Innovative Mind" award by the client for leveraging technology to provide unique solutions, and better service
- Treasurer and Publicity Co-head for Computer Society of India, University of Mumbai student chapter
- Member of Illinois Technology Association, Operations Management Group, and Enactus UIC chapter

SKILLS

Programming Languages: Python, R, SQL, C#, Java, JavaScript

ML Packages: PyTorch, TensorFlow, scikit-learn, pandas, numpy, PySpark, tidyverse, glmnet, rpart, gbm, recipes
Coursework: Deep Learning, Network Analysis, Big Data Analytics, Text Analytics, Statistical Modelling, Data Mining, Analysis of Algorithms, Data Structures, Object-oriented Programming, Database Management Systems
Data Visualization: Tableau, Power BI, matplotlib, ggplot2, seaborn