ANKIT SINHA

ankit9sinha@gmail.com | asinha33@uic.edu | 773-729-9685 | linkedin.com/in/ankit19sinha/ | ankit19sinha.github.io/

SKILLS

Programming Languages: Python, R, SQL, C#, Java, JavaScript
ML Packages: PyTorch, TensorFlow, scikit-learn, pandas, numpy, PySpark, tidyverse, glmnet, rpart, gbm, recipes
Data Visualization: Tableau, Power BI, matplotlib, ggplot2, seaborn

EDUCATION

Master of Science, Business Analytics - University of Illinois at Chicago, USA   *Aug 2019 – May 2021*
Coursework: Deep Learning, Network Analysis, Big Data Analytics, Text Analytics, Data Mining   GPA: 3.70

Bachelor of Engineering, Computer Engineering - University of Mumbai, India   *Aug 2013 – May 2017*
Coursework: Algorithms, Data Structures, Object-oriented Programming, Database Management Systems   GPA: 3.70

PROFESSIONAL EXPERIENCE

Junior Data Scientist, ClarisHealth, Chicago   *Sep 2020 – Present*
- Developing a fraud detection model for health insurance claims using TensorFlow, scikit-learn and SQL, on approximately 1.2 million observations. It would incorporate 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 graph data structure of medical procedures of various patients to determine the suspiciousness of a sequence of transition of nodes across the graph

Data Science Practicum Intern, Cresco Labs, Chicago   *Jan 2021 – Present*
- Researching methodologies to forecast demand of products at the state, brand, category, and SKU levels based on sales data
- Developing forecasting models for identifying seasonality and trends to predict sales using regression algorithms for causal inference analysis and time series modeling. Planning to present findings to stakeholders for enabling their sales strategies

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

Assistant Systems Engineer / Software Engineer, Tata Consultancy Services, India   *Jul 2017 – May 2019*
- Developed 100+ stored procedures and triggers to increase efficiency of data retrieval and querying by 30% on T-SQL for manipulation, aggregation, and analysis of clients HRMS data from across 20 different countries
- Improved stakeholder/employee engagement by 150% by developing 200+ web forms for data collection and analysis. Used ASP.NET, C#, HTML, CSS, and JavaScript to manage HRMS portal efficiently for 50,000+ users
- Collaborated with cross-functional team of developers, vendors, product managers and stakeholders to understand business requirements and specifications to build solutions for their business problems and generate renewable revenue

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
- Achieved an accuracy of 94% for additive attention and 96% for multiplicative attention, thereby improving accuracy by 66.67% as compared to models without attention

Target Marketing for donation collection 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

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