Ankit A. Sinha

GitHub: https://ankit19sinha.github.io/
LinkedIn: https://www.linkedin.com/in/ankit19sinha/
Email: asinha33@uic.edu | Mobile: 773-729-9685

SKILLS

Programming Languages:	Python, R, T-SQL, MySQL, Java, C#
ML frameworks/packages:	Python – scikit-learn, numpy, pandas, PyTorch, TensorFlow, PySpark R – tidyverse, gbm, recipes, caret, glmnet, rpart
Data visualization:	Tableau, Python – matplotlib, R – ggplot2

EDUCATION

Master of Science in Business Analytics   Aug 2019 - Present

University of Illinois at Chicago (GPA: 3.75)   Expected to graduate in December 2020

Coursework: Deep Learning & Advanced Predictive Models, Analytics for Big Data, Advanced Text Analytics, Data Mining for Business, Statistical Models and Methods, Statistics for Management

Bachelor of Engineering in Computer Engineering   Aug 2013 – May 2017

University of Mumbai, India (CGPA: 8.27/10.00)

PROFESSIONAL EXPERIENCE

Assistant Systems Engineer / Software Engineer   Jul 2017 – May 2019

Tata Consultancy Services | Mumbai, India

- Developed 100+ stored procedures and triggers using T-SQL for manipulation, aggregation, and analysis of Aditya Birla Group's (client) HRMS data from across 20 different countries
- Increased efficiency of data retrieval and querying by 30% by incorporating said stored procedures and triggers
- Developed 200+ web forms using ASP.NET, and JavaScript for data collection and analysis of information of 50,000+ users of Aditya Birla Group's HRMS portal
- Increased the traffic, and improved stakeholder/employee engagement on Aditya Birla Group's HRMS portal by 150% by developing said web forms
- 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

Technologies used: Python (torch, torch.nn, torch.optim, torchtext, argparse, numpy, pandas)

- Implemented and tuned 3 variants of Recurrent Neural Networks - namely, vanilla RNN, GRU, and LSTM for sentiment analysis of 515,000+ reviews of 1,400+ hotels across Europe. Additive and multiplicative attention was also applied to these models. GloVe's 100-dimensional representation is used for embedding. Models with attention work best with an accuracy of 94% for additive attention and 96% for multiplicative attention. An increase of approximately 66.67% in accuracy was observed by incorporating attention mechanisms.

Target Marketing for Paralyzed Veterans of America (PVA)

Technologies used: R (tidyverse, ggplot2, recipes, caret, ROSE, ranger, pROC, glmnet, gbm)

- Developed two different models 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.

NYC Temperature Forecasting

Technologies used: Python (statsmodels, matplotlib, pandas, numpy, scikit-learn)

- Implemented timeseries forecasting model based on Simple Moving Average, Simple Exponential Smoothing, Holt's Linear model, and Holt's Winter. The dataset consists of about 45,000 observations. Model performance is compared using RMSE of predicted data. Holt's Winter had the lowest RMSE of 9.85 (Fahrenheit). Approximately 55% decrease in RMSE was observed by using Holt's Winter as compared to all other methods.