SHOUVIK SHARMA

---

shouvik19@gmail.com ✉

+1-312-459-2008 📱

Chicago 📍

linkedin.com/in/shouvik-sharma19 in

github.com/shouvik19 🎧

medium.com/@shouvik19 Ⓜ

EDUCATION

Master of Data Science
Illinois Institute of Technology
08/2019 – Present
Chicago, USA
- **GPA: 3.71**

Master of Statistics
NMIMS
07/2016 - 04/2018
Mumbai, India
- **GPA: 3.35**

SKILLS

Data Science | Analytics: Linear Regression, Multiple Linear Regression, Logistic Regression, Naïve-Bayes, KNN, Time Series Analysis, AdaBoost, Ensemble Classifier, K- Nearest Neighbor, K-Means Clustering, Hierarchical Clustering, SAS Enterprise Miner, SAS Enterprise Guide, SPSS

Linear Algebra | Statistics: Z-test, ANOVA, Chi-square test

Programming Languages: Python, R, Spark, Hive, Pig

Deep Learning: Convolution Neural Network, Recurrent Neural Network, Long Short-Term Memory Network

Database: SQL Server, Snowflake, PostgreSQL, MSSQL, MYSQL

Tools: Tableau, Power BI, Pentaho, MapReduce, Visual Studio, Prefect, SSIS, SSRS, SSAS, SharePoint

Cloud: AWS Lambda, AWS S3, AWS EC2, AWS CLI, Kafka, Redshift, AWS Sage Maker

Certifications: SAS Certified Base Programmer for SAS 9 in Mar 2017, SAS Certified Predictive Modeler Using SAS Enterprise Miner 14 in Apr 2018, Practical Machine Learning in Dec 2018 from John Hopkins University, Machine Learning Specialization in Feb 2019 from University of Washington

EXPERIENCE

Data Engineer Intern
Daten Solutions Inc.
05/2020 - Present,   *Chicago, IL*
- Performed Data Migration from SQL Server to Snowflake, with dimensional modeling
- Automated ETL processes, making it easier to wrangle data and reducing the time by as much as 40% by performing large-scale data conversions, and transferring BAAN data into standardized formats for integration into Snowflake

Data Analyst
Cartesian Consulting Inc.
04/2018 – 07/2019   *Mumbai, INDIA*
- Identified the trend for improving customer retention and reducing churn rate using statistical techniques like logistic regression, which led to a two-fold improvement in the campaign response
- Performed geography-wise analysis by creating customer one view and customer profiling, and translated this analysis into business terms and actionable guidance
- Identified the 'Most Valuable Customer' by deploying Random Forest algorithm with True positive rate of 81%, this customer identification led to better customer targeting and improve the yearly top-line revenue by 13 %

Strategy and Analytics Intern
Greeksoft Technologies Pvt. Ltd.
09/2017 – 12/2017   *Mumbai, INDIA*
- Worked as a Strategies and Analytics intern to develop a price forecasting model for Greeksoft Technologies Pvt. Ltd. Mumbai
- Built an RNN Neural Network model for Live positional trading using Keras package in python where the outputs supplemented the Bull Spread Strategy in Options Trading, the built model architecture was backtested for the period from 2012-2017 where it achieved correct market prediction in 71 % of the days; this forecasting architecture is used for live trading

Associate Relationship Manager
Tata Capital Financial Services Ltd.
07/2015 – 07/2016   *Mumbai, INDIA*
- Drove acquisition channel of used car and two-wheeler dealership, by building customer scorecard after analyzing different parameters affecting the repaying capacity
- Performed customer risk assessment by analyzing financial reports and client credit history, which led to a multi-fold increase in corporate lending for two-wheeler and used cars segment, with 0% NPA cases reported over the course of 10 months

Projects

- **Stack Overflow Data Analysis (10/2019 -12/2019)** - Derived insights about questions posted by stack overflow by extracting data using Google's big query data warehouse ; identified top spammers, expert users, and most valuable customers users by leveraging big data technologies such as Apache Hive, Apache Pig and Apache Spark
- **Recommendation System using Yelp (01/2020-03/2020)** - Built a personalized restaurant recommender web app using the Yelp dataset of restaurants by testing models like Pure Collaborative, Approximate Nearest Neighbour, K-NN, Naive Bayes and Hybrid Maxtrix Factorization on different hyperparameters which were tuned using the python library scikit optimizer
- **Image Mating using CelebAMask-HQ (10/2019-12/2019) -** Performed Image Matting using the U-Net architecture of the Convoluted Neural Networks on the open-source Celeb-Mask dataset with an IOU Score of 92%
- **Inventory Optimization problem on Kaggle (10/2019-12/2019) -** Performed time series forecasting to forecast the demand for LED televisions using Holt-Winter Smoothing with MAPE of 20.760
- **Book Recommendations from Charles Darwin (07/2020 – 08/2020) –** Built a book recommendation system based on their content using the Charles Darwin's bibliography