SHOUVIK SHARMA

shouvik19@gmail.com
+1-312-459-2008
Chicago
linkedin.com/in/shouvik-sharma19
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, Mathematics

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

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

Database: SQL Server, Snowflake, PostgreSQL, MSSQL, MYSQL, Microsoft SQL Server, Microsoft Visual Studio, Visual Basic

Tools: Pentaho, MapReduce, Visual Studio, Prefect, SSIS, SSRS, SSAS, SharePoint, MS Excel, VBA

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, Snowflake Pro Certification September 2020

Data Visualization: Tableau, R shiny, Power BI

Soft Skills: Integrity, Hands-on approach towards problem solving

PROFESSIONAL EXPERIENCE

DATA SCIENTIST INTERN
Daten Solutions Inc.
05/2020 - Present **Chicago, USA**
- Developed data migration pipeline from SQL Server to Snowflake, and performed dimensional modeling on the migrated data
- Automated ETL processes using Prefect (Python), making it easier to wrangle data sets and reducing time by as much as 40% by performing large-scale data conversions, and transferring BAAN data into standardized formats for integration into Snowflake
- Developed statistical models like ARIMA using statsmodels package in Jupyter Notebook, the model achieved an overall accuracy of MAPE 5.96%

DATA ANALYST
Cartesian Consulting Inc.
04/2018 –07/2019 **Mumbai, INDIA**
- Created monthly reports identifying key metrics in MS Excel by using VBA functions, COUNT-IF, SUM-IF, HLOOKUP and VLOOKUP
- Executed geography-wise analysis by creating customer one view by extracting data from MySQL, and translated analysis into business terms and actionable guidance
- Identified the 'Most Valuable Customer' by leveraging the customer data and deploying Random Forest algorithm with True positive rate of 81%, this led to better customer targeting and improve yearly top-line revenue by 13 %

STRATEGY AND ANALYTICS INTERN
Greeksoft Technologies Pvt. Ltd.
09/2017 – 12/2017 **Mumbai, INDIA**
- Led a price forecasting project for Greeksoft Technologies Pvt. Ltd. Mumbai
- Built an RNN Neural Network model for Live positional trading using Keras package in python where outputs supplemented Bull Spread Strategy in Options Trading, the developed model architecture was backtested for the period from year 2012 to year 2017 where it achieved correct market prediction in 71 % of the days ; this forecasting architecture is utilized 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
- Led a team of 3 to construct 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 (October 2019 -December 2019)** - Analyzed insights about questions posted on stack overflow by extracting large data sets using Google's big query data warehouse ; discovered top spammers, expert users, and most valuable customers users by leveraging big data technologies such as Apache Hive, Apache Pig and Apache Sparks **(git link)**
- **Recommendation System using Yelp (January 2020 – March 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 Matrix Factorization on different hyperparameters which were tuned using the python library scikit optimizer **(git link)**
- **Image Mating using CelebAMask-HQ (June 2019 – July 2019) -** Conducted 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% **(git link)**
- **Inventory Optimization problem on Kaggle (January 2019 – February 2019) -** Forecasted the demand for LED televisions using Holt-Winter's Smoothing method with MAPE of 20.760 **(git link)**
- **Book Recommendations from Charles Darwin (July 2020 – August 2020) –** Designed a book recommendation system based on the content utilizing the open-source Charles Darwin's bibliography using the data mining techniques **(git link)**
- **ASL Recognition with Deep Learning (July 2020 – August 2020) –** Created a convolutional neural network to classify images of American Sign Language (ASL) letters **(git link)**