**Shouvik Sharma**

400 E 33rd Street, Apt 515, Chicago, IL-60616|Phone: 3124592008 | [shouvik19@gmail.com](mailto:shouvik19@gmail.com) | [Linkedin](https://www.linkedin.com/in/shouvik-sharma19/) | [Github](https://github.com/shouvik19) | [Medium](https://medium.com/@shouvik19)

**SUMMARY**

Over 3 years of comprehensive work experience in Data Science, Marketing Analytics and Business Intelligence in banking, retail, and e-commerce domains. Ability to solve complex business problems using ETL, Data Warehousing, Machine Learning and Exploratory Data Analysis by working independently, and designing analytical solutions.

**EDUCATION**

* MS in Data Science, Illinois Institute of Technology, **GPA: 3.8** **(Aug 2019 - May 2021)**

**Related Courses**: Machine Learning, Big Data Technologies, Applied Statistics, Statistical Learning, Database Management,

Data Preparation and Analysis, Introduction to Algorithm, Data Science Practicum.

* MS in Statistics, NMIMS University, **GPA: 3.35**   **(Jul 2016 - Apr 2018)**

**Related Courses:** Regression Analysis, Estimation, Testing of Hypothesis, Distribution Theory, Linear Algebra and Numerical Methods, Parametric Inference estimation, Probability Theory, Linear Models

* Certifications**:** [Snowflake Pro Certification](https://www.youracclaim.com/badges/f03d4251-13bd-4fd0-9f0d-45ff17bd718f), SAS Certified Base Programmer for SAS 9, SAS Certified Predictive Modeler

**SKILLS**

* **Programming*:*** SQL, Python, R, SAS, Pyspark, HTML, C#, Excel VBA (Macros), Regex, NLP, Adobe Analytics.
* **Big Data Ecosystem**: NoSQL databases, Spark, Hadoop, MapReduce, Hive, Pig, Kafka, Flume.
* **Cloud Technologies**: AWS (S3, EC2, Lambda, Athena, RDS, Redshift, EMR), MATLAB, GCP.
* **Tools*:*** Tableau, Power BI, Powerpoint, RStudio, Jupyter, SAS E-Miner, SPSS, SSIS, MS Office, JIRA, Spotfire, MS Access, Looker.
* **Libraries:** Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Keras, Nltk, Gensim, Scipy, Beautiful Soup, Tensorflow.

**WORK EXPERIENCE**

**Data Scientist at Daten Solutions Inc., Chicago: (May 2020 - Present)**

* Developed and automated **data migration pipeline** from SQL Server to Snowflake and performed **dimensional modeling** on the migrated data using version control in **GitHub** after breaking down strategic problems.
* Performed **customer segmentation** using **k-mean clustering** in **AWS Sagemaker**, further analyzed data to provide subject matter insights and recommended cluster-wise products using **apriori algorithm** which ultimately improved the top-line revenue by **4%.**
* Created **ad-hoc reports** and **tableau** dashboards to explain variation in success **Metrics** and **Time Series Analysis**.
* Developed statistical models like **ARIMA** using statsmodels package in **Jupyter Notebook**, the model achieved an overall accuracy of MAPE 5.96%**.**

**Data Scientist – Practicum Student at Labelmaster, Chicago: (May 2020 – Dec 2020)**

* **Predicted** department-wise sales based on seasonal and external factors, by working with business stakeholders.
* Implemented Statistical methods like **SARIMAX, VAR** along with some hypothesis testing as well as Machine Learning (Deep Learning) Time-Series techniques to large sales data.
* Achieved an accuracy of **MAPE 8%** approx. on price forecasting using Deep Learning algorithms like **LSTM** and **RNN**, to showcase results further created dashboards using Tableau.

**Data Scientist at Cartesian Consulting: (Apr 2018- Jul 2019)**

* Identified probable customer churn using Predictive Models in Python like **Logistic Regression, Decision Trees, Random Forest** and achieved a true positive rate (**recall**) of 84% for target customer retention and acquisition marketing campaigns.
* Predicted sales by **time series forecasting** using statistical conceptsin **Python** using **neural networks, ARIMAX** and **Prophet** for inventory management by eliminating understocking and reducing overstocking by 56%.
* 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 topline revenue by 13 % for a grocery client.
* Generated visualizations using **Tableau** toanalyze marketing **metrics** for making recommendations and supply chain analysis.

**Data Scientist Intern at Greeksoft Technologies Pvt. Ltd.: (Sept 2017 - Dec 2017)**

* Worked with the **Apache** **Spark** Framework for customer analytics using **Spark** **SQL** queries on large scale datasets for developing flawless **CRM** (customer relationship management) campaigns and deployed them through multiple channels.
* Built an **RNN Neural Network** model for Live positional trading using **Keras** package in python with an accuracy of 71 %.

**Data Scientist at Tata Capital Financial Services Ltd.:  (Jul 2015- Jul 2016)**

* Built **KPIs** and **Regression** models to predict **customer life-time value**, enhance propensity and scoring attributes.
* Accurately extracted insights and created dashboards using **Tableau,** **Excel VBA (Macros)**, **pivot** **tables** and **slicers**.
* Identified **fraudulent transactions** in the bank repayments using machine algorithms with an accuracy of 5%.

**PROJECTS**

**Stack Overflow Data Analysis Model (Language/Tools- Python, Jupyter Notebook, Spark, Hive, PySpark, Pig):**

* Analyzed insights about questions posted on stack overflow by extracting large data sets using GCP’s big query data warehouse ; discovered top spammers, expert users, and most valuable customers users by leveraging big data technologies such as Apache HiveQL, Apache Pig and Apache Sparks ([git link](https://github.com/rahulmnair1997/StackOverflow-Data-Analysis))

**Recommendation System using Yelp (Language/Tools- Python, Jupyter Notebook, NumPy, SciPy, pandas, scikit-learn):**

* Built a personalized restaurant recommender web app using the Yelp dataset of restaurants by testing models like Pure Collaborative, Approximate Nearest Neighbor, K-NN, Naive Bayes and Hybrid Matrix
* Factorization on different hyperparameters which were tuned using the python library scikit optimizer ([git link](https://github.com/shouvik19/Restaurant-Recommendation-System-using-Yelp-Dataset))

**Image Mating using CelebAMask-HQ (Language/Tools- Google Colab, regression):**

* Conducted Image Matting using the U-Net architecture of the Convoluted Neural Networks on the opensource Celeb-Mask dataset with an IOU Score of 92%

**Inventory Optimization problem on Kaggle (Language/Tools- Google Colab, Tableau, R studio, Adobe Analytics):**

* Forecasted the demand for LED televisions using different time-series forecasting methods with Holt-Winter’s Smoothing method as the best method with MAPE of 20.760.

**Electronic Vendor Database: (Language/Tools - MySQL, Java 8, HTML, CSS, Bootstrap):**

* Constructed the ER Model and translated into Relational Schema implemented as SQL script.