Shouvik Sharma

2829 S Wells Street, Chicago, IL-60616|Phone: 3124592008 | shouvik19@gmail.com | Linkedin | Github | Medium

SUMMARY

Over 5 years of comprehensive work experience in Data Engineering, Marketing Analytics and Business Intelligence in banking and retail domains. Ability to solve complex business problems using ETL, Data Mining, Machine Learning & Data Warehousing concepts.

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.

- **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, SAS Certified Base Programmer for SAS 9, SAS Certified Predictive Modeler

SKILLS

- **Programming:** SQL, Python, R, SAS, Pyspark, HTML, Excel VBA (Macros), Agile Methodology, PostgreSQL, MySQL.
- **Big Data Ecosystem**: Spark, Hadoop, MapReduce, Hive, Pig, Kafka, Flume, Hbase, Microsoft Azure.
- **Cloud Technologies**: AWS (S3, EC2, Lambda, Athena, RDS, Redshift, EMR), NoSQL, Cassandra, MongoDB, Kubernetes, Snowflake, CircleCI, Airflow, Prefect, Google Data Studio.
- **Tools:** Tableau, Power BI, Azure ML, RStudio, Jupyter Notebook, SAS E-Miner, SAS CI, IBM-Unica, SSIS, MS Office, JIRA, Looker.
- **Libraries:** Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Keras, Nltk, Gensim, Scipy, Beautiful Soup.
- **Datasets:** HTTP, HTML, XML, JSON

WORK EXPERIENCE

Data Management Associate at Avant LLC, Chicago*:*   (Aug 2021 –Present)

- Performed A/B tests for direct mail offers using customer attributes, this resulted in improved response rate of the direct mail campaigns by 13%.
- Extracted insights from the marketing data to create **source attribution funnel**, this helped business stakeholders to improve customer application experience, and increase application rate by 4%.
- Migrated marketing campaign pipelines from conventional marketing tools like Responsys to Segment using API supported python libraries.
- Build out refinance product data pipeline for existing loan customers based on their transunion credit scores and existing features, this increased the overall refinance book by 10%.
- Created explores, looks and dashboards using **LookML** in **Looker** to monitor expected fraud losses and emerging fraud trends for the credit card portfolio.
- Coordinated with the marketing analytics team to develop a category-based rewards product based on existing transaction data along with Argus reports and market research on competitor products to improve top of the wallet share.
- Supported debt collection practices of Avant LLC., by developing mission-critical pipelines that delivered hourly customer update information to external collection agencies like Livevox, by designing the data pipelines that implemented data cleaning, data processing, and data delivery using Pyspark and orchestration tools like Airflow.

Data Science Intern at CNH Industrial Inc., Racine*:*   (Mar 2021 – Aug 2021)

- Developed predictive analytics and statistical forecasting models such as LSTM, ARIMA, neural networks etc. to predict quarterly sales, achieved an accuracy of 73% for the back testing of the year 2020.
- Formalizing assumptions about how demand forecasts are expected to behave, creating definitions of outliers, developing methods to systematically identify these outliers, and explaining why they are reasonable or identifying fixes for them
- Development of tools to allow process automation, analysis & corrective action implementation by the business.

Data Engineer at Daten Solutions Inc., Chicago*:*   (May 2020 – Mar 2021)

- Developed and automated data migration pipeline from SQL Server to Snowflake using SnowSQL and SnowPipe, and further enhanced data quality by performing 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.
- Created Tableau dashboards to explain variation in success Metrics and Time Series Analysis to higher management.
- Automated reporting process using Excel VBA (Macros) and MySQL maintaining accuracy and saving ~ 75% of time, maintained version control Git, Mercurial, SVN.

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

- Involved in designing databases, data marts, E-R model for OLTP and multi-dimensional model for OLAP using SnowSQL.
- Optimized complex SQL scripts for quality checking of projects and populating output tables for deployment using Azure Pipelines.
- Automated hourly status report saving 10 man-hours/week, thus decreasing response time for fixes and campaign failures.
- Achieved an accuracy of MAPE 8% approx. on price forecasting using Deep Learning algorithms like LSTM and RNN, further created dashboards for presenting the forecasted values to the higher management.

Data Analyst at Cartesian Consulting, Mumbai*:*   (Apr 2018- Jul 2019)

- Predicted sales by time series forecasting in Python using neural networks, ARIMAX and Prophet for inventory management by eliminating understocking and reducing overstocking by 56%.
- Applied K-means clustering in Python for segmentation of customers, comparing it with RFM (Recency, Frequency and Monetary Value) analysis for improved campaign targeting.
- Developed dimensional data models and data warehouse adhering to integrity and normalization rules to support campaign
- data mart and customer one view for marketing campaigns. Wrote complex SQL queries (multiple joins, CTE's, subqueries).