

# Shivam Chauhan

[chauhan.shi@northeastern.edu](mailto:chauhan.shi@northeastern.edu) • (201) 282-1519 • Boston, MA • [GitHub](#) • [LinkedIn](#) • [Portfolio](#)

## SUMMARY

Data professional with 3 years of experience in problem solving with data. Adept at analytics and interpreting complex datasets. Research experience in Data Science, ML & NLP. Deep understanding of statistical modelling. Proficient in Python and SQL.

## EDUCATION

**Northeastern University, Boston, MA - Masters in Analytics, GPA: 3.8/4** Jan 2019 – Dec 2020

**Coursework:** Probability & Statistics, Predictive Analytics, Data Mining, Big Data, GIS, AWS, Artificial intelligence

**Symbiosis University, Pune, India - Bachelor of Business Administration in IT, GPA: 3.2/4** Apr 2013 – May 2016

## TECHNICAL SKILLS

**Languages:** SQL, Python (NumPy, Pandas, Scikit-learn, TensorFlow, Keras, PyTorch), PySpark, Git, R, Flask  
**Business Analysis:** Tableau, Power BI, Google Analytics, Google data studio, Microsoft Excel (Pivot Tables, V lookups)  
**Machine Learning:** Regression, Classification, Clustering, PCA, Time Series Forecasting, Natural Language Processing  
**Data Engineering:** Data extraction, manipulation, MySQL, PostgreSQL, AWS (S3, EC2, EMR, Sagemaker), KNIME, Alteryx

## EXPERIENCE

**Northeastern University, Boston, MA** Aug 2020 – Present

*Data Science Research Assistant (under Prof. Mikhail Oet)*

- Performed sentiment analysis on financial sector specific news articles leveraging NLP, ML and Deep Learning techniques
- Scraped 500,000 articles by writing Python scripts to extract URL from unstructured GDELT database based on financial sector related keywords; utilized web-scraping method to gather text from those URL and loaded to MySQL database
- Developed classification algorithms using Bag of Word, TF-IDF, and pre-trained BERT to study 5 years sentiment trend
- Designed ETL pipeline and optimized process & code for each financial sector, saves 3+ hours by auto saving results

**Point Focal, Boston, MA**

Sept 2020 – Dec 2020

*Data Analyst Capstone*

- Provided FinTech analytics to portfolio managers, investors and traders by analysing time-series data from Stock market
- Created interactive Tableau dashboard to monitor stock performance of portfolio, manage risk and improve returns
- Drafted project proposal to research on new data source and implemented overall plan for project completion in time
- Founded relation between momentum of prices of 3000+ stocks and stock-level sentiment scores from correlation stats
- Trained 3-layer LSTM Classification Model to forecast future momentum of prices; resulted in accuracy of 53.76%

**Amazon, Pune, India**

Apr 2017 – Dec 2018

*CS Associate*

- Managed responsibilities of troubleshooting amazon services and provided personalized resolutions to customers
- Analyzed customer purchasing behavior data to resolve 100+ tickets per month with customer satisfaction of 92%
- Advised fraud detection team with ad-hoc reports to minimize concession abuse and improve customer experience

## PROJECTS

**Viacom Customer Segmentation** (Skills: Python, KNIME, ML, Clustering, Tableau) Sept 2019

- Reported performance of social media ad campaigns to interpret and segment target audience for Viacom company
- Predicted increase in customer engagement by 15% by performing k-mean clustering to form marketing strategy
- Constructed a Tableau dashboard to visualize and identify each cluster demographic and behaviour information
- Applied Linear Regression on impression (CPM) price with accuracy 98% to estimate advertisement expenditure

**ETL Pipeline & Boston Air Quality Prediction** (Skills: Python, Keras, PostgreSQL, Airflow) Mar 2021

- Wrote Python script to extract daily weather data & Air Quality Index (AQI) from API and performed data transformation
- Created PostgreSQL Schema to store prepared data. Utilized Apache Airflow to design Dag to run pipeline daily
- Implemented Random Forest, XgBoost, KNN, and ANN regressor model. Achieved lowest MSE from Random Forest

**Movie Recommendation Engine** (Skills: Python, SVD) Jan 2021

- Developed recommendation system in python from user-based collaborative and content-based filtering techniques
- Evaluated results of matrix factorization to better understand latent features of customer to suggest movies to user

**Chatbot and Google Analytics on Personal Website** (Skills: Google Analytics, DialogFlow, NLP) Oct 2020

- Built personal website using HTML, CSS, Docker and linked with Google Analytics to monitor website traffic & behaviour
- Tacked user activity to better comprehend users, site performance; Visualised one month report in Google Data Studio
- Designed conversational chatbot on Google Dialogflow with intents & entities, and integrated it to website through API