

EDUCATION

Boston University

Master of Science in Computer Information Systems (Data Analytics concentration)

*Boston, MA
Expected Jan 2023*

Relevant Coursework: Advanced Database Management, Data Science with Python, Web Analytics and Data Mining, Foundation of Analytics with R

Guru Gobind Singh Indraprastha University

Bachelor of Technology (Computer Science)

*New Delhi, India
Aug 2013 – Jun 2017*

SKILLS

- **Programming Languages:** Python (NumPy, Pandas, NLTK, Scikit Learn, BeautifulSoup, Selenium), Java, C++, R
- **Data Product:** MySQL, AWS Redshift, PostgreSQL, SQL Server, ETL, MS Excel
- **Business Intelligence Tools:** Qlik Sense, QlikView, Tableau (Beginner)
- **Additional Skills:** Product Management, Data Analytics, Leadership Skills, Team Management, Analytical Decision Making
- **ML Algorithms:** KNN, Regression, Decision Tree, Random Forest, Naïve Bayes, K-means, Ensemble

EXPERIENCE

Boston University

Research Assistant

*Boston, MA
Oct 2021 – Dec 2021*

- Worked under guidance of **Dr. Eugene Pinsky** and **Dr. Yesim** on research project to develop platform for readily querying and obtaining information from more than 50 towns about issues such as planned housing across all Zoning regulations.
- Used stemming, lemmatization, and various other Natural Language Processing methods to identify key values to quote, even if values are mentioned in a different format.
- Developed process for standardizing housing data, establishing data cleaning methods for improving data quality by 68%.

Soroco

Associate Business Analyst

*Bengaluru, India
July 2019 – Aug 2021*

- Conceptualized and constructed **Workforce Intelligence** using Qlik Sense to drive actionable changes by **providing hard-to-track insights**, improving team productivity by 65% and quality by 45% across 10 F500 clients.
- Concluded **Process Cleanup** using Qlik Sense to identify and prioritize processes for consolidation, standardization, or automation and create an organizational-level digital transformation roadmap, increasing team revenue by multiple folds.
- Led an engagement with F500 company and helped **save 40% of the time** by identifying most optimal variation in a particular process, ultimately saving **\$100K per year**.

Shore Infotech India Pvt Ltd

Data Automation Engineer

*Hyderabad, India
Aug 2018 – May 2019*

- Drove solution designing and implementation for analytical projects and Spearheaded automation and analytics operations for data acquisition processes and end-to-end infrastructure automation, reducing data extraction costs by 35%.
- Managed and constructed high-end generic robots and scripts for **identifying and filtering financial data** from semi structured and unstructured data sources (Schedule of Investments), **saving manual effort, worth 35K hours per year**.

Energy Analytics Lab, IIT Kanpur

S. Research Associate – Project Associate

*Kanpur, India
June 2017 – Aug 2018*

- Interpreted statistical data and research conclusions; presented market findings and recommendations to 4 new clients.
- Scripted Python scripts using Selenium and BeautifulSoup to derive data from various static and dynamic sources and implemented Exploratory Data Analysis to identify anomalies in gathered data.
- Leveraged automated data collection techniques to collect historical power market data to be fed into ARIMA model to predict power consumption; achieved an overall accuracy of 97%.

Tatras Data

Data Science Intern

*New Delhi, India
June 2016 – July 2016*

- Created a framework to download Punjabi News articles from more than 3 websites by leveraging Selenium, and BeautifulSoup; cleaned data for analysis leveraging stemming and lemmatization.
- Developed recommendation system, content-based and collaborative filtering by using Natural Language Processing and Topic Modelling techniques.

PROJECTS

Loan Default Prediction

- Performed data mining and validation, optimizing detection algorithms, and separating genuine phenomena from spurious anomalies; undertook data treatment on missing values.
- Built financial decision model using customer's data to conclude if consumer would default on loan or not; achieved an accuracy of 86%.
- Used logistic regression to anticipate if customer would repay loan or not, increasing overall accuracy to 96%.

Analysis for Deaths due to Covid-19 and Pneumonia

- Utilized Pandas and Selenium library in Python to download and clean data; investigated data anomalies; undertook data treatment on missing values.
- Evaluated Central Limit Theorem by taking samples for size 10, 20, 30, and 40 to analyze how mean and standard deviation varied across samples.
- Presented project report on 3 statistical sampling techniques used for periodic research; evaluated benefits and limitations of probability and non-probability sampling methods including stratified and systematic sampling.

Information Retrieval on Tweets collected during Disaster Situations

- Deployed techniques including data deduplication, tokenization, and stemming to clean data and then used Term Frequency- Inverse Document Frequency and Cosine Similarity to calculate similarity between query and tweets.
- Assessed accuracy of retrieval and relevance of tweets retrieved using Mean Average Precision and published results in a conference paper; achieved MAP of 0.68 over 7 queries.

Sentimental Analyzer

- Designed and implemented an information retrieval and classification system for tweets by crawling nearly 3600 tweets per hour using API; Utilized Requests library in Python to extract JSON responses.
- Cleaned, parsed, and segmented tweets content; with purpose of creating a database to store most frequent words and hashtags associated with each emotion.
- Leveraged various lexicon based (dictionary based and corpus based) methods to identify sentiments.

CERTIFICATIONS

- Completed **Python 3 Programming**, a 5-course specialization by **University of Michigan** on Coursera.
- Completed **Executive Data Science**, a 5-course specialization by **John Hopkins University** on Coursera.
- Completed **Software Product Management**, a 5-course specialization by **University of Alberta** on Coursera.
- Completed **Data Wrangling, Analysis and AB Testing** with SQL by **UC Davis** on Coursera.
- Completed **Data Science Math Skills** by **Duke University** on Coursera.
- Completed **SQL for Data Science** by **UC Davis** on Coursera.
- Graduated with **Data Analyst Nanodegree** from Udacity.

VOLUNTEER EXPERIENCE

- Volunteered at Shades of Happiness Foundation, a non-government registered organization, to help underprivileged children cope up with education as well as develop communication skills.