

## EDUCATION

### Master of Science in Information Systems | Northeastern University, Boston, MA

Aug 2020

Coursework: Data Warehousing, Business Intelligence, Data Science Engineering Methods and Tools, Database Management and Design, Cybersecurity, Algorithmic Digital Marketing

### Bachelor of Electronics and Telecommunication | KJ Somaiya College of Engineering, India

May 2016

## SKILLS and CERTIFICATIONS

Programming Languages	Python, SQL, HQL
Databases	Hive, MySQL, SQL Server, Oracle, PostgreSQL, Redshift, MongoDB
Data Skills	Data Visualization, Data Warehousing, Data Pipelining, Data Governance
Data Integration/ BI Tools	SSIS, Alteryx, Talend, Tableau, Power BI, Spotfire, Google Analytics
Machine Learning Algorithms	Regression models, Classifiers, CNN, Recommender Systems
Python Packages	NLTK, NumPy, Spacy, Plotly, Pandas, sci-kit learn, Matplotlib
Big Data Technologies	Hadoop, Hive, Pyspark, Apache SQL
Tools	E/R studio, MS Office Suite, Git
Certification	Tableau Data Scientist, Tableau Data Analyst, ITILv3 foundation

## PROFESSIONAL EXPERIENCE

### Business Intelligence Co-op –Bristol Myers Squibb, Devens, Massachusetts

Jul 2019 – Dec 2019

- Designed visual dashboards for Managers to identify key metrics of work stoppages and troubleshoot process issues
- Built a data pipeline to extract timestamp using Regex in Python and visualized it in Spotfire to eliminate ~26 hours of weekly manual work
- Designed dashboards for senior leader to produce recommendation and insights to shape business decisions
- Classified process review PDF with NLP techniques like tokenizing, lemmatization and word cloud with 73% accuracy

### Data Analyst – Netmagic Solutions, Mumbai, India

Sept 2016 – July 2018

- Collaborated with clients to design dashboards for their website traffic, traffic source and optimize uptime of website
- Created ad-hoc analyses for stakeholders to get specific answers from the web traffic like number of active customers in given time frame
- Worked closely with Sales, PMs, Internal Operations team to identify the root cause of a problem and find best possible solution
- Integrated and cleaned data from multiple locations in SSIS for clients so they can use the data for further analysis

## ACADEMIC PROJECTS

### Big Data Analytics [AWS Sagemaker, Kinesis, Putty, NLP]

Summer 2020

- Built real time analytical dashboard by scrapping data from twitter using Tweepy API and storing results in MySQL DB
- Designed a data pipeline using AWS sagemaker and Kinesis for parallel streaming of data

### Zlymo Market Analytics [AWS Sagemaker, S3, AWS Glue, Redshift]

Spring 2020

- Created a click through rate web analytic dashboard in Tableau to identify impact of different channels, campaign and promotion on products.
- Built data pipeline to generate sales insights by collecting metadata in AWS glue and transferring to redshift

### Business Intelligence and Data Warehousing [SSIS, Talend, Tableau, Power BI]

Spring 2019

- Designed a Data warehouse by collecting data from 4 different sources and Database and loading it in SQL server
- Developed dashboards in Power BI for managing inventory of products, campaign analysis and reject analysis
- Built a performance tuned Data pipeline by allocating appropriate ram, parallelizing jobs in Talend and SSIS

### Hospital Management Database [MySQL, ER/Studio]

Fall 2018

- Normalized database till 3NF to eliminate update, delete and insert anomalies
- Built complex SQL queries - trigger, stored procedure and views to retrieve data efficiently
- Empowered ease of architecture readability by building the EER model using the reverse engineering in MySQL

### AirBnb User Booking[Python, KNN, Pandas]

Fall 2018

- Performed A/B testing on the raw dataset to understand the feature importance under various circumstances
- Implemented feature importance to increase the accuracy of the prediction
- Plotted and visualized the results of different algorithms used such as KNN, Random forest and XG Boost using matplotlib library