

## TECHNICAL SKILLS

---

- **Technologies:** AWS, Linux server, Git, Rest API, Flask, ETL, Agile, Docker, MongoDB, Elasticsearch, Sql, Hadoop
- **Programming Languages and Tools:** Python, R, C, Go, Tensorflow, Scipy, Keras, Qgis, NLTK, scikit-learn, Stanford-NLP, Tableau, SAS Visual Analytics, Power BI, AWS Data pipeline
- **Machine/Deep Learning:** Regression Models, Tree Models, Clustering Models, SVM, CNN, RNN, Gradient models, LSTM

## EXPERIENCE

---

- **Python Developer** Montreal, QC  
*SITA* *September 2019 – Current*
  - Build out the source control query service that queries information in a scalable and efficient manner.
  - Worked closely with different stakeholders of the system to understand their needs for this system integration.
  - Designing API for large scale system capable of transferring thousand data points per second from multiple sources using multithreading.
  - Rebuild driver for Lidar and Sensors from C to Python 3.
  - Upgrade the current system from python 2 to Python 3.
  - Created Technical specifications documents based on the functional design document for the ETL coding to build the data mart.
  - Automated MySQL container deployment in Docker using Python and monitoring of these containers.
  - Develop and automated testing of Back-end system which lead to decrease of testing time and bugs.
- **Data Analysis Research Assistant** Thunder bay, ON  
*Lakehead University* *November 2017 – May 2019*
  - Documented the complete process flow to describe program development, logic, testing, and implementation, application integration, coding.
  - Worked with project team representatives to ensure that logical and physical data models were developed in line with corporate standards and guidelines.
  - Involved in the project analysis, design, and implementation & debugging.
  - Worked closely with different stakeholders of the system to understand their needs for this system integration.
  - Performed troubleshooting, fixed and deployed many Python bug fixes of the two main applications that were a main source of data for both Canada health department and internal department service team.
  - Experienced in Agile Methodologies and SCRUM Process.
  - Designed and developed “Elasticsearch” on distributed server with 20 nodes which allow searching from 330M tweets/second.
  - Performed quantitative analysis and visualization of social media data for health analyzation.
  - Deployed and tested different modules in Docker containers and GIT.
  - Usage of advance features like pickle/unpickle in python for sharing the information across the applications.
- **Data Analyst** San Diego, CA  
*Datalog.ai* *January 2017 – August 2017*
  - Involved in various phases of Software Development Life Cycle (SDLC) such as requirements gathering, modeling, analysis, design and development.
  - Involved in the process of creating Use-case diagrams, Activity flow diagrams, Class diagrams and Object diagrams in the design phase and developed the Coding module.
  - Created Python and Bash tools to increase efficiency of call center application system and operations, data conversion scripts, REST, JSON, and CRUD scripts for API Integration.
  - Automation of unit testing using Python.
  - Used Unix/Linux OS based cloud for the deployment. Wrote scripts in Python for extracting data from HTML file using Python library BeautifulSoup.
  - Design and developed clustering algorithm for decision making based on chat.

- Successfully implemented 5+ visualization dashboard for backend analysis of chatbot data: semantic analysis, topic extractor & other.
- Implement asynchronous API for chatbot analysis system, which is capable of handling thousands request per second.
- Extensively involved in Data Extraction, Transformation and Loading (ETL process) from Source to target systems using MongoDB and Oracle database.
- Worked with Technical Experts to Improve Report Performance and performed System testing.
- Providing End-user Training and documentation for customer reporting services.
- Excellent experience with Python development under Linux OS (Debian, Ubuntu, CentOS SUSE Linux, Red Hat Linux, Fedora)

## • Data Analyst

Godhra, GJ

July 2014 – December 2016

*Panchamrut Dairy*

- Working real time data analysis system for raw product cost and product transportation using SAP.
- Detail report generation of different products sales and cost.
- Perform data analytic on sales data for forecasting future growth of products.
- Owned the assigned reports, worked on them and updated the Report Development Scheduler for status on each report.
- Worked on daily basis with lead Data Warehouse developers to evaluate impact on current implementation, re-design of all ETL logic.
- Worked on exporting reports in multiple formats including MS Word, Excel, CSV and PDF.
- Responsible for making the report available for scheduling or viewing on demand.
- Forecasted sales pattern of ice-cream using time-series with 71% efficiency.

## EDUCATION

---

### • Master of Science in Computer Science

Thunder Bay, ON

*Lakehead University; GPA: 3.9/4.00*

September 2017 – May 2019

### • Bachelor of Engineering in Computer Science

Vadodara, India

*Gujarat Technology University; CGPA: 7.31/10*

August 2010 – July 2014

## OPEN SOURCE LIBRARIES

---

- **Emot:** Developed and Maintainer of widely use open source "emot" python library(11k+ Users) to extract the emojis and emoticons from a text with their respective meaning. Lead library maintainer team for bug fixing and new features.
- **Sentencelab:** Developed and Maintainer open source python library for text data cleaning, sentiment analysis and topic extraction.
- **Kaggle:** Open source two database "Arxiv analysis data" and "Indian's AI growth" on kaggle with 25k+ Downloads.
- **Bug solving and test cases:** Solved more than 13 bugs and 10 test model for different open source libraries.

## FUNDING

---

- **Research grant:** 7000\$/Yearly NSERC Discovery Grant.

## PUBLICATION

---

- Shah, Neel, Darryl L. Willick and Vijay Mago. "A framework for social media data analytics using Elasticsearch and Kibana." *Wireless Networks* (2018): 1-9.
- Neel Shah and Vijay Mago. Canadian's health analysis through social media. *IEEE Machine Learning* 2019: (submitted)