

Neel Shah

<https://neelshah18.github.io>

Email : neelshah181993@gmail.com

<https://www.linkedin.com/in/neelshah18>

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

• Data Analysis Research Assistant

Lakehead University

Thunder bay, ON

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.
- Spatial analysis of Social media data to predict health problems in Canada with 94% accuracy using scikit-learn.
- Successfully implemented real-time health analysis system using social media data with 93% accuracy of multiclass classification using Random forest in tensorflow and python.
- Design, develop and test Machine learning model for analysing and predicting diabetic trend in Canada through social media data on distributed NVIDIA Titan-X CUDA server using tensorflow and keras.
- Created largest open source food database “Food in one” with 333K data points for statistical analysis of country level food buying and consuming pattern.
- 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.
- Created Technical specifications documents based on the functional design document for the ETL coding to build the data mart.

• Data Analyst

Datalog.ai

San Diego, CA

January 2017 – August 2017

- Developed backend NER functionality for NLP algorithm used in chatbots.
- 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.
- Gathered requirements from remotely based business users and defined and elaborated the requirements by holding meetings with the users.
- Providing End-user Training and documentation for customer reporting services.

• Data Analyst

Panchamrut Dairy

Godhra, GJ

July 2014 – December 2016

- 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.
- Developed Tableau visualizations and dashboards using Tableau Desktop.
- Forecasted sales pattern of ice-cream using time-series with 71% efficiency.

EDUCATION

- **Master of Science in Computer Science**
Lakehead University; GPA: 3.9/4.00
Thunder Bay, ON
September 2017 – May 2019
- **Master of Engineering in Information Technology**
Gujarat Technology University; CGPA: 8.78/10
Vadodara, India
August 2012 – June 2014
- **Bachelor of Engineering in Computer Science**
Gujarat Technology University; CGPA: 7.31/10
Vadodara, India
August 2008 – July 2012

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)