

# Siddhant Shah

shah.sid@northeastern.edu | linkedin.com/in/thesidshah | github.com/thesidshah

Experienced Software Developer using technology and data-driven decisions to design systems that enhance lives

## Relevant Experience

### Software Engineer II

TERVA Corporation

8/2023 - Present

Worcester, MA

- Spearheaded software development efforts as the first developer at TERVA Corporation, delivering a comprehensive daily report system within a strict 6-month timeline.
- Implemented Microsoft authentication, Next.js 13 frontend, and FastAPI backend, optimizing user engagement by 30% and reducing page load times by 20%.
- Automated PDF generation and email distribution, reducing manual workload by 70% and improving report dissemination efficiency by 60%.
- Explored Django integration and utilized Terraform for infrastructure management, achieving a 15% improvement in backend performance and reducing deployment time by 25%.

### Software Engineer Co-op

SS&C Intralinks

10/2022 - 12/2022

Waltham, MA

- Improved website loading time by utilizing **code splitting techniques**, resulting in a significant 30% reduction in **JavaScript** bundle size and enhanced front-end performance
- Performed **unit testing** on 30+ components, uncovered and fixed 7 critical bugs on StoryBook
- Implemented industry-standard practices like Component Story Formats (CSF) and achieved **90% test pass rate** through CI/CD practices

### Graduate Research Assistant

Center for Translational Neuroimaging Lab CTNI, Northeastern University

05/2022 - 07/2022

Boston, MA

- Designed **data warehouse** on **AWS S3** & business intelligence platform using **Flask** that streamlined data collection, integration & reporting from multiple sources resulting in 25% reduction in **data processing** time
- Conducted **data preprocessing, cleaning**, and quality control of **neuroimaging data** from 20+ experiments using MRIQC and custom Python scripts, achieving data quality score of 90% or higher using AWS

### Associate Software Engineer

Globant

12/2020 - 08/2021

Pune, IN

- Evaluated leadership requirements of Fortune 100 company across 5 regions and developed unified **ServiceNow** platform, reducing redundant reports by approximately 73%
- Improved month-over-month report viewership by 20% through standardizing local instances to global standard and resolving discrepancies
- Leveraged ServiceNow API to transfer data to **Microsoft SQL Servers** and developed interactive reports with **Power BI**, saving over 500 man-hours on report creation and circulation

### Software Engineer

Freelance

07/2020 - 09/2021

Pune, IN

- Created an **interactive adaptive** website for Suyog Life Care, resulting in 17% increase in viewership across all devices
- Implemented SEO best practices, including keyword research and optimization in **collaboration** with the marketing team, leading to a 20% increase in organic traffic
- Executed promotional campaigns for social media profiles and email marketing with sendgrid, boosting **customer acquisition** and retention by 30%

### Software Engineer - Machine Learning

Visava Labs

06/2018 - 06/2020

Pune, IN

- Secured 80% of first-round **funding** by pitching a cutting-edge product at the intersection of Technology & Architecture to address accessibility issues at scale in R&D department
- Integrated expert-based system with **Android application** reducing delivery time to 20 seconds; completing the first prototype and creating new user base
- Leveraged Generative Adversarial Networks (Big GAN) to create over 1000 possible floor plans for each input constraint through the **ML Ops** pipeline on **Google Cloud**

### Machine Learning Engineer Intern

IoTIoT.in

02/2018 - 06/2018

Pune, IN

- Used in-house arm-based Shunya OS with **custom kernels** and saving 3% power
- Analyzed different **CNN models** to minimize number of facial landmarks to 18 data points, saving 10000 faces in 15kbs, saving \$300+ in storage costs

## Education

**Master in Computer Science**, Specializing in AI, Northeastern University, MA

May 2023

**Bachelor of Engineering in Computer Engineering**, Pune University, IN

May 2020

## Technical Skills

---

### Programming Languages:

JavaScript Python Java C++ R SQL

### Databases:

Hive MySQL MongoDB Cassandra Neo4j Graph Snowflake

### Cloud services:

GCP Cloud Filestore, OAuth 2.0, Vertex AWS S3, EC2, RDS, Lambda

### AI and ML Frameworks:

PyTorch SciPy StatsModels OpenCV NLTK SpaCy

### Other Tools and Technologies:

Docker Kubernetes ServiceNow Spark Hadoop Jenkins

### Recent Projects

---

## Related to Software Development

### Personal Portfolio Website | JS, React, Tailwind, Node

[www.siddhantshah.me](http://www.siddhantshah.me)

- Creating a library of 20+ React components that are screen-adaptive using TailWind and a node.js based server to make an easy-to-duplicate template for fellow Northeastern students.
- Using 3+ JSON files as data sources in the generation pipeline for personal details to make it extensible.
- Working on an in-browser NLP model that will read a Husky resume & generate a portfolio in 60 sec.

### Author Details Database mining | R, XML, MySQL, SQLite

- Sped up read time and transformed data from a file from about 3 hours to 15 minutes.
- Created and populated a normalized relational schema with entities/tables: Articles, Journals, and Authors, using XML data and implemented in SQLite, using R to make an efficient database.
- Created and populated a star schema for author facts in MySQL by loading data from SQLite database and optimizing for interactive analytical queries with pre-computation and storage.

### Dungeons and Demons: A game of Design patterns | Java, Java Swing, Junit

- Developed a game that creates 1000+ randomized variants of maps of the dungeon.
- Applied Model View Controller and Factory design pattern to make the code adaptive and extensible.
- Followed a strong development cycle with 6 iterations of feedback to efficiently create the game.

## Related to Data Science

### Estimating student population in Marino Center | ARIMA, SARIMAX, LSTM

- Predicted floor wise distribution with ARIMA and achieved an accuracy of 55%.
- Analyzed peaks and worked with limited data for LSTMs to achieve a mean absolute error of 20%.
- Created an interactive dashboard for the rec center and increased engagement by 10%.

### Estimating Muscle Force for amputees | Spark, Horovod, PyTorch, Mllib

[\[Report\]](#)

- Used PCA over a distributed system to conclude no significant time saving above 8 partitions.
- The lowest RMSE using Regularized Linear Estimation with parallelization achieved was 11.4%.
- A 4x speed up was achieved for 2D CNN with 32 filters of kernel size 2, the RMSE achieved was 3.7%.

### Topic Modeling and Extractive Summarization | NLTK, Scikit-learn, Numpy, Gensim

- Used Non-negative Matrix Factorization, Truncated Singular Value Decomposition, Latent Dirichlet Allocation for generating top k relevant words and text summaries with rouge scores  $\sim 0.3$ .
- Normalized 1000+ text documents using Regex (re), Count Vectorizer, Term Frequency - Inverse Document Frequency (TFIDF) from sklearn library to get high quality word embeddings.

### Image convolutions | Python, Numpy, OpenCV, Matplotlib

- Developed a framework to apply image operations such as filtering from scratch using NumPy.
- Dynamically detect and remove borders and outliers of the Harris map usually at the edges of images.
- Links to all code snippets can be found on [gists @ github](#). Part of a blog available on [Medium.com](#).

## Publications

---

Shah, et. al, **An Intuitive Study: Intrusion Detection Systems and Anomalies, How AI can be used as a tool to enable the majority, in 5G era**, [IEEE](#)