

# Saumil Shah

[linkedin.com/in/saumil-shah-0257/](https://www.linkedin.com/in/saumil-shah-0257/) | [github.com/saumilsjsu](https://github.com/saumilsjsu) | [saumilsjsu.github.io/](https://saumilsjsu.github.io/)  
[saumil0257@gmail.com](mailto:saumil0257@gmail.com) | (510) 402-9585 | New York, NY

---

## Skills and Tools

**Programming Languages:** Python, SQL  
**Tools & Technologies:** Tableau, Looker, Spark, Hadoop, Kafka, Airflow, Github, Gitab, Ansible  
**Databases:** MySQL, PostgreSQL, Google Bigquery, SQL Server, Snowflake, MongoDB  
**Cloud Technologies:** GCP, AWS, Docker, Kubernetes

---

## Work Experience

**Saucelabs**, San Francisco, CA

August 2022 – May 2022

### Data Engineer

- Built and maintained multiple fault-tolerant scalable data pipelines from internal databases, SaaS applications and event streams that reduced ETL process time by 40% and improved data accuracy by implementing data quality checks.
- Implemented process improvements to increase speed, scalability, reliability and quality of existing and new ETL pipelines.
- Supported development of performance dashboards that encompass KPIs for multiple stakeholders assisting in making optimal business decisions.

**Foresight Mental Health**, San Francisco, CA

March 2021 – August 2022

### Data Engineer

- Developed cloud-first data ingestion from multiple disparate sources to Google Bigquery by identifying structural necessities and designing optimal data warehouse architecture that improved processing speed by 70%.
- Automated optimal ETL processes across millions of rows of data, which reduced manual workload by 85% monthly.
- Built multiple scalable data collection pipelines and supported the development of performance dashboards that encompass KPIs for multiple stakeholders and C-suite assisting in making better decisions.

**DeveloperDB**, Sunnyvale, CA

July 2020 – Feb 2021

### Data Engineer Intern

- Designed, developed and launched scalable ETL data pipelines to curate databases streaming 30M rows of data on a daily basis.
  - Optimized and reduced runtime of the email prediction feature, which improved response time by 92% and accuracy by 30% resulting in increased user base by 47%.
- 

## Projects

**LinkedIn Automation Bot** (*Python, Selenium, MongoDB, Tableau*)

Oct 2020 – Nov 2020

- Developed a selenium bot that can mimic human behavior and automate multiple user actions on linkedin.
- Built an ETL pipeline to extract profiles and user data and created a database design to ingest data streaming in every day.
- Performed data wrangling and exploratory analysis on user data to depict a demographic visualization of popularity of job roles in different states of USA.
- Automated processes like sending a personalized connection request, withdrawing a connection request and endorsing someone's top 3 skills resulting in improved user engagement by over 35%.

**Identifying grape leaf diseases** (*Python, Selenium, LabelMe, Tableau*)

Feb 2020 – Apr 2020

- Constructed a web crawler that collected over 10000 images of grape leaves from the web.
  - Labeled and Classified images into 4 different diseases using LabelMe and converted the annotated images to Coco dataset using LabelMe2Coco library.
  - Enforced different data augmentations like flip, shift, rotate and zoom resulting in an increase in dataset to over 50000 images and resized all the images to achieve a uniform size.
  - Implemented CNN algorithm with 3 convolutional and pooling layers followed by a fully connected layer achieving an accuracy of 95.83%.
- 

## Education

**Master of Science in Software Engineering**

Jan 2019 – Dec 2020

San Jose State University, San Jose, CA, USA.

Relevant Coursework: Software Systems Engineering, Enterprise Software Platforms, Cloud Technologies, Data Mining, Large Scale Analytics, Machine Learning, Big Data Analytics, Deep Learning.

**Bachelor of Technology in Information Technology**

Aug 2014 – May 2018

NMIMS University, Mumbai, MH, India.

Relevant Coursework: Data Structures, Designing Algorithms, Object-Oriented Programming, Operating Systems, Data warehousing, Data mining, SAS, Database management systems, Cloud Computing.