

Shreya Gokhe

📞 1-551-260-0302 | 📍 Jersey City, NJ | ✉ sgokhe@stevens.edu | 🌐 shreya-gokhe | 📺 shreyagokhe | 📧 @shreyagokhe

EDUCATION

Stevens Institute of Technology

Master of Science in Data Science, GPA 3.6/4.0

Hoboken, NJ

Expected May 2021

Shreemati Nathibai Damodar Thackersey Women's University

Bachelor of Technology in Computer Science and Technology, GPA 5.47/6.0

Mumbai, India

Jun. 2017

SKILLS

Languages & Databases Python | R | SQL | MySQL | MongoDB | AWS

Big Data & Visualization Apache NiFi | Apache Hive | Apache Spark | Hadoop | Zoomdata | Tableau

Project Management IBM Rational | Microsoft Excel | Atlassian Confluence | TIBCO JasperSoft | GitHub

PROJECTS

EDA on COVID-19 data using Spark (PySpark, AWS)

Nov. 2020

- Executed inner join on tables having 22500 records to make data efficient without cleaning
- Filtered data by geographical region, grouped relevant columns and aggregated to visualize it as bar graphs for 4 use cases

Web scraping Indeed for predicting job titles (Selenium, Python)

Nov. 2020

- Scraped 5000 job descriptions each having titles data engineer, data scientist and software engineer from Indeed applying Selenium and python by saving HTML pages and parsing the content in .csv file
- Analyzed 5000 records of job description data and transformed text to vector of token counts using count vectorization
- Predicted job titles with an accuracy of 92% applying XGBClassifier algorithm and using pickle model to serialize the algorithm

Data Pipeline for a personalized newsletter (Python, MongoDB, OpenCV, NLTK libraries)

Jul. 2020

Collaborative pet project with an industry domain expert for building a customized newsletter using Deep Learning libraries

- Extracted 20 minute audio from tech conferences on YouTube using youtube_dl library in python and converted it into text by using speech recognition library effortlessly and pushed data into multiple MongoDB collections
- Generated an insightful summary of text and converted it to speech, reducing occurrences of repeated words by 99%
- Processed YouTube videos on various tech conferences and extracted content by 110% efficiency from presentations in the form of frames from video utilizing OpenCV

Sentiment Analysis of Yelp Restaurant reviews (Python, NLTK libraries)

Jun. 2020

- Accomplished exploratory data analysis to find top 85 cities and 12 states showing states that have better restaurants
- Applied NLTK to remove stopwords that helped with a 100% clean rating of sentiment counts

WORK EXPERIENCE

Jio | Data Engineer | Navi Mumbai, India

Jul. 2018 – Jun. 2019

- Designed data pipelines using Apache NiFi processors and RegEx and ingested data in Jio's Big data lake with 40% efficiency
- Synthesized Zoomdata instructional material into a cohesive guidebook benefitting Big Data Analytics Team resulting in a 100% streamlined onboarding process
- Refined and optimized existing HiveQL queries by 40% responsible for generating tables in Apache Hive
- Assembled and documented 8 page business requirements in collaboration with business analysts for a project which dealt with ingestion of Indian Social Security data
- Collaborated with upper management as team's single point of contact for assisting in the deployment of 122 productive projects

Jio | Engineer Trainee | Navi Mumbai, India

Jul. 2017 – Jul. 2018

- Performed User Acceptance Test on the Jio's messaging and digital streaming applications against 1200 test cases on IBM Rational Tool, detected and logged 30% of bugs
- Acted as liaison between UAT and development teams to efficiently communicate around 10 types of issues

VOLUNTEER WORK

- Functioned as a Cabinet Member for a team of 20 SNTD ACM Students' Chapter representatives *Sep. 2015 – Sep. 2016*
- Headed 'Pen IT' – a Technical Paper Presentation event during undergraduate studies *Feb. 2016*
- Managed Digital India Campaign - an Indian PMO initiative, with 50 participants as a student volunteer *Jul. 2015 – Oct. 2015*

PUBLICATIONS

- Published an article on Medium about creating and visualizing Choropleth maps using Folium – a Python library *Jun. 2020*