Shreya Gokhe

1-551-260-0302 | Jersey City, NJ | sgokhe@stevens.edu | shreya-gokhe | shreyagokhe | @shreyagokhe

EDUCATION

Stevens Institute of Technology *Hoboken, NJ*
Master of Science in Data Science, GPA 3.4/4.0 *Expected May 2021*

Shreemati Nathibai Damodar Thackersey Women's University *Mumbai, India*
Bachelor of Technology in Computer Science and Technology, GPA 5.47/6.0 *Jun. 2017*

SKILLS

Languages & Databases Python | R | SQL | MySQL | MongoDB
Big Data & Visualization Apache NiFi | Apache Hive | Apache Spark | Hadoop | Zoomdata | Tableau
Project Management IBM Rational | Microsoft Excel | Atlassian Confluence | TIBCO Jaspersoft | GitHub

PROJECTS

Data Pipeline for a personalized newsletter ***(Python, MongoDb, OpenCV, NLTK libraries)*** *Jul. 2020 – Present*

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

- Extracted audio from tech conferences on YouTube using the youtube_dl library in python and converted it into text by using speech recognition library effortlessly and pushed data into multiple MongoDB collections
- Applied cosine similarity method of NLTK to generate an insightful summary of text and gtts to convert the summary to speech which reduced the use of repeated and common words by 99%
- Processed YouTube videos on various tech conferences and extracted content from presentations in the form of frames from the video using OpenCV

ASHRAE- Great Energy Predictor III ***(R, ggplot2, corrr, corrplot, imputeTS, lubridate libraries)*** *Apr. 2020*

- Utilized time series dataset having 3 years of hourly meter readings from 1000+ buildings to develop accurate models of metered building energy usage with R
- Implemented linear regression to obtain residuals, F-statistics, and standard error for each type of buildings that helped to determine whether investments made for improvements are worthwhile

MovieLens- A Movie Recommendation System ***(Python, scikit-learn)*** *Apr. 2020*

- Executed Item-based Collaborative Filtering Algorithm on 20M dataset to recommend movies based on watched movies

Sentiment Analysis of Yelp Restaurant reviews ***(Python, NLTK libraries)*** *Jun. 2020*

- Performed exploratory data analysis to find top 85 cities and 12 states that showed which states 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 used by Big Data Analytics Team resulting in a streamlined onboarding process
- Refined and optimized existing HiveQL queries responsible for generating tables in Apache Hive
- Assembled and documented 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 successful 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 types of issues raised

VOLUNTEER WORK

- Functioned as a Cabinet Member of SNDT ACM Students' Chapter *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, 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*