JIMIT DHOLAKIA

jdholakia@cs.stonybrook.edu | (631) 710-9259
www.linkedin.com/in/jimit105 | github.com/jimit105 | jimit105.medium.com

EDUCATION

Stony Brook University — **Stony Brook, NY**
Master of Science in Computer Science, GPA: 3.64/4.0 — *Expected Dec 2022*
Relevant Coursework: Data Science, Analysis of Algorithms, Data Mining, Big Data Analytics, Teaching Assistantship

KJ Somaiya College of Engineering, one of India's top engineering schools — **Mumbai, India**
Bachelor of Technology in Computer Engineering, CGPA: 9/10 — *May 2018*
Focus Areas: Machine Learning, Database Management, Software Development, Software Project Management

SKILLS

Python, Data Science, Machine Learning, Deep Learning, NumPy, Pandas, scikit-learn, Keras, TensorFlow, PyTorch, OpenCV, spaCy, Hive, MongoDB, SQL, Flask, FastAPI, Streamlit, React.js, Git, Docker, Linux, Agile Methodologies, LaTeX

WORK EXPERIENCE

Amazon Web Services — **Seattle, WA**
Software Development Engineer Intern — *May 2022 – Present*

- Developed and deployed a responsive User Interface for an upcoming product using React.js, and Python, including unit testing (snapshot testing) using the Jest framework, and other Amazon-internal tools & technologies
- Demonstrated ability to work independently & as part of a team by taking ownership of tasks & collaborating with others

SUNY Research Foundation, Dept. of Biomedical Informatics, Stony Brook University — **Stony Brook, NY**
Senior Research Project Assistant - NLP — *Mar 2022 – May 2022*

- Conducted research using NLP technologies including spaCy, BERT, and Zero-shot text classification on the PASC data
- Collaborated with professors to analyze & interpret research findings, then presented recommendations based on data

Jio Platforms, a subsidiary of **Reliance Industries Ltd.**, a top conglomerate of India — **Mumbai, India**
Data Scientist — *Jul 2018 – Jul 2021*

- Devised a Document Validation System to automate the processing of KYC documents reducing the time from 15 minutes to 2 minutes for over 5000 documents daily, using Python, OpenCV, OCR, Fuzzy String Matching, and Kafka
- Conceptualized an algorithm to find potential duplicates from Material Master Data; estimated to have 10-40% cost savings and reducing the efforts of MDM users by 50%, in 3 sprints
- Executed an Intelligent Incident Management System leveraging Natural Language Processing to automatically categorize tickets and search for past resolutions of incidents with an average response time of 20 milliseconds
- Created a prediction model for MRO Materials using Machine Learning with an accuracy of 90% for 95% of all materials
- Built a medication search service for RF Hospital utilizing Natural Language Processing and Trie data structures for ultra-fast search times with a mean response time of 30 milliseconds
- Awarded with the R-Sammaan Recognition Awards by senior leaders and received various recommendations on LinkedIn for designing and accomplishing projects that solve complex business use-cases, and delivering optimal results

ACADEMIC PROJECTS

Job Title Analysis (Stony Brook University) [Link] — **Oct 2021 – Dec 2021**

- Developed a model to predict salaries by using Natural Language Processing techniques on Job Title & Skills, and combining it with ordinal features such as Education Levels using Machine Learning Algorithms having an R^2 Score of 0.901
- Performed clustering on Job Titles by considering the required skills using Fuzzy String Matching, TF-IDF Vectorizer, and DBSCAN Clustering to achieve a Silhouette Coefficient of 0.134, and deployed the User Interface using Streamlit

Website Fingerprinting using Deep Learning (Stony Brook University) [Link] — **Oct 2021 – Dec 2021**

- Reimplemented the paper "Automated Website Fingerprinting through Deep Learning" by developing a Convolutional Neural Network using PyTorch and Random Forest Classifier using Scikit-learn to achieve an accuracy of 94.02%

Personalized Web Search based on User Profiling (KJ Somaiya College of Engineering) — **Jul 2017 – Mar 2018**

- Implemented Topic Modeling using Latent Semantic Indexing (LSI) and word2vec models; dynamically created hierarchical clusters of browsing history to re-rank the Search Engine Results Page for personalized results
- Published the paper "Mining User's Browsing History to Personalize Web Search" in 2018 ICICCT, IEEE [Link]