Samiksha Khare

skhare@scu.edu | (669) 377-6377 | Sunnyvale, CA | linkedin.com/in/samiksha-khare | github.com/samiksha-khare

SUMMARY

Currently pursuing a Master's in Information Systems with 7 years of experience as a QA Automation Engineer, now transitioning into a Data Engineering role. Proficient in Python, SQL, Apache Spark, and Machine Learning. Passionate about leveraging big data technologies and machine learning skills to build scalable solutions, tackle complex problems, and contribute to innovative projects.

EDUCATION

Santa Clara University (*Master of Science in Information Systems*) – 3.8 CGPA — Jan 2024 – Present
Coursework: Data Analytics - Python, Object Oriented Programming, Database Management System, Big Data Modeling, Natural Language Processing, Business Intelligence and Data Warehousing

R.G.P.V University, India (*Bachelor of Engineering in Electronics & Communication)* — July 2010 – June 2014

SKILLS

Programming Languages: Java, Python, MySQL, Data Structures and Algorithms, HTML, CSS, JavaScript, Node.js
Frameworks & Libraries: Apache Spark, MapReduce, Kafka, TensorFlow, Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn
Tools & Technologies: Jupyter Notebook, Eclipse, Visual Studio Code, GitHub, Jenkins, AWS, Postman, Jira
Methodologies & Models: Agile, Waterfall, Scrum, Software Development Life Cycle

CERTIFICATIONS

Machine Learning Specialization - *DeepLearning.AI & Stanford University by Andrew Ng* — Sept 2024

AWS Cloud Practitioner — June 2023

PROJECTS

Crypto Real Time Analysis Using Kafka GitHub — Nov 2024
- Developed real-time cryptocurrency analysis system using Apache Kafka for streaming, MongoDB for storage, and Python for visualization
- Integrated CoinCap API to fetch live data and visualized pricing trends using Matplotlib

IPL Data Analysis with Apache Spark GitHub — Oct 2024
- Built a data pipeline for IPL cricket data using Apache Spark in Databricks, with the dataset stored on Amazon S3
- Developed transformation logic, performed SQL-based data analysis, and generated data visualization to highlight key insights

Neural Network for Handwritten Digits Classification GitHub — May 2024
- Developed a neural network model to classify handwritten digits (0–9) using the MNIST dataset
- Applied Z-Score normalization, ReLU activation, and SoftMax and evaluated performance using a confusion matrix

Mushroom Classification Using Decision Trees GitHub — June 2024
- Built a ML model to classify mushrooms as edible or poisonous using Decision Tree, Random Forest, and XGBoost, achieving high accuracy
- Preprocessed data with one-hot encoding, and optimized model performance through hyperparameter tuning and accuracy evaluation

Sentimental Analysis GitHub — Jan 2024
- Developed sentiment analysis for employee satisfaction during COVID-19, leveraging Python, Pandas, data cleaning, text preprocessing techniques to analyze satisfaction levels

World's Billionaires Statistics GitHub — Mar 2024
- Analyzed a dataset on global billionaires by performing data cleaning, exploratory data analysis, and visualization to uncover key insights
- Performed machine learning using a Random Forest Classifier to predict their industry based on features like rank, gender, and country
- Used Python libraries: Pandas, NumPy, Matplotlib, Seaborn, and Scikit-Learn

Historical Milestones GitHub — July 2024
- Developed a full stack web application to catalog and visualize historical events with key details like date, genre, description and references
- The user interface enables users to add, view, edit, and delete events, and to include reference links with indicators whether watched or not
- Developed APIs to extract, filter, and add new events to the MySQL database

PROFESSIONAL EXPERIENCE

Research Assistant — *Machine Learning Project under Prof. Manoochehr Ghiassi* — Dec 2024 – Present
- Research on enhancing Sentiment Analysis for EV market using n-gram analysis and various supervised machine learning algorithms
- Working to improve model performance which currently shows 62% accuracy, through feature engineering and fine-tuning for optimal results in predictive analytics

Tibco Software (Pune, India) — *Member of Technical Staff* — Sept 2019 – Sept 2022
- ModelOps project, a cloud-native solution designed to automate deployment, monitoring, and management of predictive models
- Developed Selenium test cases, reducing manual effort by 90%, and integrated regression tests with CI/CD pipelines using Git and Jenkins
- Verified data transmission from source to consumer using Kafka, to ensure data integrity and accurate message flow through the pipeline
- Configured and conducted installation testing for ModelOps on AWS, ensuring seamless deployment and system functionality

Cognizant Technology (Pune, India) — *Associate* — May 2015 – Sept 2019
- Worked with clients like Credit Suisse, Wells Fargo, MUFG Union Bank, and Estée Lauder, developing Selenium test cases and integrating tools to reduce manual effort by 90%
- Collaborated with global teams, define test plans, and deliver reliable solutions for applications, ensuring smooth Agile deployments