Shreya Hagalahalli Shrinivas

242 S 10th St, San Jose, CA 95112 | (408) 824-8711| hsshreya19@gmail.com | LinkedIn | GitHub |

EDUCATION

Master of Science in Computer Engineering, Data Science**:** San Jose State University Jan 19 – Dec 20
Coursework: Machine Learning, Large scale analytics, Database Management Systems, Data Mining

Bachelor of Engineering in Electronics and Communication: Bangalore Institute of Technology Aug 13 – Jun 17
Coursework: Computer Architecture, Operating Systems, Data Structures, and algorithms in C++

SKILLS

Programming Languages: Python, SQL, C++.
Databases: MySQL, SQL Server, Oracle, PostgreSQL.
Machine Learning Libraries: NLTK, Pandas, NumPy, SciPy, TensorFlow, Scikit-learn, PyTorch, Matplotlib, Seaborn, Scrapy.
Tools: Tableau, Power BI, Anaconda suite, Hadoop, Apache Spark, MS Excel (Pivot Tables), AWS EC2, S3, Jira, Slack, GitHub.
Core Strength: Relational Databases, Data Analytics, Statistics, Data Visualization, Machine Learning, Artificial Intelligence.

PROFESSIONAL EXPERIENCE

Data Science Intern | LAUNCHPAD.AI FELLOWSHIP.AI | San Francisco, CA May 20 – Aug 20

- Participated in open challenge and created a wound image dataset that led to highest accuracy of 90%. Achieved data augmentation using fast.ai and classification employing resnet34 model in python.
- Labeled and trained image data on 50k shoe products in Platform.ai. Improved accuracy of model by 33%.
- Conducted explanatory data analysis and designed an interactive dashboard in Tableau to visualize the trends of customer subscription data for Platform.AI product.
- Wrote python code to scrape all product details and images from client website using scrapy. Performed image and data processing utilizing libraries OpenCV, Pillow and NLP.
- Implemented a visual textual embedding in python for multimodal style search. Tuned the PyTorch model and documented all versions of models that were trained and tested.

Associate Software Engineer – Data Analytics | Accenture, Bangalore, India Jan 18 – Dec 18

- Scripted SQL queries to retrieve cross-functional data from oracle database. Developed and automated business report on SAP Business objects achieving 70 % acceleration in data reporting and data visualization.
- Designed an interactive KPI dashboard in tableau to visualize the trends and insights of sales data in real time.
- Developed a collaborative filtering recommendation engine with nearest neighbor approach in Python with F-1 score of 87% to analyze personalized products of interest to the customers.

ACADEMIC EXPERIENCE Jan 19 – Dec 20

LSTM Neural Networks for Stock Market Predictions [Tableau, Python, Pandas, Keras, LSTM, scikit-learn, matplotlib]

- Built a deep learning model in python with mean square error of 1.90 to predict the closing price of stock for next day.
- Created an interactive dashboard in tableau to visualize the daily, monthly, and yearly changes in close price of all the stocks. The model performance is visualized by comparing the actual and predicted close price values of stocks.

Health Insurance Database Design [MySQL, ERD, Relational Schema, Normalization, Stored procedures, triggers]

- Designed database of health insurance in MySQL. Analyzed and mapped functional requirements to ERD diagram, created schema normalized to 3NF, populated tables with sample data to build functionality using stored procedures and triggers.

Research Paper Recommendation System [Python, Pandas, NLTK, regex, Sklearn, TF-IDF, Cosine Similarity]

- Devised a pipeline hybrid model in Python for recommendation of research papers combining content and collaborative filtering approach. Data consisting of 30000 samples is scraped through arXiv API and is pre-processed.

Customer Transaction Prediction [Python, Pandas, sklearn, seaborn, Correlation, Hyperparameter tuning]

- Anticipated transaction of a customer using XgBoost and CatBoost models in python with best AUC-ROC score of 90%.
- Performed exploratory data analysis on the data set consisting of 200,000 samples with 200 anonymous attributes.
- Optimized performance of these algorithms utilizing hyperparameter tuning to select the best evaluation metrics.

Classification of Handwritten Digits [Python, Numpy, PCA, Pandas, scikit-learn, Confusion matrix, matplotlib]

- Classified handwritten digits applying KNN, Logistic Regression, and Random Forest models in Python.
- Accomplished pre-processing of data with DE skewing, centering, and shrinking dimensions employing PCA with 95% and 85% variance. Determined KNN model executed best with 94% accuracy and 95% data variance.