Malvika D Shetty

(734) 846-7974 | malvikad@umich.edu | Github: malvika-shetty | LinkedIn: malvikadshetty

Summary

Result-driven graduate student with a strong passion for analyzing big data and problem solving. Experience working with a variety of large data sets, data environments, demonstrating strong statistical and analytical skills and capabilities. Broadly skilled in **Data Science, Data Analytics, Machine Learning, Data Mining, Data Modeling, and Visualization**. Seeking full- time opportunities in the field **of Data Science and Analytics.**

Education

University of Michigan, Ann Arbor — *Apr 2021*
Master's in Computer Science and Electrical Engineering — **CGPA: 4.0**
Relevant Coursework: Data Manipulation and Analysis, SQL and Databases, Computational Data Science, Machine Learning, Computer Vision, NLP

National Institute of Technology Karnataka (NITK), India — *May 2019*
Bachelor of Technology in Engineering — **CGPA: 3.85**

Technical Experience

Programming: Python, SQL, MySQL, DASK(Big Data Tool), MS Excel, Microsoft Office, MATLAB, DataGrip, Django, HTTP, HTML, CSS, Julia
Tools and Packages: Pandas, NumPy, Scikit-learn, SciPy, NLTK, Regex, JSON, SQLite, OpenCV, TensorFlow, PyTorch, Jupyter, Google Colab
Technical Experience: Data Collection, Data Management, Data Processing and Statistical Modeling, Relational Databases, Git Control

Work Experience

Software Engineer Intern, Intrepid Control Systems, Troy, MI — *Jan 2021 - Present*

Developing a Django based Test and Data Management System

- Developing a new test management system using Django for manual and automated testing of new versions of company's data analysis software
- Deploying a full featured web site by creating forms in HTTP, HTML and CSS with to improve data collection by a factor of at least 20%
- Reporting to the Director of Engineering to finalize system features like bug tracker integration, test automation framework & visual reports

SHA(Secure Hash Algorithms) bit extraction

- Specialized in programming of data analytics platforms that communicate with vehicle ECUs with a 18% improvement in troubleshooting
- Developed a Python script with UNIX command line functionality to parse through VSB files and extract CAN(Control Area Network) messages
- Co-led the project with 2 senior team members to develop unit testing on the python scripts and find root cause of defects before software release

Data Science Intern, Intrepid Control Systems, Troy, MI — *Jun 2020 - Dec 2020*

Object Detection & Image Recognition Using DataSpy / IPA Video Analysis Feature – Video Link

- Developed a video analysis feature to detect objects using the YOLO (You Only Look Once) algorithm with OpenCV & Python with a 92% accuracy
- Performed quantitative analysis of image statistics and signals from vehicles on a common timeline in DataSpy for better insight of the signal changes
- Conducted data analysis of the signals using SQL to query DB files to ensure data quality obtained from data loggers in vehicles

Advanced Data Visualization to generate Matplotlib Charts

- Programmed a Python script that overuses a given template excel file and overhauls its individual pages with matplotlib charts
- Communicated with the Director of Business Development to enhance the company's data mining tool(DataSpy) sales by 36%
- Utilized SQL to clean, format, and aggregate data from database files containing over 3 million rows of raw data

Extracting and Analyzing Control Area Network (CAN) Statistics

- Designed a python script to extract, model and maintain data integrity of CAN signals from files containing vehicle transmission data
- Performed in-depth analytics of vehicular CAN network files to detect unique ArbId using the ICS-VSBIO python library
- Arranged, refined, and formatted the extracted data with data mining python script to store it as a table using macros in MS Excel

Projects

SI 564 - SQL and Databases, University of Michigan — *Jan 2021- Present*

- Programmed efficient SQL techniques to retrieve and manipulate data from databases using joins, group by & aggregate functions
- Developed advanced indexes by creating relationships between tables using keys & handling duplicate data in a large data sets using normal forms
- *Project*: Analyzing over 800000 rows of New York Stock Exchange data by designing 5 normalized databases to derive insights into market trends

SI 618 - Data Manipulation and Analysis, University of Michigan — *Jan 2021- Present*

- Implemented the ETL(extract, transform and load) process for data processing using text parsing, graphing & clustering methods
- Derived statistical analysis using correlation, Ordinary Least Square(OLS) regression, linear models, pivot tables, ANOVA & hypothesis testing
- Created data visualizations like line & scatter plots, bar charts, histograms, box plots & time series with Matplotlib & Seaborn packages
- Implemented and visualized parallel processing of python code with DASK Framework, Git Control and Docker to handle Big Data

EECS 504 – Computer Vision, University of Michigan — *Oct 2020- Dec 2020*

- Implemented foreground-background image segmentation optimizing max-flow/min-cut over super pixels generated by SLIC
- Extracted image features (blob and corner detection) using difference of Gaussian scale space and Harris corner detector
- *Project*: *Object Tracking for Safety* - Detected & tracked objects in a live video stream using YOLOv3 machine learning model & Deep SORT algorithm

DNA base calling with Convolutional Neural Networks (CNN) & Gated Recurrent Units (GRU) — *Jan 2020- Apr 2020*

- Worked with a team to implement configs of modified CNN models to build a computationally less expensive DNA base caller
- Successfully gained a 5x speedup within 90% of the state-of-the-art Bonito model's accuracy using CNN+GRU model

EECS 505 – Computational Data Science and Machine Learning, University of Michigan — *Sept 2019- Dec 2019*

- Applied computational methods for data modeling, identifying patterns and outliers in large data sets with graphs, clustering & linear models
- Devised python codes for eigenvalue decomposition, independent component analysis, non-linear model fitting, forecasting and deep neural networks
- *Projects:* Handwriting recognition, foreground & background subtraction in videos, filling missing entries and denoising images, & unmixing sounds

Certifications

Applied Data Science with Python:

- Introduction to Data Science in Python
- Applied Plotting, Charting and Data Representation in Python
- Applied Machine Learning in Python
- Applied Text Mining in Python

Django for Everybody:

- Web Applications Technologies and Django
- Building Web Applications in Django
- Django Features and Libraries
- Using JavaScript, JQuery, and JSON in Django

Achievements

- Gold Medalist at the 57th National Rifle Shooting Competition
- Captain of the NITK Women's Volleyball Team
- Secretary and Events Representative of NITK Artists' Forum Club
- Student Council President at Vissanji Academy