Smit Chandarana

857-472-4746 | github.com/smitchandarana

Career Objective

Data Analyst looking to learn more about Data Engineering, Machine learning. 3+ years of experience with understanding data and optimizing to be used for analysis. Looking to deep dive into Supervised, Unsupervised and reinforcement learning methods looking forward to building scalable machine learning products and data processing pipelines to be used for data analysis.

Skills

Languages:	Python (Pandas, NumPy, Scikit-Learn, Pyspark), SQL
Machine Learning	A/B Hypothesis Testing, Predictive and Sentiment Analysis, NLP, Tensorflow, Keras, PyTorch, OpenCV
Database:	MS Excel (VLOOKUP, Power Query), MySQL, MongoDB, Tableau, PowerBI, Seaborn, Matplotlib
Others:	AWS, GitHub, Jira, Jenkins, BitBucket, Shell Scripting

Professional Experience

Data Analyst Engineer – Fuel Cell Lab, General Motors   **Jan'22 – Oct'22**
- Collaborated with Stakeholders cross-functionally across teams to identify business problems and developed a Data Pipeline for collecting data and dashboard using Power **BI and JIRA** which reduced the man power by 20% and increased accuracy by 38%.
- Used Exploratory Data Analysis (EDA) using **Python, Numpy, Pandas,** and **Matplotlib** to identify important KPI's which had higher effect on test results of Fuel Cells.
- Developed ETL (Extract, Transform and Load) code components like mapping, workflow, and database objects like stored procedure, functions from business requirements and design plan.
- Worked with **REST API** and **ODBC** connection to manipulate and load data from the application for Complex Event Processing and Real-Time Event Processing.
- Worked with peers to implement full Agile methodology to manage projects and reduced downtime by 28% in first three months.
- Designed multiple Dashboards in PowerBI for process improvements and increasing efficiency of the business

Data Analyst Engineer, Microsoft   **Sept'21 – Jan'22**
- Implemented data warehouse for data modeling, Functions, Store procedure and normalization on Microsoft SQL server.
- Developed **ETL framework using PySpark** (including daily runs, error handling, logging) to useful data.
- Coordinated with team and Developed framework to generate Daily adhoc, Report's and Extracts from enterprise data.
- Automated the data mining and cleaning by creating ETL pipeline, reduced the process time by 85% and increasing the accuracy by 30%
- Compared VM attributes from different touch points and identify discrepancies using Excel and VLOOKUP
- Collaborate with Project Managers to identify correct attributes for the data extracted, which is updated in a CSV files.

Data Analyst, DreamingCode, Boston   **June'20 – Sept'21**
- Generated a series of sales report daily using advanced MS Excel for supply chain team reducing operational time by 12%
- Established KPI by collaborating with other cross functional teams to optimize the financial ad hoc reports increased the efficiency of 38%

Machine Learning - Research Assistant, Northeastern University   **April'19 – Sept'19**
- Created **ETL** pipeline for data collection, cleaning and visualization of data from sensors using python.
- Built **TensorFlow Lite** and **Keras** based deep learning model with 84% accuracy for predication of laptop location on the wireless charging.
- Analyzed research papers for its best fit models, presented and defended the finding of the same to fellow peers and PhD scholars.

Reporting and Insight Analyst, Softvan, India   **Sept'17 – Aug'18**
- Created visualization dashboards on **Tableau** for more than a dozen online reports helping clients identify opportunities for more ads.
- Worked closely with the UX group to upgrade and enhance our platform and the user experience by creating new interfaces and infographics depicting complex data sets using **Python, Pandas, Seaborn and Matplotlib** and **PowerBI**

Academic Projects

Analysis of FinTech Hiring Pattern (EDA, BeautifulSoup, Numpy, Scipy, Pandas, Matplotlib, Tableau, ETL)   **Jan '19 - Feb '19**
- Collaborated with 6 teams to gather, clean and summarize the reports about FinTech and created list of 100 keywords using TF-IDF and Text rank, Numpy and Pandas and created CSV files for the same using data wrangling.
- Visualized outcomes like key areas, skills, hiring trends based on location and job sectors using Tableau and Seaborn.

Prediction of Interest Rate (Linear Regression, Neural Network, AutoML, Pandas, Numpy, Tableau)   **Feb '19 - Mar '19**
- Prepared data by extensive EDA, data cleansing and preprocessing, applied Multiple Imputations by Chained Equations (MICE) to fill missing values, used LassoCV Regularization for feature selection in Python environment with 56 different variables.
- Built models like Multiple Linear Regression, Random Forest and Neural Networks to predict lending club interest rates and summarized their respective mean absolute percentage errors (MAPE) , RMSE and MAE, for both training and test data.
- Analysed data with Tableau for various findings, results gained out of project and presented to an audience of 50 students.

Education

Northeastern University, Boston, MA   **Sept'18 - May '20**
Master of Science in Computer Engineering (*Internet of Things)*
Gujarat Technological University, Ahmedabad, Gujarat, India   **Aug '14 - Jun '18**