Shubham Jain
shubhamj@andrew.cmu.edu | 412-915-3982 | LinkedIn | GitHub

EDUCATION

Carnegie Mellon University, Pittsburgh, PA 12/2022
Master of Information Systems Management
Courses: Data Mining, Unstructured Data Analytics, Data Focused Python, Computational Data Science, Object Oriented Programming in Java, Distributed Systems

Vellore Institute of Technology, Vellore, India 05/2018
Bachelor of Technology in Computer Science
Courses: Data Structures and Algorithms, Agent Based Intelligent Systems, Soft Computing

SKILLS

Languages: Java, Python, C, C++ | **Databases:** MySQL, Oracle 11g | **Scripting**: PowerShell, Perl
Libraries: Numpy, Pandas, Matplotlib, Seaborn, Scikit-Learn, Spacy, BeautifulSoup, PyTorch, Tensorflow
Functional: Linear Regression, Logistic Regression, Decision Trees, Random Forest, PCA, Gradient Boosting, XGBoost, SVM, Hypothesis testing**,** Clustering, Hyperparameter tuning
Tools: GIT, MDT, TFS, SAS Enterprise Miner, Tableau, Rapid7, Jupyter Notebook, Eclipse IDE, IntelliJ

ACADEMIC PROJECTS GitHub

Amazon Review Rating Prediction (Logistic Regression, SVM, KNN, Random Forest) 04/2022
- Conducted EDA, data preprocessing and feature construction by implementing BOW and TF-IDF models for textual data.
- Trained models such as Logistic Regression, Random Forest, SVM, KNN; finalized with random forest with an accuracy of ~88% for rating prediction of reviews.
- Deployed it to public API endpoint using Azure ML studio.

Image Classification (Logistic Regression, LeNet, AlexNet, Azure, Pytorch*)* 11/2021
- Trained classification neural networks such as Logistic Regression, LeNet and AlexNet to carry out image classification task on 60,000 color images dataset.
- Performed hyperparameter tuning and evaluated model performances; with highest accuracy of 82% on AlexNet CNN model.
- Created a public endpoint API and deployed on Azure.

Instacart Customer Order Forecasting (Python, Tableau, Market Basket Analysis) 02/2021
- Built interactive dashboards on Tableau to do initial data analysis on 8,000,000+ rows of data of Instacart grocery business.
- Executed feature engineering to aggregate most important attributes to identify consumer purchasing patterns.
- Built a random forest model in python to predict the customers' next order items with an accuracy of 88.5%.
- Implemented market basket analysis to identify associations between products and provide recommendations to customers.

Predicting Status of H-1B Applications (Python, Naïve Bayes, Logistic Regression and Random Forest) 06/2020
- Conducted exploratory analysis, data preprocessing, and feature selection for 3,000,000+ rows of H1b Visa applications data.
- Experimented with models like Naïve Bayes, Logistic Regression, Random Forest; evaluated models using AUC curve as evaluation metric.
- Enhanced the efficiency of the models using cross-validation and tuning hyper parameters by using Grid Search.

WORK EXPERIENCE

Data Analyst Intern, Highmark Health, Pittsburgh, USA 05/2021–08/2021
- Offered approaches to simplify the analysis process and created an automated framework with Python and SQL for vulnerability analysis of different business platforms, lowering their turnaround time from 45 to 4 minutes and saving $120,000 - $190,000.
- Designed a data pipeline consuming raw scans of vulnerability data generated from various discovery sources and reporting them for remediation from the findings, cutting the remediation turnaround time by ~50% using Python.
- Presented project findings to technical and non-technical stakeholders in a succinct manner and incorporated feedback.

Software Engineer, Philips HealthCare Systems, Bangalore, India 07/2018-06/2020
- Worked on upgrading of windows operating system's bios type from 'Legacy' to a secure type 'UEFI' boot on MR systems.
- Reduced annual spendings by 25% by upgrading the MR backup and restore tool originally integrated with 'Acronis' to 'Wbadmin' for volume level backup using C#, WPF.
- Developed an automation suite for windows operating system deployment with complete MR SW configuration for different hardware platforms, decreasing the turnaround time from 14man days to 4 hours in C#, Perl, PowerShell.
- Optimized processes of MR workflow programmed in C, C++ and managed to increase the efficiency from 'E' to 'A'.

ACHIEVEMENTS

Team Up to Win Award, Philips Culture Awards, Philips Healthcare Systems 04/2019
Student Head - Institution of Electronics and Telecommunication Engineers (IETE-VIT) 05/2017 - 04/2018