VARAD S. TUPE

varad.s.tupe@gmail.com | 716-328-8518 | about.me/varadtupe | linkedin.com/in/varadtupe | github.com/varadtupe

EDUCATION

University at Buffalo, The State University of New York Sep 2018
Master of Engineering Sciences (Data Science) **GPA: 3.67/4.00**
Relevant coursework: Machine Learning, Data Intensive Computing, Database System, Statistical Data Mining

University of Mumbai, India July 2015
Bachelor of Engineering, Information Technology **First Class**
Relevant coursework: Data Structures & Algorithm, Object Oriented Analysis & Design

SKILLS

Programming Languages: Python, R, JavaScript, Scala, Shell Script, PL/SQL, Java, PHP, CSS, C++
Packages: NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, Folium, Plotly, Tensorflow, Keras
Databases: Oracle, PostgreSQL, MySQL, MS-SQL
Operating Systems: Ubuntu, RHEL, Windows OS, Mac OS
Tools: Tableau, Docker, Apache Airflow, SVN, GitHub, GitLab, SQL Developer, Eclipse, MS Excel
Big Data: Hadoop, MapReduce, Hive, Pig, Spark
ML: Regression, Decision Trees, Random Forest, SVM, PCA, Neural Networks, NLP, Dimension Reduction, Deep Learning

WORK EXPERIENCE

Analytics Developer Intern, Quantworks Inc, *Raleigh, NC, May 2018 – Aug 2018. (Python, Postgres, Docker, ETL, AWS, Airflow)*
- Development of Smart Abstractor Assistant for extracting key highlights from medical data in PDF format using NLP techniques.
- Research of negation detection algorithm in order to reduce false positives.
- Built Data Pipeline using Apache Airflow and integration with the Fanalytical system.
- Provisioned the infrastructure and templatized code for loading data from CSV, Salesforce, Mailchimp REST API, Amazon S3 bucket.

Software Engineer, Vistaar Systems, *Mumbai, Jul 2015 – Jun 2017 (JavaScript, SQL, ETL, RestAPI, Shell Script, Price Science)*
- Gathered, analyzed, designed and implemented enhancements by clients.
- Built, enhanced & maintained the science-based recommender engine responsible for delivering target prices.
- Responsible for planning and estimation of releases and represent Vistaar in SCRUM meeting with clients.
- Developed utilities to track the system state, usage, performance with report generation capabilities which was further used to analyze the system.
- Pivotal role in migration of solution from on-premise to AWS server.
- Key contributor in optimizing the performance of operations, ETL, REST API calls, migration of system improving execution time by 40%.
- Collaborated for documentation and best practices template to be used by teams.
- Lead and mentor junior team of 3.

PROJECTS

Spotify Music Recommendation using Clustering Algorithms, Spring 2018 *(Python, R, Clustering Algorithms)*
Analysis of experimental recommendation system of songs included in playlist using audio features extracted from Spotify & Gracenote's Web-API. Data was transformed using T-Distributed Stochastic Neighboring Entities (T-SNE) for dimension reduction and fed into Hierarchical clustering, K-Means, K-Medoids model. *(Currently working on making it a standalone application)*

New York Times News classifier using SparkML, Spring 2018 *(Python, PySpark, SparkMLib, Classification Algorithm)*
Built a News classifier on articles fetched from NY Times Article Search API and Beautiful Soup. Trained Logistic Regression and Naïve Bayes model with accuracy of 91% post cross validation in Spark Environment.

Predicting Employee Satisfaction & Attrition using Random Forest *(Source: Kaggle)*, Fall 2017 *(R, CART)*
Built models to predict employee satisfaction using regression and decision tree techniques. Trained a Random Forest model for predicting employee's exit from company with accuracy of 94%. Recommended focus areas to lower attrition rate and keep employees satisfied.

Chicago Taxi Trip Analysis *(Source: Chicago Data Portal)*, Fall 2017 *(Python, MySQL, Folium, Regression Algorithms)*
Analyzed trends of users' taxi rides and recommended hot zone for taxi companies to increase pickups. Understood taxi traffic flow by community area and duration of day which helped identifying high frequency taxi routes maximizing driver's trips. Extracted the correlation of trip distance and duration on fare of the trip and built linear regression model predicting fare for trips.

Location Based Advertising, 2015 *(Java, PHP, MySQL, Android Studio, HTML, CSS)*
Android application to facilitate user to search for offers/deals provided by local businesses to increase footfall. Scope focusing on small scale businesses in order to compete with large scale and online giants. Devised an algorithm to fetch the offers based on user's location, category preference and popularity of offer. Developed a website to administer clients and their offer data.

CERTIFICATIONS

Python for Data Science and Machine Learning Bootcamp, Hadoop Starter Kit, Natural Language Processing (NLP) in Python,
Udemy, 2018