YATHARTH GARG

ygarg020@gmail.com | (682)-521-6466| linkedin.com/in/yatharthgarg | github.com/yatharthgarg

EDUCATION

MS in Computer Science – *University of Texas at Arlington.* **Jan 2017 - Dec 2018**
Specialized in Data Science and Artificial Intelligence

B.Tech in Computer Science – *ITM University* **Jul 2012 - May 2016**
Minor in Data Science and Economics

EXPERIENCE

NLP Data Scientist – *Verizon (Dallas, TX)* **Jul 2019 – Present**

- Developed machine learning models using **Python and Tensorflow** to predict the probability of email escalation to decrease email processing time by upper management.
- Collected and analyzed millions of customers' email data using **Tableau** to identify patterns among the features and predicted escalation probability and reduced the false positives by 80% to minimize the workload for the business team.
- Collaborated with both data science and business teams and communicated the technical details to business leaders and helped to put the models in the production stage using **REST API and Python Flask**.
- Documented, managed, and tracked using agile methodologies for project portfolio using **Confluence, Jira, and GitLab.**

Research Assistant - *University of Texas at Arlington (Arlington, TX)* **Feb 2019 – Jul 2019**

- Perused the accents of people from 177 countries and built a speech-accent recognition application using **CNN** to predict the backgrounds from their accents.
- Worked in a team of three and converted the audio files to different formats to extract the maximum features and minimized the noise in the files to set the threshold for audio.
- Extracted data by web-scraping using **Beautiful Soup** from Speech-Accent archive and analyzed the MFCC features from over 2000 audio samples resulting in an improved model accuracy from **65% to 80%.**

Data Scientist – *Acequare (Gurgaon, India)* **Aug 2015 – Dec 2016**

- Designed and built machine learning classification models using **Python** to predict the application user type (new, frequent, sporadic) to provide them with reasonable deals and benefits based on user preferences and increase company growth.
- Partitioned terabytes of data using **Hive** to reduce the computational power by almost 90% for examining and rendering visualizations on **Tableau**.
- Used and developed ensemble learning methods and improved model performance by 10% and helped other software engineers in the deployment of the project on **Python Flask.**

SKILLS

- **Programming Languages -** Python (NumPy, Pandas, SciPy, Scikit-Learn, Matplotlib, NLTK), MATLAB, C++, SQL, Python Flask, R
- **Frameworks and Visualization Tools-** Keras, TensorFlow, PyTorch, H2O.ai, D3.js, Tableau
- **Environments and OS–** CI/CD (Jenkins), Windows, UNIX, LINUX (Ubuntu), Agile Methodology
- **IDE and Tools -** Jupyter Notebook, Spyder, Visual Studio, XCode, Git
- **Cloud Technologies –** Amazon Web Services, Google Cloud
- **Big Data Technologies –** Hive, Spark (PySpark)
- **Techniques and Algorithms -** Regressions, Classification (Logistic Regression, CART, XGBoost, Naive Bayes, SVM), Deep Learning, Neural Networks, Regularization (Ridge, LASSO), NLP (TF-IDF, Word2Vec, Gensim, SpaCy), Dimensionality Reduction (SVD, PCA), Web Scrape (Beautiful Soup), REST API, Statistics (Hypothesis testing, p-test, t-test) and Mathematics.
- **Essential tools and methods –** Microsoft Office (Excel, PowerPoint, Word), Putty/WinSCP, JIRA, Confluence

PROJECTS

Reinforcement Learning using WoLF-PHC algorithm

- Varied the learning rates for two agents to support the convergence of the algorithm. The main idea behind this algorithm is learn quickly while losing and slowly while winning. The specific method for determining when the agent is winning is by comparing the current policy's expected payoff with that of the average policy over time.