ARUN AMBALLA

510-994-9343| arunamballa24@gmail.com |LinkedIn | Medium | GitHub

SUMMARY

- Hands-on experience in **Machine Learning** and **Natural Language Processing** techniques and principles.
- A self-motivated, detail-oriented team player with good communication skills and an active **Data Science open resource contributor.**

EDUCATION

- **MS in Computer Science from California State University, East Bay** (GPA:3.97). (Aug 2021 - May 2023)
- **Bachelor's in Computer Science from MGIT, India** (GPA:3.86) (Aug 2016 - Sep 2020)

TECHNICAL SKILLS

- **Languages** : **Python**, Java, C, C++.
- **Databases** : **Oracle**, **MySQL, MongoDB.**
- **Web** : HTML5, CSS3, JavaScript, ReactJS, Spring MVC, Spring REST, Hibernate, JPA,PHP, Django, Flask.
- **Cloud** : **AWS (**IAM, EC2, Elastic Bean Stalk, S3, Lambda, SageMaker)
- **Tools** : **TensorFlow, Keras, fast.ai, PyTorch, Gensim, Scikit-learn,** OpenCV, NLTK, Scikit-Multilearn.

EXPERIENCE

Social Tek, Hyderabad, India. (Machine Learning Intern) (Jan 2021 – April 2021)

Quora Question Pair Similarity (Python,NLTK, Scikit-learn, Gensim, Keras, fuzzywuzzy)

- Developed a Machine Learning model to find **semantic similarities** between pairs of questions.
- Term Frequency-Inverse Document Frequency (TF-IDF) weighted **Word2Vec** is used as a word embedding technique along with features extracted using Fuzzy-Wuzzy.
- A Majority Voting Classifier with **Random Forests, Adaptive Boosting** and **Gradient Boosting** Techniques are used for classification.

Amazon Apparel Recommendation System (Python,NLTK, Keras, Gensim)

- Developed an Apparel Recommendation System based on the title, brand, color, price, and image of the product.
- Text Vectorization is performed using **average-weighted fastText**,**TF-IDF weighted fastText** and feature extraction from images is performed using Transfer Learning and Fine Tuning and experimenting with **VGG-16** and **weighted Euclidean** distance is used as a metric to find similar products.

Saven Technologies, Hyderabad, India. (Machine Learning Intern). (April 2019 – May 2019)

Stack Overflow Question Tagging (Python, Scikit-multilearn, NLTK,Scikit-Learn)

- Developed a Machine Learning Model to Predict Tags Based on the **title and description** of a question.
- Used **Binary Relevance** to transform **Multi-Label** problem into **Multiple Binary Classification problems.**
- **Logistic Regression** with **L1- regularization** was used for Classification and **GridSearchCV** was used for hyper-parameter tuning.

ACADEMIC PROJECTS

Plant Disease Detection from Images (Python, Keras, TensorFlow, fast.ai, Flask, S3, EC2)

- Developed a Deep Learning model by performing Transfer Learning and Fine-Tuning and is experimented with **ResNet50** using **Differential Learning Rates** for different parts of the network and achieved state of the art results with an **accuracy of 99.56%.**
- Deployed model on EC2 Instance and S3 bucket is used to store the input images for retraining the model.

Personalized Cancer Diagnosis (Python, NLTK, Scikit-learn, Gensim, Keras, Flask, EC2)

- Developed a Machine Learning Model to classify the genetic variations based on evidence from text based clinical literature.
- **Extreme Gradient Boost** (XGBoost) and **Gradient Boosting** with hyperparameter tuning are used for classification and text representation is performed using average weighted fastText and average weighted word2vec.

CERTIFICATIONS

- **Elite Gold Certification** in "**Deep Learning**" from Indian Institute of Technology Kharagpur.
- **Elite Gold Certification** in "**Python**" from Indian Institute of Technology Madras.