Gayathri Pulagam

gayathri.pulagam@sjsu.edu github.io/gyeah3 gyeah3

EDUCATION

San Jose State University
MS Software Engineering (Data Science) 2021
Web and Big Data Mining
Deep Learning
Reinforcement Learning
Systems Engineering

Jawaharlal Nehru Technological University
BS Computer Science 2017
Data Structures and Algorithms
Design Patterns
Operating Systems
Distributed Systems

SKILLS

TECHNICAL SKILLS: Python, Java, C++, Spark, Data Analysis, PyTorch, TensorFlow, SQL, React.js, Docker, R, Data Modelling and Evaluation, Data Visualization

PROJECTS

Credit Card Fraud Detection — May 2020 - July 2020

- Built a binary classification model using **SageMaker's LinearLearner** that can identify transactions as either fraudulent or valid
- Improved the accuracy of the model upto **95.9%** by decreasing class imbalance and tuning the hyperparameters

Sentiment Analysis on User Reviews — June 2020 - July 2020

- Designed an interactive web application which takes in text as an input and analyzes it to predict the sentiment
- Processed the data using **NLP** to homogenize the data
- Developed an **XGBoost** model to classify the input data's sentiment quotient

Dog Breed Classifier — Mar. 2020 - Apr. 2020

- Built a data processing pipeline which can be used within web or mobile app to process real-world, user-supplied images to identify an estimate of a canine's breed
- Trained the model using **CNN (Transfer Learning)** to improve accuracy of the image classification

Plagiarism Detector — Feb. 2020 - Mar. 2020

- Built a custom **PyTorch neural network classifier** to detect varying levels of plagiarism in the input text data
- Extracted meaningful features like containment features to calculate n-gram range and **LCS** to catch varying levels of plagiarism in the input text
- Deployed the model to a web application using **AWS API Gateway** and **AWS Lambda**

Population Segmentation — July 2020 - Aug. 0200

- Employed two unsupervised learning algorithms to do population segmentation to find natural groupings in population data that reveals some feature-level similarities between different regions in the US
- Using **PCA**, reduced the dimensionality in original census data and then used **k-means clustering** to assign each US county to their respective clusters

Facial Expression Recognition — Jan. 2020 - Feb. 2020

- Built and trained a **CNN in Keras** from scratch to recognize facial expressions
- Used **OpenCV** to detect faces in images and classify each face based on emotion into 7 different categories
- Deployed the trained model to a web interface to perform real-time facial expression recognition on video and image data