Tavishi Priyam

(720) 951-9786
Boulder, CO
tavishi.priyam@colorado.edu

Data Scientist

GitHub: tpriyam
LinkedIn: tavishi-priyam

Inquisitive data scientist passionate about solving real-world problems using Machine Learning and Artificial Intelligence. Proficient in researching, analyzing, processing and visualizing big datasets and training efficient, well-performing models

Skills

Tools and Languages Python, R, SQL, TensorFlow, PyTorch, Keras, scikit-learn, SciPy, Pandas, NumPy, seaborn, ObsPy, Weka, Jupyter Notebooks, Excel, C/C++, HTML, CSS, MATLAB, Java, RStudio, Anaconda, Git, AWS

Courses Machine Learning, Natural Language Processing, Data Structures & Algorithms, Data Science and Big Data Analytics, Probability and Statistical Inference, Database Management, Deep Learning, Neural Networks

Technologies Natural Language Understanding, Information Retrieval, Machine Translation, Predictive and Generative Algorithms, Classification, Sentiment Analysis, Data Mining, Feature Engineering

Projects

Speech and Opinion Recognition using an Ensemble Classifier — *Python, NLTK, scikit-learn, numpy, matplotlib, Flask*

- Designed and trained an ensemble natural language classifier to detect sentiment of spoken text with an accuracy of 94%
- Ensemble comprised of Naive Bayes, Support Vector Machine, Logistic Regression and Decision Trees with dynamic weights
- Performed dimensionality reduction, removed duplicates and stopwords, created vectors using TF-IDF on Amazon Reviews
- Deployed the classifier using web app with flask backend and used CMU Sphinx API to convert speech-to-text

Visual Question Answering using Neural Networks — *Python, TensorFlow Keras, Feature Extraction, VGG16, Binary-encoding*

- Trained deep neural network using Tensorflow Keras to predict an object using an image and a question as input
- Cleaned and pre-processed a subset of the VizWiz - VQA dataset with 2000 training and 500 validation images
- Extracted features from images using VGG16 pre-trained on imagenet data and from questions using binary encoding
- Achieved a validation accuracy of 51% using the VQA accuracy metric which is more than the 47% reported by the VQA paper

Online Loan Prediction and Banking Application — *Python, SVM, Flask, PostgreSQL, Object-Oriented Design, React*

- Implemented a Support Vector Machine model to predict loan amount for clients
- Designed a react web app to register and process loan applications using flask and PostgreSQL as database server

Education

Masters of Science in Computer Science, *University of Colorado Boulder* — *GPA: 3.9/4.0* — Dec 2022

Bachelors in Technology, *SRM Institute of Science and Technology Kattankulathur* — *GPA: 9.0/10.0* — Jun 2020

Experience

Research Analyst Intern / Data Analytics and Modelling — **May 2022 — Present**
Ruder Finn – Cision, BrandWatch, Excel, Python — *New York, NY*

- Performed market research for over 10 clients using social and traditional media data
- Analyzed and investigated the data to find insights that form the basis of improved market strategies
- Implemented data science techniques like sentiment analysis to investigate public opinion of clients and their products
- Delved deep into enterprise analytics tools such as Cision and Brandwatch to analyze and process big data

Graduate Research Assistant / Seismic Event Detection from Volcanic Regions using Machine Learning — **Jan 2022 — May 2022**
University of Colorado Boulder – Python, PyTorch, NumPy, pandas, scikit-learn, matplotlib — *Boulder, CO*

- Extracted and processed one year of seismic time-series data from GeoNet using ObsPy, visualized using dayplots and trace plots
- Trained a deep neural network using PyTorch with multiple convolutional, pooling, and fully connected layers
- Processed the time-series data and engineered different features using Wavelet transform
- Used engineered features as input to the multilayer perceptron instead of the raw time-series
- Analyzed performance to identify the best model based on accuracy, precision, recall, and computational efficiency
- Presented a poster at CIRES Rendezvous 2022

Academic Intern / Artificial Neural Networks — **Jan 2018**
National University of Singapore – Python, R, Weka, scikit-learn, ggplot, matplotlib — *Singapore*

- Designed a gradient boosting model (XGBoost) model to predict New York City Taxi trip duration from a dataset on Kaggle
- Created a random sample of 10000 from 1.4 million initial data points and cleaned the data by removing noise, outliers
- Visualized the spatial-temporal data in R and Python using ggplot, matplotlib on attributes including speed, month, time
- Performed feature engineering to extract features such as haversine distance, speed, total distance, maneuvers made by car
- Experimented with various predictive learning algorithms such as logistic regression, support vector machine, naive bayes classifier, and multilayer perceptron using Weka, scikit-learn, TensorFlow keras

Academic Intern / Big Data Analysis — **Dec 2017**
Hewlett Packard Enterprise Education Services – Hadoop, MapReduce — *Singapore*

- Built a multiple node Hadoop cluster from scratch and performed MapReduce operations

Publications

T. Priyam, AMJ Muthukumaran and H. Vinayak, **"Speech and Opinion Recognition from a Conversation,"** International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-6, April 2020