HUA, Yuchen (April)

linkedin : https://www.linkedin.com/in/yuchen-hua-70505917b

Mobile: (+1) 413-461-8023 | Email: yuchen_hua@brown.edu / yuchenhua96@gmail.com

EDUCATION

Brown University (GPA:3.8/4)	**M.S. in Data Science**	**08/2020 - 12/2021**
Brown University	**M.S. in Physics**	**08/2018 - 05/2020**
East China University of Science and Technology (GPA: 3.5/4)	**B.S. in Applied Physics**	**08/2014 - 06/2018**

TECHNICAL SKILLS

- **Programming language & Applications:** Python, SQL (Postgres, MySQL), Scala, Tableau, R, Matlab, Javascript, HTML, Github, Latex, Mathematica, Julia
- **Data visualization:** Tableau, Power BI, Tensorboard, Matplotlib, Seaborn, Ggplot2
- **Packages:** Tensorflow, Keras, Pandas, Numpy, Scikit-learn, BeautifulSoup, Gym, SciPy, PySpark, Dplyr, Tidyverse

WORK EXPERIENCES

- **AI/ML Research Intern at SoftInWay Inc.** *(Burlington, MA)* **07/2021 – 10/2021**
- **Axial compressor performance prediction project**
  - Use **AutoKeras** to build deep learning models to predict axial compressor performance, with metrics efficiencies and pressure rate. Automatically find the best model architecture and hyperparameters accordingly given different compressor features as input results in prediction efficiencies and pressure rate error (RMSE) around 1e-5.
  - Use **Gaussian Mixture models** (GMM) to generate a synthetic dataset that has similar statistical distribution with the current dataset to improve models' performance on a small dataset. Result in RMSE improves from 1e-4 to 1e-5 by adding the synthetic data.
  - Deliver a ready-to-use model for automating turbomachinery design that enables people with limited machine learning backgrounds to use the models easily, saving the cost of design and smoothing the collaboration between the AI research team and the engineering team. Deliver reports, visualize results, and explain to non-machine learning background audiences and other collaborative teams.
- **Research Assistant at Brown University** **06/2021 - current**
- **Predict Patient Anxiety Using Machine Learning during Radiation Cancer Treatment**
  - Use **Python**, R, and MATLAB to preprocess physiological responses data acquired from wearable sensors of patients, perform **feature engineering** to get valuable features. Check correlation using **Pearson correlation** between skin conductance and acceleration data to investigate if motion affects skin conductance data and denoise the data accordingly.
  - Use Python to implement machine learning models, including **unsupervised** learning and **supervised** learning models. **Cluster** segments data into different categories and compares them with tagged events during the CT (cancer treatment) sessions, providing insightful information and evaluating the marking method.
  - Build classification models use **Logistic regression**, **XGBoost**, **Random Forest** algorithms to detect whether patients have anxiety during their CT sessions. Visualize and analyze the prediction results and give insightful information on how to help better plan for more efficient and safe radiation therapy treatment.
- **Teaching Assistant at Brown University** **01/2021 - 05/2021**
  - Assist design **statistical testing procedure**, including using **multiple hypothesis tests** such FWER (Family-wise error-rate) control to analyze real-world clinic trial data, aims to explore if the newly developed drug alleviates cancer disease symptoms by comparing between control and experiment group
  - Assist in designing and implementing **MapReduce** algorithms equivalent with mappers and reducers in **Spark** using **PySpark** from Python. Apply the MapReduce framework to the movie **recommendation** system.
  - Assist design and implement **topic modeling**/latent semantic analysis (LSA) using tf-idf scores in natural language processing (NLP), implement **classification** (KNN) and **clustering** (k-means) algorithms to category documents.

PERSONAL & COURSE PROJECTS

- **Movie Revenue Forecast Before Its Release**
  - **Web scraping** IMDB movies data and perform sentiment analysis using Twitter API, pre-process, clean missing data and obtain 300k data points. Implement tree-based models such as **Random Forest**, **XGBoost** dealing with multiple categorical data for predicting movie revenue to get insightful information to help the investor make a better investment plan. Achieve R2 0.56 from the best model and successfully detect revenue peaks and trends.
- **Credit Card Fraud Detection** *(Self-motivated project)*
  - Implement Tree-based models, such as Random Forest, GBMs to predict transaction type (fraudulent/normal). Implement SMOTE algorithm to upsample the data to deal with highly imbalanced data nature. Improve prediction accuracy from baseline 98.4% to 98.6%
- **Image Classification with Transfer Learning** (Cassava Leaf Disease Classification)
  - Use Tfrecord to load the large dataset. Implement **transfer learning** with pre-trained **ResNet**, **Inception**, and **VGG** as feature extractors and fine-tune them with current data for faster training and more accurate results. Achieve prediction accuracy of 0.87 from baseline accuracy of 0.6. Use prediction results to reduce production costs by avoiding using defective leaves.
- **Human Motion Recognition (Classification) with 51 classes**
  - Combine sequential model (**bi-directional LSTM**) with pre-trained **MobileNet** to capture data sequential nature, fine-tune with 51 classes video dataset. Boost model performance with ensemble method from baseline 20% to 74% accuracy.