Swayanshu Shanti Pragnya
7209498070 swayanshuspragnya@gmail.com swayanshu ssp Denver

RESEARCH & PUBLICATIONS (ResearchGate)

- The implication of Statistical Analysis and Feature Engineering for model building using Machine Learning algorithms.
- Clustering: Review on Partitioned Clustering Algorithms
- Study of Information Extraction and Optical Character Recognition

EDUCATION

University of Colorado, Denver **Aug 2019 – 18th Dec 2021**
Master of Science in Computer Science with Data Science and Biomedicine **(GPA: 3.7/4.0)** **EAD**: 14th Feb 2022
Courses: Big Data Mining, Big Data Science, Machine Learning, Longitudinal Data Analysis, Biostatistical Methods
Teaching Assistant: Deep Learning, Bioinformatics, Mergers and Acquisitions
Centurion University of Technology and Management, India **Aug 2014 -May 2018**
Bachelor of Technology in Computer Science and Engineering **(CGPA: 9.02/10.0)**

SKILLS

Programming Languages/ Database/Tools: Python, R, Spark, MySQL, Jupyter Notebook, Google Colab, PyCharm, RStudio, GitHub PyTorch, Keras, TensorFlow
Modelling: Predictive modeling, Non-Linear (LMM, GLM, MLR), RNN, Regression (Logistic, Linear), NLP, **Timeseries** (ARIMA, AR(1)), RFE, SVM, Tree (Decision, Random forest, KNN), PCA, ensemble (catBoost, LightGBM, XGBoost). **Clustering**: Bisecting K-means, DBSCAN, K-means, K-means++
Deep Learning & Pattern Recognition Algorithms: MAML, GAN, CNN, ANN, Neural Network-based models like Feed forward neural network, SNN, LSTM with dropout and L1, L2 regularization, Pattern mining(Apriori, Eclat, FP Growth, Naive Bayes)

EXPERIENCES

Research Experience, **University of Colorado, Denver** **Jan 2021 – Nov 2021**
Thesis topic: Artificial neural networks to formulate the underlying dynamics of neural activities of reaching kinematics.

- For faster computation, prototyped a novel low-power constrained Spiking neural network (SNN), and RNN models (LSTM and vanilla RNN) to conduct multiple sequential regression analysis for muscle activity prediction using time series data. Nengo-DL library was used to integrate the models in neuromorphic hardware, such as Nengo Loihi which used only 1.29 Joules.
- Visualized high dimensional data by using **Neural Tuning Curve**, jPCA, Cosine Curve fitting for pattern identification.

Data Analytics Intern, Seagate Technologies, CA **Jun 2020 – Aug 2020**

- Designed Machine Learning Pipeline in a distributed environment for a new product using PySpark with more than 20 million multivariate time series data.
- Assisted in the transition to faster processing for clustering billions of images with an unique Bisecting K means algorithm. Detected anomaly from Billions of Images using VGG models.
- During weekly meetings with team, learned how to write production-ready code and manage GitLab repositories.

Research Assistant (NLP & Risk Factor), University of Colorado Business school **Sep 2019 – May 2020**

- Risk Factor analysis on 5Million Aircraft Crashes dataset by Implementing Recursive Feature Elimination (RFE) for Feature Selection among more than 1000 features.
- Created a language translator for reviewing feedback in the English language. Classified Sentiments from Twitter conversations and Optimized machine learning models (KNN and Lasso Regression) using Spark ML-Lib to predict crashes.

Data Scientist Intern, Loop Reality Pvt Ltd, Hyderabad **Feb 2018 - Mar 2018**

- Collaborated with 5 different teams to collect real-time data. Classified humans cognitive skills using brain signals in more than 10lakhs sequential data and predicted person's next movement based on a 5-minute VR game, achieved 93.8% accuracy.
- Enabled HRs to analyze patterns of user's personality by building a Python module to create visualizations and statistical reports.

ACADEMIC PROJECTS

Recommendation engine | Python, LSH, NLTK

- Developed knowledge extraction pipeline to get topics from public reviews and created a model to recommend movies based on conceptualization for 2 million users using collaborative filtering method. Achieved 92% validation accuracy.

Sleep stage classification | Python, SVM, LSTM

- Classified sleep stages in 4 categories using machine learning Random forest, SVM on complex medical data, and facilitated early detection to prevent sleep apnea.
- Performed PCA with clustered people based on optimal sleeping hours using K means algorithm. Link Here.

Emotion Recognition from Speech and Video (MAML, GAN, CV2, Segment level feature extraction with Deep NN)

- Preprocessed 5K audio-visual samples using spectral centroid and wiener filtering method.
- DCGAN and MAML methods are used for identifying fake images based on face shape, nose alignment, eye blink and lip movement features. Link Here. Eliminated fake images and created GAN to eliminate Gender and Color Bias from samples and Pitch estimation with face shape detection and got 98% accuracy in 6 class emotion recognition. Link Here