KUNTAL PAL

Riverside, CA | 951-640-1691 | kuntal.pal@email.ucr.edu | LinkedIn | Github | Kaggle

SUMMARY: Machine Learning Researcher & physics Ph.D. candidate with strong math background. Experienced in developing deep learning models & focussed on contributing to current ML research and applying them for impactful solutions.

EDUCATION

- **UNIVERSITY OF CALIFORNIA, RIVERSIDE** **Sep 2017 - Dec 2023**
  Ph.D., Physics
  Cumulative GPA: 3.86/4.0
  Research topics: Data analysis and predictive modeling in high-energy physics using statistical and ML-based techniques. Neural architecture search using tensor completion.
  Relevant Coursework:
  - Data Mining Techniques
  - Computational Methods for biomolecular data
  - Probabilistic Models for AI
  - Optimization in Machine Learning
  - Introduction to Deep Learning
  - Advanced Computer Vision
- **INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, KOLKATA** **Aug 2012 - May 2017**
  BS-MS Dual Degree, Physics
  Cumulative GPA: 8.98/10.0
  Master's Thesis: Automated Quantum Field theory calculations using SciPy.

TECHNICAL SKILLS

- **Programming Languages**: Python (Advanced), SQL (Advanced), MATLAB (Intermediate)
- **ML/Data Analysis frameworks:** Keras, TensorFlow, Pandas, PyTorch, Scikit-learn, OpenCV
- **Other:** Github, Bash scripting

WORK EXPERIENCE

➔ Graduate Student Researcher | University of California, Riverside **Sep. 2017 - Present**

- Examine Tensor-Train decomposition with EM-algorithm for low-rank tensor completion in efficient hyperparameter search for neural networks.
- Developed search strategies and engineered features for applying the XGBoost decision tree algorithm. Achieved 80% accuracy in separating rare signal events from background noise. [2]
- Constrained model parameters using hypothesis testing and confidence interval estimation. [1][3]
- Demonstrated strong communication and teamwork skills through multiple research collaborations.

➔ Kaggle Competitor | Notebooks Expert **Jun. 2020 - Present**

- Forecasted sub-seasonal temperatures for multiple US locations using autoML library PyCaret for tuning CatBoost and TabNet regressors, achieving a top 30% ranking out of 709 teams. **[Github]**
- Deployed a large protein language model (ESM-2) and 3D CNN architecture (ThermoNet) for predicting enzyme variant stability based on melting temperature data, incorporating protein structure analysis with Rosetta and HTMD, placing in the top 40% among 2482 teams. **[Github]**

- Applied transfer learning with transformer-based models on student essays for assessing the language proficiency of 8-12th graders finishing with a mean column-wise RMSE score of 0.47.

➔ Team Lead- Data Science Challenge | Lawrence Livermore National Lab Jun. 2022 - Jul. 2022

- Mentored four undergraduate students in ML techniques, including data preprocessing, hyperparameter tuning, and Neural Networks, tailored to each team member's prior experience.
- Achieved the highest AUC score of 0.89 among six teams with a pretrained transformer model for classifying ligand molecules via SMILES strings.
- Trained 3D CNNs based on voxel representation of ligand structures, achieving an accuracy of 71%.

ML / DATA SCIENCE PROJECTS

➔ Web Scraping Job Descriptions | Python [Github] Mar. 2023

- Scraped LinkedIn to create a dataset of 2000 job profiles for data scientist and ML engineer positions.
- Analyzed the text data to identify in-demand skills and qualifications for employers in these domains.

➔ Forward-forward algorithm | TensorFlow | [Github] Jan. 2023

- Implemented the recently introduced forward-forward algorithm as a new neural network learning procedure. Obtained an accuracy of 94% for supervised tasks on the MNIST dataset.

➔ Gravitational wave time-series classification | TensorFlow | [Github] Sep. 2021 - Dec. 2021

- Analyzed 550k training and 200k test samples of simulated time-series data, using bandpass filtering and spectrogram imaging techniques like constant-Q transform and short-time Fourier transform.
- Trained accurate neural-network classifiers like ResNet, LSTM/GRU + CNN, and an Encoder, achieving the best AUC-ROC score of 0.86.

CERTIFICATIONS

- **IT Automation with Python Professional Certificate -** Google | Mar. 2023
- **TensorFlow Developer Certificate -** TensorFlow Certification Program | Feb. 2023
- **Deep Learning Specialization -** DeepLearning.AI | Sep. 2021

PUBLICATIONS

[1] Yoav Afik, Shaouly Bar-Shalom, **Kuntal Pal**, Amarjit Soni, Jose Wudka. Generic tests of CP-violation in high-pT multi lepton signals at the LHC and beyond. *Preprint arXiv: 2212.09433 (2022)*

[2] Subhaditya Bhattacharya, Sanjoy Biswas, **Kuntal Pal**, Jose Wudka. Associated production of Higgs and single top at the LHC in presence of the SMEFT operators. *Preprint arXiv: 2211.05450 (2022)*

[3] Yoav Afik, Shaouly Bar-Shalom, **Kuntal Pal**, Amarjit Soni, Jose Wudka. Multi-lepton probes of new physics and lepton-universality in top-quark interactions. *Nucl.Phys.B 980 (2022) 115849*

[4] ***K. Pal,*** L .V. Sales, and J. Wudka. Ultralight Thomas Fermi dark matter. *Phys. Rev. D.* ***100****,083007 (2019)*