Arunkumar Venkataramanan

#14/6, Pothaiya Naidu Street, Kaveripattinam, Krishnagiri, Tamil Nadu – 635112
arunkumar.ramanan@gmail.com | +91 90080 41191 | arunkumarramanan.github.io
Kaggle: arunkumarramanan | LinkedIn: arunkumarramanan | GitHub: ArunkumarRamanan

SUMMARY

- I'm a **Machine Learning Practitioner** (also known as **ML Software Engineer** and **Data Scientist**) have over **6 years** of **experience** in the fields of **Machine Learning**, **Deep Learning, Artificial Intelligence, Data Science**, and **Predictive Analytics** & **Modeling**.
- I founded and manage **DeepBrainz**, *an Independent **AI Initiative/Team Project*** as "***AI Lead***" who's been solving various business & real-world problems across industries using **Cutting-edge Technologies** and the **State of the art AI** in **Computer Vision** and **NLP**.

EXPERIENCE

Machine Learning Practitioner (ML Engineer & Data Scientist - ML) (July 2013 – Present)
DeepBrainz (An Independent AI Initiative/Team Project) Bengaluru, India

- **Advanced** the *SOTA* in **generative images creation** based on **MiFID** as **50.42142 CV** by training **GAN** models such as **BigGAN**, **StyleGAN**, **DCGAN**, **ACGAN**, **RaLS BigGAN** on Stanford Dogs Dataset along its Annotations. (**TF/PyTorch**)
- **Automated** multi-label **audio tagging** that scored **0.759 CV** as per **label-weighted label ranking AP** by building, training custom **CNN** models on Freesound & Yahoo FlickrCC datasets with augmentations to preprocessed Mel-Spectrogram. (**PyTorch**)
- **Predicted demand for an online advertisement** as **RMSE 0.216 CV** by designing, implementing **LGB**, **Ridge** models, **MLP**s, **RNN**s, on Avito's Ad datasets with pre-trained **word vectors, DenseNet, vgg19** for feature extraction. (**sklearn XGB TF Keras**)
- **Classified** and **Forecasted the future web traffic** for approximately 145,000 Wikipedia pages using **RNN seq2seq, LSTM/GRU** with **attention** by **Time Series Analysis** and **Inference** that evaluated on **SMAPE** as **test score 35.48065**. (**TF Keras**)
- **Recommended events to the users** on Event Recommendation Engine datasets from Kaggle as **0.712 CV** over **mean Average Precision mAP @200** by building, training **Logistic Regression**, **RF**, **k-Means** models for feature transformation. (**sklearn**)

PROJECTS

Large Scale Computer Vision and NLP with Deep Learning (Jan 2019 – Oct 2019)

- **Diagnosed diabetic retinopathy** as **0.936129 test score Quadratic Weighted Kappa** by fine-tuning, augmenting, ensembling **EfficientNet**, **SE-ResNet**, **SE-ResNeXt/50/101**, **SE-Net154** on large set of retina images. (**PyTorch**)
- **Detected Objects Automatically** as **mean AP 0.59 CV** by using pre-trained **ResNet/EfficientNet**, **NAS-FPN** & **RetinaNet** models on large-scale Open Images v5 with **TFX** pipeline for training, testing, deploying models on **TPU**s. (**TF**)
- **Recognized Images** as **Global AP @ k 0.2515 CV** by fine-tuning and augmenting pre-trained **SE-ResNeXt50** & **Faster R-CNN** models on large-scale Google Landmark & CVDF datasets with Feature Engineering. (**PyTorch**)
- **Detected toxicity** and **minimized Bias in Toxicity Classification** as **0.947CV** of custom **AUC** metric by developing custom **LSTM**s with pre-trained **BERT, XLNet, GPT2** models on Jigsaw's dataset and preprocessed embeddings. (**PyTorch**)
- **Improved gender-fairness** for **gender bias pronoun resolution** by designing **BERT** based models on GAP Coreference Dataset with preprocessed embeddings by augmentation measured as **multi-class log loss** of **0.1916 test score**. (**TF Keras**)

EDUCATION

Independent Coursework - MOOCs (July 2013 - Present)

Coursera	*Udacity*	*edX & Others*	*Stanford University*	*MIT*	*Others*
Stanford University: • **Algorithms** • **Machine Learning** *deeplearning.ai:* • **Deep Learning** • **TF in Practice** • *UToronto* **Learn to Program** *Series* • *UCSD* **Big Data**	• **Machine Learning Engineer** NanoDegree (*ND)* • **Data Scientist** *ND* • **Data Engineer** *ND* • **Design** of **computer programs,** *Norvig* • **Software Testing** & **Debugging**	• *UCSDx* **Data Science** • *IBM* **Python DS** • *ColumbiaX* **Business Analytics** *CS50x HarvardX:* • **Computer Science** • **Web Programming** • CMU **Distributed Systems**	• **Machine Learning** • **Mining** Massive **Datasets** • **CNN**s for **Visual Recognition** • **NLP** with **Deep Learning** • **Information Retrieval**	• **Design** and **Analysis** of **Algorithms** • **Computational thinking** & **Data Science** • *MITx*: **Stats & Data Science** • **Deep Learning**	• *Google AI* - *Education: Learn with Google AI* • *Datacamp* Track: **Data Scientist** with **Python** & **R** • *Kaggle* Courses: Faster **Data science** education

B.Tech. Information Technology (Aug 2009 - July 2013)
University College of Engineering Tindivanam Anna University, Chennai, India

SKILLS

- **Programming**: **Expert** in **Python, Java, SQL;** Have **Experience** with **C++, C, R, MATLAB, C#, BASH/Shell script, JavaScript**
- **Libraries & Frameworks**: **Expert** in **TensorFlow** & Ecosystem **Keras PyTorch scikit-learn XGBoost LightGBM pandas NumPy SciPy IPython Matplotlib**; **Experienced** in MXNet MapReduce Spark Hadoop Ecosystem (HBase Hive Pig Flume Sqoop Hume) OpenCV NLTK spaCy genism Caffe2 fastai Plotly ONNX; Familiar with Beam Kafka Storm Knime Weka RapidMiner Neo4j
- **Tools**: Colab Anaconda PyCharm VS Code Excel Linux Git MySQL Tableau MongoDB Cassandra RStudio GCP AWS Azure

AWARDS & HONORS

- Top **0.03%** ranked **38th/102k** & **49th** ***Kernels Master*** and Top **0.2%** ranked **235th/122k** ***Competitions Expert*** on Kaggle
- Won *Medals* on Kaggle: **62nd/468** in *Google AI* ***Inclusive Images Challenge,*** **74th/201** in *Google AI* ***Open Images 2019 – Visual Relationship,*** **90th/521** in ***iMet Collection 2019 FGVC6***