

HARIKRISHNAN RAGHUKUMAR

College Station, TX 77840 | ☎ (979) 344 8480

✉ harikrishnankr16@gmail.com | 🌐 kr-hari.github.io | 🔗 linkedin.com/in/kr-hari | 🐙 github.com/kr-hari

EDUCATION

Texas A&M University, College Station, TX

Aug 2022 – Dec 2023

MS in Data Science (Computer Science)

CGPA – 4.0 / 4.0

- *Relevant courses: Information Storage and Retrieval, Machine Learning, Data Mining, Computational Tools for Data Science*
- *Texas A&M Institute of Data Science scholarship (\$1,000 with in-state tuition waiver)*

Indian Institute of Technology Madras (IIT- M), India

Aug 2015 – Jul 2020

Dual Degree - BTech. in Engineering Design and MTech. in Automotive Engineering

CGPA – 8.92 / 10

- *Relevant courses: Natural Language Processing, Machine Learning, Deep Learning, Reinforcement Learning, Time Series*

EXPERIENCE

Texas A&M University (TAMU)

College Station, Texas

Graduate Research Assistant

Aug 2022 – Present

- Built an end-to-end web-based recommender system application to recommend TAMU scholars for federal and state funding opportunities in Python in less than 3 months
- Compiled dataset consisting of 60K+ government grants and 6K+ scholars through web scraping [BeautifulSoup]
- Utilized BERT to extract keywords and developed content-based recommender system
- Automated dataset updation [Cronjob] and deployed it in production; containerized the virtual server using Docker
- Modularized the existing codebase, documented [Sphinx] and hosted it publicly through Readthedocs
- Optimized the code and enabled parallelization [Multiprocessing], cutting down execution time by 80%

American Express

Gurgaon, India

Risk Analyst

Aug 2020 – Jul 2022

- Utilized **SQL** and **NoSQL** databases to analyze 1B+ rows of customer data, discovered **trends** for 15+ financial products and delivered presentations to senior leaders across business units through impactful **visualizations** [Tableau]
- Improvised existing customer entity resolution algorithm and **reduced underlinkage** by 8%
- Built a batch-process data pipeline to **merge, deduplicate, clean** and **analyze** dataset of 200M+ customers
- Collected **data lineage** information for 10+ consumer products and documented the same in Confluence to assist **data governance** efforts
- Took initiative, proposed and implemented a time-critical algorithm to derive time zone of each customer which currently serves 80 MM+ customers; awarded '**Critical thinker – 2021**' for structured problem solving
- Collaborated with business and IT teams, and strategized upgradation of MATCHBOX, increasing user base by 300%

PhotoGAUGE

Chennai, India

Machine Learning Engineer (Intern)

Jan 2019 – Jun 2019

- Developed a program to locate irregularities on the surface of freshly manufactured equipment using **deep learning** [PyTorch] and **computer vision** [OpenCV]
- Conducted comprehensive **research** and implemented several state-of-the-art deep learning models to investigate effectiveness of advanced machine learning paradigms like continual learning, lifelong learning and multitask learning to build a one-for-all network
- Achieved **93% accuracy** in detecting irregularities from equipment images by implementing **ResNet-50** model
- **Reduced training data** demand by 18% through active learning; built a **Graphical User Interface** to aid data labeling

KEY PROJECTS

Hot Topics in Machine Learning (Natural Language Processing)

Dec 2022

- Performed topic modeling and topic classification to analyze trends in Machine Learning using Python [Jupyter notebook]
- Extracted, cleaned and compiled NeurIPS research paper dataset; parallelized code to cut down execution time by 85%
- Implemented LDA algorithm with a coherence of 0.69 and created visualization [pyLDAvis] to model research topics
- Utilized Gensim Zero-shot classifier to track evolution of top 6 Machine Learning topics through the years

IT Salary Survey for Europe (EDA, Visualization)

Oct 2022

- Performed **data wrangling** (merging, cleaning, manipulation and transformation) to prepare dataset from 3 sources
- Utilized Python and Jupyter notebook to conduct a detailed **exploratory data analysis** (including univariate, bivariate and multivariate analyses), **feature engineering** and **hypothesis testing** for analyzing job trends in the European IT industry
- Visualized feature distributions and **correlations** using scatterplot, boxplot, pairplot, QQplot, catplot and heatmaps [Tableau, Seaborn]

Cancer Diagnosis Web Application (Deep Learning, Computer Vision)

Mar 2020

- Built end-to-end prototype of cancer detection tool based on **deep learning** [PyTorch] that can identify potential malignant cells in Whole Slide Images and incorporated features like cancer cell segmentation and report generation
- Trained a U-NET **image segmentation** model on DigestPath 2019 dataset with an accuracy (IOU) score of **0.64**
- Created a web application [Flask, HTML, CSS] and containerized it using Docker

SKILLS

Programming Languages: Python (PyTorch, TensorFlow, NumPy, Pandas, Scikit-learn, SciPy, Matplotlib, CUDA), R, C++, C, MATLAB, SQL

Machine Learning Techniques: Regression, Decision Tree, SVM, GMM, Hierarchical clustering, KNN, Decision Tree, Random Forest, XGBoost, AdaBoost, CNN, RCNN, GAN; Data science pipeline (cleaning, visualization, modeling, interpretation), Statistical testing

Tools: Google colab, Jupyter Notebook, Dask, Tableau, Power BI, AWS, Rally, Git, Docker, IPython, MS Office

ACHIEVEMENTS

- 6th Rank, Gartner All India HackElite Competition, Gurgaon
- National Finalist, Philips Code to Care Challenge, Bengaluru

Nov 2019

Sep 2019