Ravi Kiran Selvam

Los Angeles, CA 90007 | rselvam@usc.edu | www.sravikiran.com | +1 2138006664

EDUCATION

M.S in Applied Data Science
University of Southern California — CGPA - 4.0/4.0 — **May 2021***

B.E. in Computer Science *(Among top 5% out of 180 students)*
Anna University - CEG Campus — CGPA - 9.47/10 — **April 2019**

RESEARCH AREAS - Machine Learning, Deep Learning, Natural Language Processing, Semantic Web

SKILLS

Languages and Technologies: Python, SQL, C++ (Proficient); C, JAVA, Bash, pySpark, MapReduce (Intermediate);
Data Analysis: Exploratory Data Analysis, Time Series Analysis, Model Evaluation
Data Management and Engineering: MongoDB, Snowflake, Azure Data lake, MySQL
ML Frameworks: Tensorflow, Keras, scikit-learn
Data Visualization: Plotly, matplotlib, seaborn
Others: Git, Markdown, Flask, Bootstrap, JavaScript, Android, Software testing, OO design skills

EXPERIENCE

***Graduate Researcher*, Information Sciences Institute, USC** — **October 2019 - Present**

- Working at the **Center on Knowledge Graphs Lab** under the mentorship of Prof. Mayank Kejriwal.
- Our project is focussed on creating better and improved neural representations of text data by incorporating domain knowledge with word embeddings to solve some of the most interesting NLP problems in the e-commerce domain funded by Yahoo Research! Currently, we are working to solve “large scale product categorization” and “product matching” with these new representations.

***Data Science Intern,* Motorq** — **December 2018 - June 2019**

- Set up the **initial Data Science workflow** and infrastructure for building ML models
- Analyzed **large-scale connected car data** from IoT devices and solved problems such as refueling event detection, idling time detecting, trip completion event detection
- Designed and Implemented **various engine hour metrics, meta-metrics** for different parameters of car data and analyzed the trends across time
- Build the **battery voltage failure prediction model** based on the number of parameters from car data
- Tech stack: **python, numpy, pandas, matplotlib, plotly, scikit-learn, tensorflow, snowflake, Azure Data lake, Azure VM**

***Undergraduate Researcher,* College of Engineering Guindy, Anna University** — **August 2018 - November 2018**

- Worked under the mentorship of Prof. G.S. Mahalakshmi.
- Our project aimed at proposing a state-of-the-art model to identify Named Entities in the Indian Culinary Science Text dataset using Deep Learning sequence models.
- Our paper titled **"Exploiting Bi-LSTMs for Named Entity Recognition in Indian Culinary Science"** has been accepted in the 5th International Conference on Next Generation Computing Technologies (NGCT 2019)

***Machine Learning Intern,* Kenome Technologies** — **May 2018 - June 2018**

- Built a deep learning model to perform **sequence tagging** for colors, materials, and patterns in text documents
- Built a method for data-annotation by reducing the time complexity of string matching from a naive algorithm using a modified version of the Trie data structure. Observed a maximum F1 score of 0.94 for tagging colors and materials in the testing data set
- Built a dashboard to **visualize the crypto-currency prediction model**
- Tech stack: **python, TensorFlow, Keras, AWS EC2, plotly, d3**

***Undergraduate Researcher,* College of Engineering Guindy, Anna University** — **August 2017 - December 2017**

- Worked under the mentorship of Prof. G.S. Mahalakshmi.
- Implemented a modified version of the Power Graph data structure to represent the relationships of author and co-author in the citation network data set in a memory-efficient manner.
- Deduced algorithms to perform queries like finding the bonding value between authors (to find the type of citations between papers) and a similarity index between research papers.

***Software Development Engineer Intern*, Amazon** **May 2017- July 2017**

- Developed prototype features for **Amazon-Fire TV Stick** to integrate marketing notifications using Amazon's internal library and to integrate IMDB ratings with Amazon Prime videos
- Tech: **Java, XML, Software testing, Software design principles**

OPEN SOURCE CONTRIBUTIONS

***Google Summer of Code 2018 Student Developer*, CERN** **April 2018 - August 2018**

- Provided support for advanced deep learning optimizers in the open-sourced **ROOT-TMVA, a data analysis software framework** by CERN
- Implemented deep learning optimization algorithms (SGD, RMSProp, Adam, Adagrad, etc.) in CPU & GPU architectures by exploiting the parallel programming capabilities; my code has been successfully integrated into the new production release of ROOT version 6.16
- Tech stack: **C++, Blas, CUDA, CuBlas**

PROJECTS

Sign Language Translator from Video to Speech (ISL) **December 2018 - March 2019**

- Developed a deep learning model for classifying sign language images to corresponding sign words by transfer learning of the Inception V3 architecture and extended it to videos for generating simple English sentences
- Created the sign language data set for 30 sign words and 25 alphabets which could be classified using just one image frame and trained our model using it
- Got an accuracy of 62.5% on the test set under normal lighting conditions
- Technologies: **python, tensorflow, keras, plotly, opencv, scikit-learn**

Customized Adversarial Image Generator **August 2018 - October 2018**

- Implemented a variation of Fast Gradient Sign Method (FGSM) algorithm to perturb the input image to fool the pre-trained Deep Convolutional Neural Network model so that it will misclassify it to the desired target class; Produced perturbed images that are indistinguishable to the human eye.
- Demonstrated the drawbacks of CNN architecture and suggested ways to overcome it.
- Technologies: **python, numpy, pandas, matplotlib**

Credit Card Fraud Detection **February 2018 - March 2018**

- Developed an ML model using multivariate Gaussian distribution to detect fraudulent credit card transactions
- Trained the model using standard credit card dataset available on Kaggle; Achieved accuracy of 95% on new test data
- Technologies: **python, numpy, pandas, scikit-learn, matplotlib**

CERTIFICATIONS

- Deep Learning Specialization (series of 5 courses) by Deeplearning.ai, Coursera, **March 2018**
- Machine Learning by Stanford University, Coursera, **December 2017**
- Codechef Certified Data Structures and Algorithms Program (CCDSAP) - Advanced Level, CodeChef, **November 2017**

AWARDS

- Ranked 35th among 250 teams (Amritapuri Regionals) and 30th among 120 teams (Chennai Regionals) **across India** in ACM International Collegiate Programming Contest, **December 2017**
- Won 25 coding competitions in 12 inter-college tech fests (by securing 1st among ~400 participants), **October 2016 - March 2019**

EXTRA-CURRICULAR ACTIVITIES

- **Founder, CEG Codechef Campus Chapter** - Delivered lectures on competitive programming to many college students and trained them to participate in the ACM-ICPC, **September 2018 - March 2019**
- **Problem Setter, Abacus'17 & Abacus'18**, departmental inter-collegiate national-level technical symposium- Organized 5 intercollegiate onsite & online programming contests (HackerRank, CodeChef), Anna University, **March 2017 & March 2018**
- **Authored 2 blogs** for beginners on Algorithms and Data Structures with ~25,000 page views, (Link1, Link 2), **March 2016 - May 2017**