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 - 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: pytorch, 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
- My primary research was focused on **learning representations for sparse taxonomies**. Our project aims at reconstructing a taxonomy given a set of concepts in a particular domain using a background resource and exploiting the zero-shot learning paradigms. We are planning to submit the paper, **"Zero-Shot Taxonomy Induction Using Representation Learning: An Empirical Study"** in a journal
- Studied the data quality issues in the Schema.org Product-specific data and formulated a set of best practices for consuming this dataset for downstream applications in the e-commerce domain. My first-authored paper ***"On using Product-Specific Schema.org from Web Data Commons: An Empirical Set of Best Practices"*** is accepted in the E-commerce and Knowledge Graph workshop at KDD 2020

***Data Scientist Intern,* Amazon** — **May 2020 - August 2020**

- Built a deep learning model to automatically **identify the reason categories for low-star customer reviews** for Amazon gift card products. The reason categories represent the potential problems faced by the customers. Since there was a lack of labeled data, we utilized learning approaches that work well in a low data regime. We performed a variety of text data augmentation techniques to increase the labeled dataset size. We achieved a **test accuracy of 74% for the aforementioned few-shot text classification problem**. We implemented the uncertainty estimation framework for estimating the uncertainty of the text classification model and increased the overall accuracy to **above 90%** by giving a small portion of the reviews with uncertain predictions for manual labeling
- Build an **end-to-end pipeline** starting from the ETL operations to query the data and make the inference on a weekly basis and send automated emails using native AWS Services **(SageMaker, Lambda, and Step functions)**; Also, built a dashboard for visualizing the trends in the predictions using **Amazon QuickSight**
- **Helped the Gift cards business team** by reducing the manual tagging time of around 40 hours every month to less than 1-2 hours per month **(20x faster)** by giving a small portion of reviews with uncertain predictions for manual labeling and trusting the model's predictions for the remaining reviews
- Tech stack: **python, numpy, pandas, pytorch, AWS, scikit-learn, SageMaker**

***Data Scientist 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
- My first-authored 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**

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**

PUBLICATIONS:

- **Ravi Kiran Selvam** and Mayank Kejriwal. 2020. ***On using Product-Specific Schema.org from Web Data Commons: An Empirical Set of Best Practices***. In Proceedings of KDD '20: Workshop on Knowledge Graphs and E-Commerce (KDD '20), San Diego, CA, USA.
- Mahalakshmi, G.S.* and Sreedhar, Makesh Narsimhan* and **Selvam, Ravi Kiran*** and Sendhilkumar, S, ***Exploiting Bi-LSTMs for Named Entity Recognition in Indian Culinary Science*** (February 27, 2020). 5th International Conference on Next Generation Computing Technologies (NGCT-2019), Available at SSRN: https://ssrn.com/abstract=3545088.

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**