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, Knowledge Graphs, Information Extraction

SKILLS

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

PUBLICATIONS:

1. **Ravi Kiran Selvam** and Mayank Kejriwal. *On using Product-Specific Schema.org from Web Data Commons: An Empirical Set of Best Practices*. **Knowledge Graphs and Ecommerce @ KDD'20.**
2. Wangchunshu Zhou*, Dong-Ho Lee*, **Ravi Kiran Selvam**, Seyeon Lee & Xiang Ren. *Pre-training Text-to-Text Transformers for Concept-centric Common Sense*. **ICLR'21**
3. Wangchunshu Zhou*, Dong-Ho Lee*, **Ravi Kiran Selvam**, Seyeon Lee, Bill Yuchen Lin, & Xiang Ren. *Pre-training Text-to-Text Transformers to Write and Reason with Concepts*. **Self-Supervised Learning @ NeurIPS'20**
4. Mayank Kejriwal, **Ravi Kiran Selvam**, Chien-Chun Ni, Nicolas Torzec. *Locally Constructing Product Taxonomies from Scratch Using Representation Learning*. **ASONAM'20**
5. Mayank Kejriwal, **Ravi Kiran Selvam**, Chien-Chun Ni, Nicolas Torzec. *Empirical Best Practices On Using Product-Specific Schema.org*. **IAAI'21**
6. 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*. 5th International Conference on Next Generation Computing Technologies (**NGCT-2019**)

EXPERIENCE

***Graduate Researcher*, Intelligence and Knowledge Discovery Lab, USC** **September 2020 - Present**

- Working under the mentorship of **Prof. Xiang Ren** in the USC Intelligence and Knowledge Discovery Lab
- My primary project is focused on lower resource NER using weakly supervised learning and developing methods for improving the interpretability in Neural Sequence Models. Also collaborating as a part of the MACROSCORE project (https://usc-isi-i2.github.io/macro-score/) at USC Information Sciences Institute.
- Contributed to another project that is related to generative commonsense reasoning in language models where we aim to inject commonsense knowledge into pre-trained language models. Refer to [2] and [3] in the Publications section.

***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**

***Graduate Researcher*, Information Sciences Institute, USC** **October 2019 - May 2020**

- Worked at the **Center on Knowledge Graphs Lab** under the mentorship of **Prof. Mayank Kejriwal** and collaborated with the **Yahoo Product Knowledge Graph team.**
- 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. Refer to [4] in the Publications section.
- 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. Refer to [1] and [5] in the Publications section.

***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**

***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**

***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
- Created a prototype to integrate IMDB ratings with Amazon Prime videos, where the user's rating for a particular video will get accumulated to the IMDB rating
- 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**

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**
- **Authored 2 blogs** for beginners on Algorithms and Data Structures with ~25,000 page views, (Link1, Link 2), **March 2016 - May 2017**