I am a Master of Computer Science student with a solid mathematics foundation with hands-on experience including in-depth knowledge of C, C++, Java, R. I previously graduated from Rensselaer Polytechnic Institute with a dual major in both math and computer science. I am looking for a summer internship where I can leverage my background in computer programming, mathematics, data analysis and fundamental research. . I am particularly interested in Machine Learning. I am really interested in learning how to design algorithms to fit different types of Datasets. I also want to learn how to choose which approach from many possible models. I am also interested in learning how deductive statistical based approaches can be fused with the current inductive based Machine Learning approaches. I find maximum likelihood and maximum entropy models to be very interesting. I want to know if there is some way to use supervised Machine Learning to uncover latent variables. I hope to expand my knowledge. Moreover I hope to be able to work on some interesting projects. In the past, I can implemented a calendar that can run on distributed systems using the Paxos algorithm. Moreover, I used a message passing based Graph Neural Network to act as a recommendation system. In my previous internship over the summer, I worked at a 3D Printing company. The Company’s main business was to manufacture 3D printers. I worked on the software that was packaged with the printers. I was able to use linear algebra and divide and conquer based algorithm on high dimensional spatial trees to dramatically improve performance. In another internship, I designed a machine learning algorithm that could predict which people were most likely to qualify for disability benefits.

**Karan Sarkar**

**978.908.8841.** [**karan@sarkarfamily.org**](mailto:karan@sarkarfamily.org)

**7 Country Rd, Westford MA 01886**

**Summary**

Talented technical MSCS student with a solid mathematics foundation with hands-on experience including in-depth knowledge of C, C++, Java, R. Looking for a summer internship where I can leverage my background in computer programming, mathematics, data analysis and fundamental research. While at high school, I was selected to a highly competitive yearlong research program at MIT. My research paper was published in the “International Journal of Game Theory”.

**Languages:** C, C++, Java, Python, R, MATLAB, SQL, Powershell

**Libraries:** STL, Eigen, numpy, pandas, nltk, keras, matplotlib, tbb, bnlearn, ggplot2,

**Tools:** Eclipse, RStudio, Visual Studio, SQL Developer, MATLAB**,** GitHub

**Courses:** Data Structures, Algorithms, Data Analytics, Natural Language Processing, Data Mining, Machine Learning

Intro to AI, Numerical Computing, Advanced Calculus, Econometric Methods for Big Data

**Projects**

**Formlabs (Software Engineer Intern) May 2020 – August 2020**

* **Used spatial tree to compute support failures**
  + Worked on algorithmic enhancements to the Preform 3D-Printing desktop software.
  + Computing support locations is one of the most expensive compuations
  + Used a modified Cholesky Decomposition combined with Branch-and-Bound on 6D-RTree
  + **Sped up support generation by 40%.**

**Healthcare Financial Inc. (Data Scientist)** **August 2018 – October 2019**

* **Developed rule mining algorithm to find high disability clusters**
  + Generated candidate rules use pairwise support and lift
  + Constructed custom rule linked prefix-tree data structure to allow fast evaluation of rules
  + Used C++ STL containers such as unordered-map to obtain high performance
  + **Use of data structures reduced run-time by 75% from 4 hours to 1 hour**
* **Constructed intuitive conversion from diagnosis descriptions to social security determinants**
  + Used PubMed Word Embeddings to convert diagnosis descriptions to quantitative domain
  + Generated intuitive categories from spherical K-means
  + **Resulted in 0.25 Purity**

**Rensselaer Polytechnic Institute, Troy NY (Student) August 2016 – Present**

* **Constructed Graph Neural Network for Link Prediction**
  + Used Message-Passing Gated Recurrent Unit based GNN to predict existence of edges in Large Scale Networks
  + Out-Performed traditional heuristics like Katz and Adar-Adamic with **96% AUC**
* **Created C++ STL Style Graph Library**
  + Used C++ template arguments to allow choice in internal representation i.e. vector vs hash-table vs linked-list
  + Created de-linked graph algorithm templates that allow dependency injection. Can be used to yield Djikstra’s vs DFS vs BFS with just change of parameter
* **Implemented Paxos Algorithm using Python TCP Programming**
  + Maintains distributed log across many machines
  + Used appropriate data structures such as hash-table to maintain efficiency
  + Made use of Python threading library to maintain safety
  + Constructed flexible Event class to maintain extensibility

**Key Milestones**

**Dual Major in Math and Computer Science at RPI 2020**

**Program for Research in Mathematics, Engineering, and Science, Massachusetts Institute of Technology 2015-2016**

Wrote a research paper that was **published at a MIT conference and published in the International Journal of Game Theory.** The paper is called “On P-Positions in Modular Extensions to Nim” [http://link.springer.com/article/10.1007/s00182-016-0545-7](http://www.springer.com/-/0/AVYUtZb3RWOIidwVFEop)