

Hirak Sarkar

Stony Brook – NY 11790

☎ 631-520-8131 • ✉ hsarkar@cs.stonybrook.edu • <http://www.hiraksarkar.com>

OBJECTIVE

I am interested in applying machine learning techniques such as Bayesian inference and deep learning to analyze and extract information from large corpus in the field of genomics, social and computer networks.

EDUCATION

Stony Brook University (SBU) <i>Ph.D in Computer Science (3.99/4)</i>	Stony Brook, NY <i>2014-2019(exp)</i>
Indian Statistical Institute <i>M.Tech in Computer Science (1st class Hons.)</i>	Calcutta, India <i>2011-2013</i>
West Bengal University of Technology (Kalyani Govt. Engg. College) <i>B.Tech in Computer Science (8.88/10)</i>	Calcutta, India <i>2007-2011</i>

EXPERIENCE

COMBINE-Lab (Stony Brook University) <i>Research Assistant (https://github.com/COMBINE-lab)</i>	<i>Jan 2015-Present</i>
<ul style="list-style-type: none">○ Application of machine learning methods for publicly available massive genomic databases. (Python, sklearn)<ul style="list-style-type: none">- Development of SeaDragon (under development) involved application of different dimensionality reduction techniques, and gradient boosted trees for detection of population type from GEUVADIS dataset.○ Development of graph based k-mer mapper, Pufferfish (C++)<ul style="list-style-type: none">- Genome sequences (string in the order of gigabytes) are difficult to index and search in bounded memory, used a minimum perfect hash based, rank-select algorithm to implement a fast query scheme for nucleotide sequences.○ Developed an intermediate solution for accurate mapping of read sequences. (C++)<ul style="list-style-type: none">- Alignments involves dynamic programming and therefore are costly. Mapping of reads are first yet not accurate, to carry best of the both worlds we developed a selective-alignment based algorithm.○ Developed alignment free methods for sequence reads. (C++)<ul style="list-style-type: none">- We developed <i>RapMap</i>, an ultra fast mapper which builds a suffix array over transcriptomic sequences.○ Graph based clustering for novel organisms. (C++, Python)○ State of the art compression tool for RNA-seq reads. (C++)	
Collaboration Siepel-Lab (Cold Spring Harbor Lab)	<i>June 2016-Aug 2016</i>
<ul style="list-style-type: none">○ Developed probabilistic graphical model for inferring transcription rate from multi-assay dataset. (Python)	
Collaboration with Wings Lab (Stony Brook University)	<i>Aug 2017-Present</i>
<ul style="list-style-type: none">○ Data driven inference models for spectrum sensing. (Python)	

PUBLICATION

- Journals
 - Bioinformatics'16,17 (*impact factor: 7*), Journal of Theoretical Computer Science'15 (*impact factor: 0.8*)
- Conferences
 - ISMB'16, RECOMB-seq'16, WALCOM'13
- Posters
 - BioData'16, WABI'17

AWARDS

- Special CS Chair Fellowship (\$10K) from *SBU*, Post-Graduate Scholarship from *Govt. Of India*

SKILLS

- Programming: *C++, Python, R*
- Data analysis: *Jupyter, Pandas*
- Machine Learning Tools: *sklearn, tensorflow*