

Aryaman Bhagat

+91 (807) 604 2213
f20180248@pilani.bits-pilani.ac.in
https://www.linkedin.com/in/aryamanbhagat/

Education

2018–2022 **B.E. Computer Science**, GPA: 8.17/10 , BITS Pilani, Pilani, India.

Research Interests

- Deep Learning, Distributed Computing, High Performance Computing

Honors & Awards

- Merit Scholar at BITS Pilani for three semesters for being within top 3% of the batch by CGPA
- Placed 21st in India and 377th worldwide out of 10724 participants in Google Hashcode 2020

Experiences

Jan 2022 to Present **Undergraduate Researcher**, DREAM Lab, IISc.

- Project Description** Working towards two publications targeting top conferences
 - "Optimizing the Interval-Centric Distributed Computing Model for Temporal Graph Algorithms" targeting EuroSys
 - "A Distributed Streaming Graph Partitioner To preserve Community Structure" targeting ICDCS
- Technologies used: Hadoop, Matplotlib, Apache Thrift, Apache Giraph, Apache Kafka

May 2021 to Jun 2021 **Software Engineering Intern**, Walmart Global Tech, Bangalore.

- Project Description** Built Phronesis, an NLP based pipeline to suggest panelists for interviewing candidates based on their resume.
 - Used state of the art NLP models, BERT, spaCy, LSTM for Named Entity Recognition to extract various candidate skills, contact details and qualifications.
 - Increased efficiency of the NER task using IOU tagging for parts of speech in the resume.
 - Built an ensemble model to increase recall after noticing that certain models had high precision on certain entities but low recall overall.
 - Converted the problem of suggesting panelists to an information retrieval problem.
 - Created a TF-IDF scoring system to rank panelists based on fuzzy matches with candidate skills.
 - Used Cosine Similarity threshold matching to match semantically similar skills.
 - Obtained a precision@5 of 90% on a relatively small data-set of 200 resumes.
- Technologies Used: NLTK, spaCy, scikit-learn, python, azure.

May 2020 to July 2020 **Research Intern**, CDAC Pune.

- Project Description**
 - Built containerized versions of the HPL and STREAM benchmarks, allowing quick and uniform bench-marking of any machine.
 - Obtained 80% efficiency compared to bare metal benchmarks, trading benchmark performance for portability.
 - Conducted extensive literature review in the field of system benchmarking
- Technologies Used: CUDA, MPI, C++, Docker, Singularity, openMP

Jan 2021 to May 2021 **Research Assistant**, BITS Pilani, ADAPT Lab.

- **Project Description** Accelerating DBSCAN on GPU nodes.
 - Wrote a GPU based implementation of the DBSCAN algorithm
 - Obtained similar performance as the state of the art G-DBSCAN implementation
 - Conducted extensive literature review in the field of GPU accelerated clusters and data clustering algorithms
 - Technologies Used: CUDA, MPI, C++

Aug 2018 to Dec 2020 **Simulation Engineer**, BITS Pilani, Team AcYut.

- **Project Description**
 - Ported Legacy Code to ROS for our mid-sized humanoid robot AcYut-VII.
 - Wrote an open-loop walk for our mid-sized humanoid robot AcYut-VII.
 - Used Co-variance Matrix Adaptive Evolutionary Strategy to tune parameters for a simulated humanoid agent.
 - Created Low-Level primitives(kick, run, walk) for a simulated humanoid agent.
 - Created high-level behaviour algorithms and developed game-play strategy for a team of simulated humanoid soccer playing agents
 - Technologies Used: ROS, Gazebo, C++

Course Projects

- **Foundations of Data Science**

Active Learning in Clustering Problems

- Wrote Clustering Algorithms for classifying fault types in steel plates
- Converted Clustering Problem to an Active Learning problem.
- Implemented Active Learning techniques like Query By Committee, Uncertainty Sampling, Version Space.
- Implemented a Novel Cluster Labelling approach to the Active Learning Problem.
- Compared all Approaches to a random sampling baseline.

- **Machine Learning**

Data Representations for Machine Learning

- Wrote various representations for visual, audio and textual data.
- Wrote simple kernel transformations for visual data.
- Compared various super-pixel algorithms using a simple neural network classifier on image-net data
- Used a Mel-Frequency Cepstrum transformation to identify voice in audio.
- Used Word-Vector embedding to train a simple spam classifier.

- **Machine Learning**

Time Series Temperature Forecasting using LSTMs

- Used LSTMs to forecast temperature on New Delhi weather data.
- Improved Performance by augmenting data with simple tags like season and time of day.

- **Principles Of Programming Languages**

Compiler in C:

- Built a rudimentary compiler to recognize a simple language as a part of CS F301

- **Information Retrieval**

Query Based IR System

- Implemented a simple IR system using TF-IDF scoring and pattern matching.
- Improved the performance using word proximity and synonym matching
- Implemented Cosine Similarity Threshold on word vectors to improve semantic matches

- **Applied Statistical Methods**

Stock Market Prediction using Time Series Models

- Used state of the art Time Series Models Facebook Prophet, Amazon Sagemaker DeepAR for stock market prediction.
- Compared to traditional approaches like ARIMA and SARIMA
- observed the effects of stationarity on time-series data prediction algorithms.
- recognized weekly trends in the behaviour of the stock price.

Teaching Experiences

Jan 2020 to **Teaching Assistant**, C Programming.

- May 2020
- Conducted weekly labs for a batch of 40 students
 - Administered weekly quizzes and other major evaluative components

Aug 2020 to **Teaching Assistant**, Object Oriented Programming.

- Dec 2020
- Conducted weekly labs for a batch of 40 students
 - Wrote and administered weekly quizzes and other major evaluative sections

Aug 2021 to **Teaching Assistant**, Object Oriented Programming.

- Present
- Conducted lab sessions for a batch of 40 students.
 - Wrote and Administered weekly quizzes, online tests and other major evaluative components.
 - Made JAVA project assignments for students
 - Mentored a group of 20 students towards 3 different projects

Coursework

- NVIDIA NSM DLI(set of 10 MOOCs towards accelerating Deep Learning using NVIDIA GPUs)
- Deep Learning Specialization(MOOC)
- Reinforcement Learning(Lecture Series)
- Cloud Computing
- Game Theory(MOOC)
- Machine Learning
- Information Retrieval
- Data Storage and Networks
- Applied Stochastic Processes
- Applied Statistical Methods
- Parallel Computing
- Foundations of Data Science

Skills

- **Languages**

Python, C/C++, JAVA, MATLAB

- **Technologies**

Linux/Unix, Tensorflow, MPI, CUDA, Keras, hadoop, STL, YARN, Giraph, Thrift

Languages

- **English:** Native
- **Hindi:** Native