Vignesh Nanda Kumar

Contact Information
Email: vigneshnandakumar1997@gmail.com Phone: +91-9999491454
LinkedIn: Vignesh Nanda Kumar

Education

Birla Institute of Technology and Science Pilani, Pilani Campus, India *Aug '15 - Jun '19*
Bachelor of Engineering (Honours), Computer Science
CGPA: **9.98/10** — **Silver Medalist of 2019 batch**

Professional Experience

AI Labs, American Express (Amex) **Bengaluru, India**
Research Engineer *Jul '19 - Present*
Developing new features and maintaining existing features for the AXGBoost algorithm (have contributed to 4 releases till date); exploring new research problems relevant to the business for future products.

Work Projects

Improvement of XGBoost Distributed Algorithm *May '20 - Present*
- Researching on weights used in Weighted Quantile Sketch algorithm by comparing trees built using Single Machine and Distributed XGBoost algorithm.
- Proved empirically and theoretically the scope for optimising weights used by the Distributed XGBoost algorithm. Working on deriving weights more faithful to the XGBoost objective function.

Development of a Universal Search Pipeline *Sep '19 - Aug '20*
- Developed an end-to-end system that enables context-aware search for enterprise-wide unstructured information retrieval using Neo4j, Machine Reading Comprehension algorithm, and Django. (Amongst the 5 projects selected out of 13 as part of an internal ideation workshop).
- Set up internal end-user tests and found that 73% of the queries are answered in the top 10 results.

Development of features for AXGBoost *Jul '19 - Present*
- Optimized the distributed algorithm code base from a 2400+ lines code to <500 lines for the new version of AXGBoost (Amex XGBoost) in C++ for better maintainability (has been used in building 10,000+ models internally).
- Working on providing GPU support using CUDA for the new version of AXGBoost algorithm.

Internships

AI Labs, American Express (Amex) **Bengaluru, India**
Research Intern *Jan '19 - Jun '19*
- Researched on open-source XGBoost algorithm with a focus on the distributed algorithm (in the Amex context).
- Improved the Approximate Split Point Proposal Algorithm used in distributed AXGBoost, which improved the capture rate on Amex datasets by 4%.
- Inherited functionality from XGBoost to design and implement the architecture for CSV data reading in AXGBoost.
- Improved column distributed data reading of CSV files in AXGBoost so that no column is skipped while reading.

Goldman Sachs **Bengaluru, India**
Summer Analyst *May '18 - Jul '18*
- Developed a generic parallel email scanner to enable easy access to conversations that went down for a deal.
- Developed the scanner using Microsoft Exchange Web Services and Java Spring Framework.
- Set up RabbitMQ queues for storing mails at intermediate steps, processed the mails to remove redundant information using text processing techniques, and finally stored them in MongoDB.

Gnowledge Lab, Homi Bhabha Centre for Science Education **Mumbai, India**
Summer Intern *May '17 - Jul '17*
- Developed a feature-rich offline search engine using Django for a digital learning platform (CLIx) to enable quick content access in schools with poor internet connectivity (deployed in 500 government schools).
- Worked on document ingestion and database initialization using Elasticsearch for diverse types of documents.
- Implemented functions to support suggestions, advanced triplet search, contribution search, and search filters.

Research Projects

Parallelization of K-Medoids Clustering Algorithm **BITS Pilani, India**
Advisor: *Prof. Poonam Goyal* *Aug '18 - Dec '18*
- Developed parallel K-Medoids algorithm using Adaptive Gridding for spatial partitioning in Spark Java.
- Improved the algorithm's efficiency of selecting initial medoids without compromising the clustering error (average sample size is 10x less than the state of the art - PAMAE) given any skewed data set.

Parallelization of Union-find Algorithm **BITS Pilani, India**
Advisor: *Prof. Poonam Goyal* *Jan '18 - May '18*
- Developed a communication efficient distributed Union-find algorithm using Open MPI in C++.
- Reduced the number of message passing operations between processes using deferred bulk updates.

Course Projects

Kinship Verification from Facial Images of Parents and their Kids
Machine Learning — **Instructor:** *Dr. Navneet Goyal* *Nov '18 - Dec '18*

- Compared qualitatively and quantitatively the existing techniques (Artificial Neural Networks, SVM, CNN, ensemble of SVMs) for Kinship Verification in R using Keras library.
- Used the results to design and implement an ensemble of Metric Learning based CNN architecture.
- Improved accuracy by 2.8% on the KinFaceW-1 dataset and by 3.1% on the KinFaceW-2 dataset.

Data Analysis and Modelling of Student Course Grades
Machine Learning — **Instructor:** *Dr. Navneet Goyal* *Sep '18 - Oct '18*

- Created a Bayesian Belief Network using bnlearn library in R based on grades of students, incorporating various hypotheses as to how attributes in data are related.
- The network can answer complex queries without being adversely affected by missing values, irrelevant attributes, and size of data.
- The network can be used to assess teaching pedagogies by modelling natural language queries as conditional probabilities.

Foster's Design Methodology on a Range-Queryable Distributed Data Structure (RAQ)
Parallel Computing — **Instructor:** *Prof. Shan Sundar Balasubramaniam* *Apr '18 - May '18*

- Designed a parallel algorithm to facilitate joining and leaving of peers from a peer to peer network (represented as RAQ data structure) using Foster's Design methodology with a commodity cluster as the target platform.
- Obtained logarithmic speedup and improved time complexity of joining mechanism compared to sequential execution.

Compiler for C-Like Language
Compiler Construction — **Instructor:** *Prof. Vandana Agarwal* *Jan '18 - Apr '18*

- Developed lexical, syntax, semantic analyzers, and code generator modules of a compiler for a language in C.
- Implemented functionalities to support simple functions, simple matrix operations, and conditional statements.

Design Word Document Index Creation for Shared Memory Systems
Parallel Computing — **Instructor:** *Prof. Shan Sundar Balasubramaniam* *Jan '18 - Feb '18*

- Designed a PRAM algorithm for document index creation using OpenMP in C++ for a UNIX based file system.
- Developed a scalable divide and conquer algorithm on a file system with up to 160,000 files.
- Reduced time taken to create an index from 43 seconds on 1 CPU core to 9 seconds on 32 CPU cores.

Teaching Assistantships

Discrete Structures for Computer Science *Aug '18 - Dec '18*

- Created take home assignments for a class of 150 students.

Database Systems *Jan '18 - May '18*

- Created lab sheets and conducted lab sessions for a batch of 40 students.

Object-Oriented Programming *Jan '17 - May '17*

- Conducted lab sessions and invigilated final lab exam for a batch of 40 students.

Certifications

- Mining Massive Datasets, offered by Stanford Online, eDX *May '20*
- CUDA Programming Masterclass, Udemy *Apr '20*
- Functional Programming Principles in Scala, EPFL, Coursera *Nov '18*

Technical Skills

C, C++, Java, Python, R, Scala, Scheme, MPI, OpenMP, Spark Java, Django

Awards and Scholarships

- **BITS Merit Scholar**: Received 80% Scholarship for being in the top 1% in all semesters. *Aug '15 – Jun '19*
- Awarded **Commendation Letter** from HRD Minister for being among the top 0.1% in class 12^{th} exam. *Jun '15*

Extra Curricular Activities

- Google AI Summer School HCI+AI for Social Good Track: among the 50 students selected for the school. *Aug '20*
- Member, Organizing Committee, American Express: Organized colleague engagement events. *Oct '19 - Present*
- Planted tree saplings as part of a Tree Plantation Drive by American Express. *Aug '19*
- Volunteered to teach underprivileged kids at an NGO as part of Community TeamWorks(Goldman Sachs). *Jun '18*