

Arnav Kumar

✉ a8kumar@uwaterloo.ca 🌐 arnavcs 🌐 arnavcs.github.io in Arnav Kumar

Skills

Languages: Go, Python, C++, JavaScript, GDScript, Haskell, Racket, Java, Bash, \LaTeX , HTML, CSS, Markdown, RegEx
Tools: Linux, Unix, Godot, Protobuf, Gomega, Ginkgo, Pandas, Matplotlib, NumPy, Beautiful Soup, Tensorflow, Scikit-learn, OpenCV

Education

Candidate for BSc. Computer Science (term 3A) @ University of Waterloo 2022 - Present
• 3.99/4.0 GPA after second year with all advanced level courses and received term distinction.
• Received the Ronald G. Dunkley National Scholarship (\$18,000), the President's Entrance Scholarship (\$3,500), the NSERC undergraduate research scholarship (\$6,000), and the Cenovus Energy STEM Scholarship (\$18,500).

Experience

Combinatorics Researcher @ University of Waterloo May 2024 - Aug 2024
• Worked on novel research exploring random graph orders and poset dimension utilizing probabilistic methods including configuration models, approximation techniques, and convergence in probability.
• Examined a conjecture about poset dimension, designing and running programs to search for counterexamples, and devising constructions to improve random graph order dimension bounds.
• Engaged in regular meetings to discuss and advance research, demonstrating capacity for effective collaboration.

Blockchain Software Engineer @ Dandelion Networks May 2023 - Aug 2023
• Developed and implemented an enhanced lattice syncing and node discovery algorithm in Go, querying peers with Protocol Buffers to identify and address missing blocks in the local lattice with secure and concurrent updates.
• Created and deployed a locally hosted debugging website in Go to be run by each node of the blockchain network.
• Designed multiple mock servers, clients, and services with Ginkgo and Gomega to test the algorithms.
• Created Jenkins pipeline which automatically build, test, vet, and format pushed code to catch regressions.

Data Science Intern @ Tektorch.ai Apr 2022 - Aug 2022
• Extracted features such as job title, location, classification, and pay from a job listing website using Python, Beautiful Soup 4, NumPy, and Pandas to create a dataset for analysis.
• Developed an end-to-end data visualization pipeline to ingest data and display trends using Matplotlib, and NumPy.

Machine Learning Research Intern @ University of Alberta Jul 2021 - Aug 2021
• Trained and employed Natural Language Processing (NLP) models to help determine the cause of a change in depressive language in Tweets as part of a psychology study.
• Created an evaluation framework with graphs and videos to analyse subject mouse position data during exams.
• Developed automation pipeline in Bash to reduce time spent by $20\times$ on the application of the evaluation framework.

Achievements

- 335th (top 10%), with score 26, **William Lowell Putnam Mathematical Competition** 2022
- Estuary Sponsor Prize Winner, **UofTHacks X** 2023
- 10th in Canada, **Asian Pacific Mathematics Olympiad (APMO)** 2022
- 18th in Canada, 2 time qualifier, **Canadian Mathematical Olympiad (CMO)** 2021 - 2022
- Invited to write, **United States of America Mathematical Olympiad (USAMO)** 2022
- Bronze medalist for machine learning project on prognosing IPF, **Canada Wide Science Fair** 2021

Projects

Bloom (Game) | *Godot*
• Game jam submission made in 72 hours with Godot: “you are a bee that pollinates flowers in a field to plant new hybrids”
• Implemented colour mixing, screen wrapping, movement, and flower spawning mechanics

InterPlanetary File Explorer (IPFE) | *Go, Python, Scikit-learn, Estuary, Co:here, Three.js*
• Created vector embeddings for files with their headers using Co:here's NLP embeddings to facilitate classification of files.
• Performed principal component analysis of the vector embeddings to reduce the dimensionality from 4096 to 3 to be plotted and displayed interactively in 3D space using Three.js.

Prognosing Idiopathic Pulmonary Fibrosis (IPF) | *Python, Tensorflow2, Pandas, Scikit-learn*
• Implemented an auto-encoder, linear regression, dense neural network, and bayesian model in order to accurately predict future lung capacity and give a confidence value using initial lung capacity data, age, sex, smoking status, and more.
• Obtained a Laplace Log Likelihood score of -6.9 (much better than the baseline score -8.1) with $\sigma \approx 200\text{mL}$.

Custom Language Interpreter | *Haskell*
• Created an expression evaluator for a self-made stack-based language.
• Implemented zipper traversal to determine which part of the stack has already been processed.