

Avishek Kumar

DATA SCIENTIST

5455 S Ingleside Ave 1E, Chicago IL, 60615, United States of America

732-439-4602 | avishek.raman.kumar@gmail.com | [avishekrk](https://www.linkedin.com/in/avishekrk) | [avishekrkumar](https://www.linkedin.com/company/avishekrkumar) | [@avishekrk](https://twitter.com/avishekrk) | [Avishek Kumar](#)

Summary

I am a data scientist that has worked on problems in fields such as fintech, healthcare, city planning, criminal justice, and genetic disease prediction using methods from machine learning and data science. I have built and debugged predictive models, performed model evaluation and selection, and deployed models into production settings. I have also trained data scientists through executive training programs and led teams of data scientists as a technical lead.

Professional Experience

Intuit

San Diego, CA

STAFF DATA SCIENTIST

Jun. 2019 - Present

- **Overdraft Early Warning System (Mint):** Built and deployed a machine learning system to predict when customers are likely to overdraft and prevent the overdraft through an email notification. This feature saved Mint customers 12 million dollars in a single year and was highlighted in both CNET and The Wall Street Journal.
- **Transaction Categorization (Mint):** (Tech Lead) Revamped the current Mint Transaction Categorization by developing a new BERT-based model. This model enabled the first Venmo and Zelle support as well as increased the overall accuracy of the model.
- **Forecasting for Personal Finance (Mint):** (Tech Lead) Led a team of three data scientists to develop an RNN-based forecasting model to forecast future spending of Mint customers.
- **Dynamic Upsell (TurboTax):** Built a model to select if a customer should be shown an upsell or not in TurboTax. The model decreased churn due to upsell fatigue by suppressing upsells for customers that are the least likely to be upsold.

Center for Data Science & Public Policy, The University of Chicago

Chicago, IL

DATA SCIENTIST/RESEARCH SCIENTIST

May. 2016 - Jun. 2019

- **Retention of HIV Patients in Medical Care (Healthcare):** Created a risk assessment tool to predict which HIV+ patients are likely to drop out of care for use by the University of Chicago HIV clinic and Chicago Department of Public Health. This work was published in Nature: Scientific Reports.
- **Prevention of Childhood Lead Poisoning (Healthcare):** Deployed a machine learning model for predicting which homes in the city of Chicago are likely to have lead hazards that lead to early childhood lead poisoning. This work was in partnership with the Chicago Department of Public Health, the Chicago Department of Innovation and Technology and Alliance HealthCare, awarded the Academy Health Local/State Innovation Prize.
- **Preventing Water Main Breaks (City Planning):** Built and deployed a machine learning system to identify which city blocks are most at risk of having water main breaks for the city of Syracuse, NY. This work has been featured in State Scoop, Water Online, and Politico and replicated in several cities as well as published in KDD.
- **The Cost of Technical Recidivism (Criminal Justice):** Developed a point-of-service model for predicting general recidivism for the Illinois Department of Corrections. Conducted a retrospective study to understand the effect of technical recidivism violations on an offender's ability to obtain future employment.
- **Executive Training:** Wrote and taught the curriculum for the Coleridge Initiative, a 3-month long course to train the heads of city and state government agencies on the use of data science methods to solve public policy problems.

Arizona State University, P.I. Associate Professor S. Banu Ozkan

Tempe, AZ

POSTDOCTORAL RESEARCH ASSOCIATE/SYSTEMS ADMINISTRATOR

August 2014 - September 2016

- Worked on problems related to protein dynamics, protein structure refinement, genetic disease prediction, and antibiotic resistance using molecular dynamics and machine learning methods.
- Wrote several software packages for studying protein dynamics and analyzing genetic disease in Python.
- Built and maintained a 1408 node supercomputer as systems administrator.
- Authored six publications in peer reviewed journals and mentored two doctoral students.

Arizona State University, P.I. Professor Michael F. Thorpe

Tempe, AZ

GRADUATE RESEARCH ASSISTANT

September 2009 - August 2014

- Developed software packages in C++, Python, and Fortran to study amorphous materials.
- Authored five publications in peer reviewed journals. Awarded over \$60,000 through multiple fellowships to fund research.

Education

Arizona State University

PH.D. IN PHYSICS (COMPUTATIONAL CONDENSED MATTER AND BIOLOGICAL PHYSICS)

Tempe, AZ

December 2014

Arizona State University

M.S. IN PHYSICS (PHI KAPPA PHI)

Tempe, AZ

December 2012

Carnegie Mellon University

B.S. IN PHYSICS WITH HONORS

Pittsburgh, PA

May 2009

Publications

Peer Reviewed Journal Articles

- Ose NJ, Butler BM, Kumar A, Kazan IC, Sanderford M, Kumar S, et al. (2022) Dynamic coupling of residues within proteins as a mechanistic foundation of many enigmatic pathogenic missense variants. PLoS Comput Biol 18(4): e1010006. <https://doi.org/10.1371/journal.pcbi.1010006>
- Ramachandran, A., Kumar, A., Koenig, H. et al. Predictive Analytics for Retention in Care in an Urban HIV Clinic. Sci Rep 10, 6421 (2020). <https://doi.org/10.1038/s41598-020-62729-x> (Equal Contribution)
- Brandon M. Butler, Avishek Kumar, S. Banu Ozkan. Dynamic allosteric residue coupling reveal disease mechanism for Gaucher disease and nSNVs across the proteome. (Submitted, Equal Contributor)
- Butler BM, Kazan IC, Kumar A, Ozkan SB (2018) Coevolving residues inform protein dynamics profiles and disease susceptibility of nSNVs. PLoS Comput Biol 14(11): e1006626. <https://doi.org/10.1371/journal.pcbi.1006626>
- Avishek Kumar, Syed Ali Asad Rizvi, Benjamin Brooks, R. Ali Vanderveld, Kevin H. Wilson, Chad Kenney, Sam Edelstein, Adria Finch, Andrew Maxwell, Joe Zuckerbraun, and Rayid Ghani. 2018. Using Machine Learning to Assess the Risk of and Prevent Water Main Breaks. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '18). ACM, New York, NY, USA, 472-480. DOI: <https://doi.org/10.1145/3219819.3219835>
- Avishek Kumar, Brandon M. Butler, Sudhir Kumar, Sefika B. Ozkan. Integration of structural dynamics and molecular evolution via protein interaction networks: a new era in genomic medicine. Current Opinion In Structural Biology 35, 135 (2015)
- Sean L. Seyler, Avishek Kumar, M.F. Thorpe, and Oliver Beckstein. Path Similarity Analysis: A method for Quantifying Macromolecular Pathways. PLoS Comput Biol (10): e1004568 (2015).
- Avishek Kumar, Paul Campitelli, M.F. Thorpe, S. Banu Ozkan, Partial Unfolding and refolding for structure refinement: A unified approach of geometric simulations and molecular dynamics. Proteins: Structure, Function, and Bioinformatics 83(12), 2279 (2015)
- Avishek Kumar, T.J. Glembo, S. Banu Ozkan, The Role of Conformational Dynamics and Allostery in Disease Development. Biophysical Journal 109(6), 1273 (2015)
- W.G. Ellenbroek, Varda F Hagh, Avishek Kumar, M.F. Thorpe, Martin van Hecke, Rigidity Loss in Disordered Systems: Three Scenarios. Phys. Rev. Lett. 114, 135501 (2015)
- Avishek Kumar, Mark Wilson, David Sherrington, M.F Thorpe, Ring statistics of silica bilayers. Journal of Physics: Condensed Matter 26, 395401 (2014)
- Mark Wilson, Avishek Kumar, David Sherrington, M.F. Thorpe, Modeling vitreous silica bilayers. Phys. Rev. B 87, 214108 (2013)
- D Van Tuan, Avishek Kumar, Stephan Roche, F. Ortmann, MF Thorpe, P Ordejon, Insulating behavior of an amorphous graphene membrane. Phys. Rev. B 86 (12), 121408 (2012)
- Avishek Kumar, Mark Wilson, M.F. Thorpe, Amorphous graphene: a realization of Zachariasen's glass. Journal of Physics: Condensed Matter 24(48), 485003 (2012)
- Y Li, F. Inam, Avishek Kumar, M.F. Thorpe, D.A. Drabold, Pentagonal puckering in a sheet of amorphous graphene. physica status solidi (b). doi: 10.1002/pssb.201147195 (2011)

Book Chapters

- Using GPUs to Study Protein Evolution in Order to Understand Antibiotic Resistance, *Strategies in Biomedical Data Science: driving force for innovation* by Jay A. Etchings
- Workbooks *Big Data and Social Science: Data Science Methods and Tools for Research and Practice* by Ian Foster, Rayid Ghani, Ron S. Jarmin, Frauke Kreuter and Julia Lane

Teaching Experience

Coleridge Initiative, The University of Chicago

Chicago, IL

INSTRUCTOR

October 2017 - Present

- The Coleridge Initiative is an executive education course that is taken by members of city and state government in order to learn how to use data science tools and methods to solve public policy problems.
- I developed and taught the curriculum for the class. Topics include Command Line, Git, SQL, Machine Learning, Text Analysis, GIS Analysis, Network Analysis

Department of Physics, Arizona State University

Tempe, AZ

TEACHING ASSISTANT

January 2013 - August 2014

- Held Online Recitations using Adobe Connect.
- Recorded Recitations using ScreenCast and developed modus operandi for Recorded Recitations.
- Proctored Examinations.

Department of Physics, Arizona State University

Tempe, AZ

INSTRUCTOR PHY112

August 2012 - December 2012

- Lectured three times a week to 120 students.
- Responsible for all aspects of the course—lectures, examinations, and recitations.

Honors & Awards

2016	Data Science for Social Good Fellowship
2014	Winner Inaugural Clinton Global Initiative University Codeathon
2011-2014	ARCS Foundation Fellowship for College Scientists
2013	Outstanding Graduate Teaching Assistant Award
2012	Wally Stoetzel Physics Fellowship
2011	Molecular Imaging Corporation Fellowship
2011	Phi Kappa Phi
2009	Graduate Assistance in Areas of National Need (GAANN) Fellowship
2009	Sigma XI, The National Honor Society in Scientific Research

Service

2014-Present	Instructor Software Carpentry Workshops
2010-2014	Organizer Graduate Seminar Series
2011	Judge Arizona Middle School Science Bowl
2010-2014	Reviewer GPSA Grant Committee

Skills

Proficient	Machine Learning, Data Mining, GIS, Text Analysis, Network Analysis, Parallel Computing
Programming	Python, SQL, R, C/C++, Java, FORTRAN90, JavaScript
DevOps	AWS, Docker, Kubernetes, CircleCI
Software	Git, PostgreSQL, SQLite, \LaTeX , Mathematica, Matlab, NumPy, SKlearn, Tensorflow, Pandas