

## Bach Viet Do

---

CONTACT INFORMATION	311 West Hall Department of Statistics University of Michigan Ann Arbor, MI 48109-1107 USA	<i>E-mail:</i> vietdo@umich.edu <i>Google Scholar Profile:</i> Bach Viet Do <i>Linkedin:</i> <a href="https://www.linkedin.com/in/bach-viet-do-632097b7/">https://www.linkedin.com/in/bach-viet-do-632097b7/</a>
EDUCATION	<b>University of Michigan</b> , Ann Arbor, Michigan USA Doctor of Philosophy, Statistics, May 2023 <ul style="list-style-type: none"><li>Advised by: Prof. Yang Chen and Prof. Long Nguyen. Thesis: Mixture Modeling: Solar Application and Misspecification Behaviors.</li></ul> <b>Columbia University</b> , New York, New York USA Master of Science, Computer Science, Dec 2016	
PUBLICATION	Uncovering Heterogeneity of Solar Flare Mechanism With Mixture Models, <i>Frontiers in Astronomy and Space Sciences</i> (2024).  Prediction of Protein–ligand Binding Affinity from Sequencing Data with Interpretable Machine Learning, <i>Nature</i> 2022.	
PREPRINT	Forecasting Automotive Shortfalls Chain Disruption with Heterogeneous Time Series, presented at INFORMS' 2024 Annual Meeting and under review targeted for INFORMS Data Science and	
PROFESSIONAL EXPERIENCE	<b>Meta Platforms, Inc</b> , Menlo Park, CA USA <i>Machine Learning Research Scientist</i> <b>August 2024 - Present</b> Building cutting-edge recommendation systems at scale, leveraging the latest research in sequential recommendations, large language models, and multimodal foundation models. The technologies used include PyTorch and Meta in-house ML platforms.  <b>Ford Motor Company</b> , Dearborn, MI USA <i>Research Data Scientist, Global Data Insights &amp; Analytics</i> <b>June 2023 - July 2024</b> Developed models for complex time series data to optimize the supply chain at Ford. My models leverage modern deep learning architectures, such as Sequence-to-Sequence and Transformers, to transform time series data into suitable latent representations before performing inference with advanced statistical models like Parametric and Semi-Parametric Survival Analysis. The technologies used include PyTorch, Google Cloud Platform (GCP), and Vertex AI.  <b>University of Michigan</b> , Ann Arbor, MI USA <i>Graduate Student Research Assistant</i> <b>August 2017 - April 2023</b> Conducted original research in Statistics, Data Science, and Machine Learning to further literature knowledge and contribute to the field of Statistics.  <b>Amazon</b> , Seattle, WA USA <i>Applied Scientist Intern</i> <b>June 2020 - August 2020</b> Implemented Machine Learning and Deep Learning models to enhance the recommendation engine for Amazon.com. The objective was delivering optimal personalized purchasing advice for online customers. Utilized technologies included TensorFlow, AWS, and SageMaker.  <b>eBrevia</b> , New York, NY USA	

*ML/NLP Software Engineer*

**Mar 2017 - May 2017**

Utilized machine learning and natural language processing (NLP) techniques to develop artificial intelligence software automating the legal contract analysis traditionally conducted by junior lawyers in law firms.

TEACHING  
EXPERIENCE

**University of Michigan**, Ann Arbor, Michigan USA

*STATS 306 Introduction to Statistical Computing, Undergrads, size: 160, GSI* **Winter 2023**  
*STATS 470 Introduction to the Design of Experiments, Undergrads, size: 60, GSI* **Fall 2023**  
*STATS 504 Practice and Communication in Applied Statistics, Masters, size: 30, GSI* **Fall 2021**  
*STATS 551 Bayesian modeling and computation, Masters, size: 80, GSI* **Winter 19, 20**  
*STATS 506 Computational Methods & Tools in Statistics, Masters, size: 50, GSI* **Fall 2019, 20**  
*STATS 499: Honors Thesis, Undergraduate, size: 5, GSI* **Fall 2019**  
*STATS 449 Topics in BioStats, Undergraduate, size: 80, GSI* **Winter, Fall 2018**  
*STATS 412 Intro to Probability & Statistics Undergraduate, size: 200, GSI* **Fall 2017**

**Columbia University**, New York, New York USA

*COMS4771 Machine Learning, Undergraduate, size: 200, TA* **Fall 2016**  
*COMS3203 Discrete Mathematics, Undergraduate, size:150, Head TA* **Fall 2016**  
*COMS4701 Artificial Intelligence, Undergraduate, size: 60, TA* **Summer 2016**

HONORS AND  
AWARDS

Graduate Student Instructor Excellence in Teaching, 2022  
University of Michigan Ann Arbor PhD Fellowship, 2017  
Columbia Computer Science CA Fellowship, 2016

SERVICE

University of Michigan Student Seminar Coordinator, Fall 2019, Winter 2020  
University of Michigan PhD Recruitment Visit Day Helper, Winter 2016  
Columbia University Computer Science Peer Advisor Fall 2016  
Columbia University Computer Science MS 2016 Orientation Advisor  
Columbia University Computer Science Peer Advisor Spring 2016

COMPUTING SKILLS

- Statistical Packages: R, Matlab
- Languages: Python, R, SQL, Java, Unix-Linux/Shell