RISHABH MISRA

r1misra.eng.ucsd.edu | rishabhmisra.github.io | misrarishabh | @rishabh_misra_

Education

University of California, San Diego — *La Jolla, CA*

Master of Science - Computer Science (Artificial Intelligence specialization), GPA: 3.93/4.0 — *June 2018*

Thapar University — *Patiala, India*

Bachelor of Technology - Computer Engineering, GPA: 9.88/10.0, Rank 1, Gold Medalist. — *July 2015*

Experience

Twitter — *San Francisco, CA*

Senior Machine Learning Engineer — *July 2019 - PRESENT*

- Working in the Content Quality team that drives foundational infrastructure, core ML modeling, and thought leadership on ML opportunities across the products that enable creation and conversations on Twitter.
  - Engineering better features and models to improve user engagements on conversational products of Twitter.
  - Building and scaling end-to-end Machine Learning system to rank tens of millions of candidates.
  - Performing Data Science analyses to identify potential problem statements and their impact on user experience.
  - Driving ML tooling (in BigQuery and GCP) adoption to improve the velocity of ML lifecycle in the team.
- Technologies: Python, Java, Scala, Tensorflow, Scalding, Airflow, Kubeflow, BigQuery, Google Cloud Platform (GCP).

Amazon — *Seattle, WA*

Software Development Engineer — *July 2018 - July 2019*

- Worked in Amazon Global that enables customers to buy products internationally based on export eligibility.
  - Improved the infrastructure scalability by designing solutions using Native AWS technologies.
  - Conducted experiments to improve the eligibility prediction of products using Machine Learning models.
  - Assisted courses taught in Amazon's Machine Learning University.
- Technologies: AWS Technologies, Java, Python, Jupyter Notebook.

McAuley Lab at UC San Diego — *La Jolla, CA*

Graduate Researcher — *April 2017 - June 2018*

- Worked under the guidance of Prof. Julian McAuley towards several novel user behavior modeling and NLP problems.
  - Product size recommendation: Proposed a framework based on Latent Factor Model and Metric Learning to predict fit of different catalog sizes of clothing products. Contributed a public dataset and improved upon an algorithm developed by Amazon by 18%.
  - Fine-Grained Spoiler Detection: Developed a Hierarchical RNN architecture to detect spoiler sentences in review corpora we collected as part of this work. Attention mechanism in the architecture reveals interesting spoiler cues. We beat strong baselines by 3%.
  - Addressing Marketing Bias in Product Recommendations: Developed a framework to address potential marketing bias that significantly improves the recommendation fairness across different market segments, with a negligible loss (or better) recommendation accuracy.

Amazon — *Seattle, WA*

Software Development Engineering Intern — *June 2017 - September 2017*

- Worked in the DataForge team that provides a platform for smartly scheduling Big Data operational workloads within SLAs.
  - Designed support for primary key constraint and batch inserts/updates, using append-only table and multi-version concurrency control concepts, while ensuring consistent reads in Hive.
  - Support for transactional operations in Hive.
  - Support for compaction (carefully discarding old data) without blocking other operations.
- Technologies: Java, Hive, DynamoDB.

University of California, San Diego — *La Jolla, CA*

Graduate Teaching Assistant — *January 2017 - March 2018*

- *CSE 258: Recommender Systems* — Guided graduate students on various Recommender Systems concepts, helped them figure out challenges for session-end projects and graded tests.
- *CSE 110: Software Engineering* — Hosted info sessions on Software Engineering Design Patterns and aided Android project development.

Arcesium (A D.E. Shaw Company) — *Hyderabad, India*

Software Engineer — *July 2015 - July 2016*

- I worked in the Arcesium/Tech division's "Straight Through Processing" team. Some of my important responsibilities include:
  - Adding support for self-sanitization, self-recovery and fault tolerance in the new infrastructure built using JAVA.
  - Adding a self-aware triggering mechanism for Blotters, greatly minimizing data completeness issues.
  - Profiling and optimizing (around 40%) code (using concurrency) and database (using index and partitions).
- Technologies: Java, Spring, MyBatis, SQL Server.

Indian Institute of Technology, Madras *Chennai, India*

Research Intern *June - August 2014, January - May 2015*

- Worked under the guidance of Prof. Balaraman Ravindran towards developing scalable Bayesian algorithms for Recommender Systems.
  - Scalable Bayesian Matrix Factorization: The proposed algorithm reduces the cubic time complexity of existing algorithm to linear.
  - Scalable Variational Bayesian Factorization Machine: Framework for factorization machines that converges to optimal solution in an order of magnitude less iterations.

Research Publications

212 citations as of June 20th, 2022.

- **Do not fake it till you make it! - Synopsis of trending fake news detection methodologies**: Book Chapter by Rishabh Misra and Jigyasa Grover, accepted for publication in book "Deep Learning for Social Media Data Analytics" of Springer book series "Studies in Big Data", ISSN 2197-6503 Sep. 2022.
- **Sculpting Data for ML: The first act of Machine Learning**: Book by Jigyasa Grover and Rishabh Misra, Jan. 2021. ISBN-13: 979-8585463570.
- **Addressing Marketing Bias in Product Recommendations**: Mengting Wan, Jianmo Ni, Rishabh Misra, Julian McAuley, in Proceedings of 2020 ACM Conference on Web Search and Data Mining (WSDM'20), Houston, TX, USA, Feb. 2020. (15% acceptance rate)
- **Fine-Grained Spoiler Detection from Large-Scale Review Corpora**: Mengting Wan, Rishabh Misra, Ndapa Nakashole, Julian McAuley, in Proceedings of 57th Association for Computational Linguistics 2019 (ACL'19), Florence, Italy, Jul. 2019. (18% acceptance rate)
- **Sarcasm Detection using Hybrid Neural Network**: Rishabh Misra, Prahal Arora, arXiv preprint arXiv:1908.07414 (2019).
- **Decomposing Fit Semantics for Product Size Recommendation in Metric Spaces**: Rishabh Misra, Mengting Wan, Julian McAuley, in Proceedings of 2018 ACM Conference on Recommender Systems (RecSys'18), Vancouver, Canada, Oct. 2018. (25% acceptance rate)
- **Hotel recommendation system**: Aditi A Mavalankar*, Ajitesh Gupta*, Chetan Gandotra*, Rishabh Misra*, arXiv preprint arXiv:1908.07498 (2018). *equal contribution
- **Scalable Variational Bayesian Factorization Machine**: Avijit Saha, Rishabh Misra, Ayan Acharya, and Balaraman Ravindran, preprint 2017.
- **Scalable Bayesian Matrix Factorization**: Avijit Saha*, Rishabh Misra*, Balaraman Ravindran, In Proceedings of 6th International Workshop on Mining Ubiquitous and Social Environments (MUSE), co-located with the ECML PKDD 2015. *equal contribution

Dataset Publications

74k+ downloads & 500k+ views on Kaggle; Used in DeepLearning.AI's "Natural Language Processing in TensorFlow" course on Coursera & Youtube (taken by 300k+ people to date)

- **News Headlines Dataset**: Misra, Rishabh, DOI: 10.13140/RG.2.2.16182.40004 (2018).
- **News Category Dataset**: Misra, Rishabh, DOI: 10.13140/RG.2.2.20331.18729 (2018).
- **IMDB Spoiler Dataset**: Misra, Rishabh, DOI: 10.13140/RG.2.2.11584.15362 (2019).

Research Committees

Invited Program Committee Member at some of the leading research conferences. Completed 31 peer reviews as of June 2022.

- International AAAI Conference on Web and Social Media (ICWSM) 2023.
- International Conference on Machine Learning (ICML) 2022.
- ACM's Conference on Knowledge Discovery and Data Mining (SIGKDD) 2022.
- ACM's Special Interest Group in Information Retrieval (SIGIR) 2022.
- IEEE International Conference on Data Mining (ICDM) 2022.
- ACM's Conference on Recommender Systems (RecSys) 2022.
- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) 2022.
- ACM Transactions on Recommender Systems (TORS) 2022.
- International AAAI Conference on Web and Social Media (ICWSM) 2022.
- International Conference on Advances in Databases, Knowledge, and Data Applications (DBKDA) 2022.
- Scientific Computing with Python (SciPy) 2019.

Other Notable Achievements

Blogs Machine Learning blogs on Towards Data Science online publication have 116k+ views. Additionally, content on personal website has been viewed 80k+ times by people from 169 countries.

Media Coverage Spoiler Detection research featured in TechCrunch, Gizmodo, NBC, Geek.com, TechXplore, and UC News. which have up to 16M monthly readership.

Patents Two US patents with application numbers 63/267,780 and 63/362,556 are under review.

Scholarships/Awards Python Software Foundation Grant (2x), Merit scholarships throughout 4 years of undergraduate education, University Medal, Summer Fellowship from Indian Institute of Technology, Madras, Yuuvis SF Hackathon Winner.

Conference Talks PyCon US, RVA Tech Data Summit, Re-Work Enterprise AI Summit, All Things Open (+ *book signing*), LeadDev Live, ML Conference, ACM Conference on Recommender Systems, This Week in ML & AI