RISHABH MISRA

+1-650-686-9605 | r1misra@eng.ucsd.edu | rishabhmisra.github.io | rishabhmisra | misrarishabh

5+ years of engineering experience in shipping low-latency massive-scale systems to production. Past experience at Twitter, Amazon, and Arcesium (a D.E. Shaw Company). Seeking backend engineering positions preferably in San Francisco Bay Area or Remote. I would require H1B sponsorship.

Experience

Twitter
San Francisco, CA

Senior Machine Learning Engineer — *July 2019 - PRESENT*

- One of the founding engineers of the Conversations Quality team that works on tweet replies ranking service to drive meaningful conversations.
  - Shipped a Light Ranking module in reply ranking service for graceful degradation and scaled the service to systematically rank tens of millions of candidates per second. This led to 20% gain in p9999 latency while improving health metrics by 5% overall.
  - Scaled up reply ranking service further by designing solutions like conditional feature population request, batching requests, caching, and sharding.
  - Drove modernization of ML stack by architecting framework using KubeFlow to train next generation native Tensorflow models. Utilized other GCP technologies like DataFlow and BigQuery to improve end-to-end model training duration by 10x.
  - Shipped novel features & predictive models based on user engagement graph on Twitter to drive >30% gains in key product metrics.
- Technologies: Python, Java, Scala, Hadoop, Airflow, Tensorflow, Kubeflow, BigQuery, DataFlow, Google Cloud Platform (GCP), Scalding.

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.

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.

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*

Research Publications

277 citations as of November 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", ISBN: 978-3-031-10868-6 Sep. 2022.
- **Sculpting Data for ML: The first act of Machine Learning**: Book by Jigyasa Grover* and Rishabh Misra*, Jan. 2021. Independently published. ISBN-13: 979-8585463570. *equal contribution
- **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.
- **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.
- **Sarcasm Detection using Hybrid Neural Network**: Rishabh Misra, Prahal Arora, arXiv preprint arXiv:1908.07414 (2019).
- **Hotel recommendation system**: Aditi A Mavalankar*, Ajitesh Gupta*, Chetan Gandotra*, Rishabh Misra*, arXiv preprint arXiv:1908.07498 (2019). *equal contribution
- **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.
- **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

80k+ 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

- TheWebConf 2023, ICML 2022, SIGKDD 2022, SIGIR 2022, ICDM 2022, RecSys 2022, ECML-PKDD 2022, ICWSM 2023, 2022, TORS 2022.

Other Notable Achievements

Blogs Machine Learning blogs on Towards Data Science online publication have 125k+ views. Additionally, content on personal website has been viewed 85k+ times by people from 169 countries.

Media Coverage Spoiler Detection research featured in TechCrunch, Gizmodo, Times of India, NBC, Geek.com, TechXplore, and UC News. which have up to ∼ 148M 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