Rishabh Misra

rishabhmisra.github.io r1misra@eng.ucsd.edu

EDUCATION

University of California San Diego **GPA: 3.93/4**
MS Computer Science (Specialization in Machine Learning) *Sep 2016 - Jun 2018*

Thapar University, India **GPA: 9.88/10**
BE Computer Engineering *Jul 2011 - Jul 2015*

RECENT INDUSTRY EXPERIENCE

Machine Learning Engineer II @ Twitter Inc., San Francisco *Jul 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 offline and online user engagement metrics.
  - **Building and scaling** end-to-end Machine Learning pipelines to rank millions of candidates.
  - Performing **Data Science analysis** to identify potential problems and their impact on user satisfaction.
  - Driving ML tooling adoption (BigQuery and GCP) to speed up the exploratory analysis process.
- **Technologies**: Python | Scala | Tensorflow | Scalding | Hadoop | Airflow | BigQuery | GCP

Software Development Engineer @ Amazon.com, Seattle *Jul 2018 - Jul 2019*

- Worked for **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.
- **Technologies**: AWS Technologies | Java | Python | Jupyter Notebook

Past experiences are summarized at rishabhmisra.github.io/experience.

RECENT ML PUBLICATIONS

Sculpting Data for ML: The first act of Machine Learning *(Book published via Amazon)*

- The book introduces the readers to the first act of Machine Learning, **Dataset Curation**. This **step-by-step guide** accompanies **code examples** in Python from the **extraction of real-world datasets** and **practical tips to identify valuable information** on web. In addition, it also dives deep into how **data fits into the Machine Learning ecosystem** and highlights the impact of **data-centric approaches** on ML system's performance.
- Supported by leading experts in Academia and Industry: Julian McAuley, Laurence Moroney, and Mengting Wan.

Addressing Marketing Bias in Product Recommendations *(Published at WSDM 2020)*

- Recognizing that consumer interaction might be **biased by how product is marketed**, we sought to understand how that affects the classic Recommender Systems algorithms and how to correct for this bias.
- We study this phenomenon for different **consumer-product market segments** on two e-commerce datasets.
- We develop a framework to address this potential marketing bias that **significantly improves the recommendation fairness** across different market segments, with a **negligible loss (or better) recommendation accuracy**.

Fine-Grained Spoiler Detection from Large-Scale Review Corpora *(Published at ACL 2019)*

- Contributing large-scale book review dataset that includes fine-grained spoiler annotations at the sentence-level.
- Incorporating the findings from exploratory analysis, we developed a **Hierarchical RNN architecture** to detect spoiler sentences in review corpora. Attention mechanism in the architecture reveals interesting spoiler cues.
- Experimental results demonstrate that our method outperforms strong baselines by nearly **3%**.

Decomposing Fit Semantics for Product Size Recommendation *(Published at RecSys 2018)*

- Proposed a framework based on **latent factor model** and **metric learning technique** to predict fit of different catalog sizes of clothing products for recommendation.
- Contributed the only publicly available datasets (at the time) for the catalog size recommendation problem.
- Observed an improvement of up to **18%** over an algorithm developed by Amazon.

Other publications are summarized at rishabhmisra.github.io/publication.

KEY ML PROJECTS

Sarcasm Detection using Hybrid Neural Network ***Python | PyTorch***

- Collected a news headlines-based dataset which improves upon frequently used Twitter datasets by removing the noise in label and language.
- Developed **interpretable hybrid neural network architecture** (CNN + RNN) with attention mechanism which improves baseline by **5%**. Attention module provides insights about the cues that make sentences sarcastic.

Jointly Modeling Aspects, Ratings and Sentiments with Temporal Dynamics ***Python***

- Implemented a **probabilistic graphical framework** which utilizes data from product reviews to jointly model aspects of the products, user sentiment on products and associated ratings to predict the unknown ratings.
- For interpretability, model **produces insights** on the various aspects of products and user sentiment on them.
- Incorporated **temporal information** into the joint model which improves performance by **1%** and additionally provides insights into **how users' preference of different product aspects change over time**.

Hierarchical Attention Network for Rating Prediction ***Python | Keras***

- Implemented a **hierarchical RNN** with attention mechanism that uses product reviews to predict the product ratings.
- Attention mechanism allows the RNN to focus on words and sentences that **best explain the rating** given to an item and uses this knowledge to predict unknown ratings.

Other major projects are summarized at rishabhmisra.github.io/projects.

ML TEACHING EXPERIENCE

Teaching Assistant @ Amazon's Machine Learning University *Jan 2019 - Apr 2019*

- **Introduction to Data Science** | **Instructor**: Zachary Levin (Senior Data Scientist)
- **Text Mining** | **Instructor**: Pascual Martinez-Gomez (Applied Scientist II)

Teaching Assistant @ UC San Diego *Fall 2017*

- **Recommender Systems and Web Mining (CSE 258)** | **Professor**: Dr. Julian McAuley

ACHIEVEMENTS AND POSITIONS OF RESPONSIBILITY

- My research publication have over 120+ citations with an h-index of 6.
- Delivering talks and workshops on dataset curation and data-centric approaches in AI: recently at All Things Open, LeadDev Live, ML Conference, and Algorithm Conference. Others are noted at rishabhmisra.github.io.
- Ranked in **Top 100** dataset contributors on the Kaggle Platform. My datasets have collectively **1300+** upvotes, **60K+** downloads, and **400K+** views.
- My Sarcasm Detection dataset was used in **Deeplearning.ai**'s NLP in TensorFlow course on Coursera for teaching purposes.
- My research work on **Spoiler Detection** got featured in TechCrunch, NBC, Gizmodo, and Geek.com among other channels.
- I won the Yuuvis SF Hackathon for building an Alexa skill to easily store, retrieve and share documents using the Yuuvis API.
- Served as a **Program committee member** and **reviewer** for the SciPy 2019 conference.
- Served as a **reviewer** for Amazon's Machine Learning Conference (**AMLC**) 2019.
- Received a **financial grant** from Python Software Foundation to attend PyCon 2019 in Cleveland, Ohio.
- I write blogs on Machine Learning concepts with Towards Data Science online publication. My stories have **90K+** views and **30K+** reads.
- **Mentored** first-generation undergraduate students at UCSD under the JUMP mentorship program.
- **Mentored** newly joined graduate students at UCSD as part of Graduate Women in Computing.
- Received **university medal** for being the topper of Computer Engineering batch at Thapar University.
- Regional finalist for **ACM-ICPC** Asia Region, Kanpur site 2013 held at IIT Kanpur.