

Rishabh Misra

🌐 rishabhmisra.github.io

✉ r1misra@eng.ucsd.edu

EDUCATION

University of California San Diego

MS Computer Science (Specialization in Machine Learning)

GPA: 3.93/4

Sep 2016 - Jun 2018

Thapar University, India

BE Computer Engineering

GPA: 9.88/10

Jul 2011 - Jul 2015

RECENT INDUSTRY EXPERIENCE

Machine Learning Engineer II @ Twitter Inc., San Francisco

Jul 2019 - Present

- Member of **Timelines Quality** team working towards developing and deploying ML models to improve the user experience on the Home Timeline and Tweet Detail page.
 - Working on **engineering better features and models** to drive up the content quality on Twitter's Home Timeline and Tweet Detail pages.
 - Performing **Data Science analysis** to identify potential problems and their impact on user satisfaction.
 - Working on building end-to-end **ML training and deployment pipelines** using Airflow.
 - Developing **ML tooling** using BigQuery and GCP to **accelerate exploratory analysis by 10x**.
- **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.
 - Improving the **infrastructure scalability** by designing solutions using Native AWS technologies.
 - Conducting experiments to **improve the eligibility prediction** of products using **Machine Learning** models.
- **Technologies:** AWS Technologies | Java | Python | Jupyter Notebook

Software Development Engineering Intern @ Amazon.com, Seattle

Jun 2017 - Sep 2017

- Interned at **Financial Intelligence Systems** team which provides a platform for running big data operational workloads consistently within service level agreement. Worked towards designing and implementing:
 - Support for **ACID properties and non-blocking compaction** (discarding stale data) with consistent reads in Hive.
 - A solution based on the concepts of append-only table and multi-version concurrency control.
- **Technologies:** Java | Hive | DynamoDB

RECENT ML PUBLICATIONS

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 collected.
- 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** for the catalog size recommendation problem.
- Observed an improvement of up to 18% over an algorithm **developed by Amazon**.

SELECTED 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.

ML TEACHING EXPERIENCE

Teaching Assistant @ Amazon's Machine Learning University

Jan 2019 - Apr 2019

- **Introduction to Data Science**
 - **Session:** 1-2019 | **Instructor:** Zachary Levin (Senior Data Scientist)
- **Text Mining**
 - **Session:** 2-2019 | **Instructor:** Pascual Martinez-Gomez (Applied Scientist II)

Teaching Assistant @ UC San Diego

Sep 2016 - Mar 2018

- **Recommender Systems and Web Mining (CSE 258)**
 - **Session:** Fall 2017 | **Professor:** Dr. Julian McAuley

ACHIEVEMENTS AND POSITIONS OF RESPONSIBILITY

- Co-hosting a workshop on **How to curate quality datasets for machine learning** at UT Austin for **Algorithm Conference 2020**. (Postponed due to COVID-19)
- Ranked in **Top 20** dataset contributors on the **Kaggle** Platform. My datasets have collectively **900+** upvotes, **30,000+** downloads, and **150+** kernels.
- **Sarcasm Detection dataset** was used in DeepLearning.ai's **NLP in TensorFlow** course on Coursera for teaching purposes.
- Won the **Yuuvis SF Hackathon** for building an **Alexa skill** to easily store, retrieve and share documents using the **Yuuvis API**.
- Research work on **Spoiler Detection** got featured in **TechCrunch**, **NBC**, **Gizmodo**, and **Geek.com** among others.
- **Program committee member** and **reviewer** for the **SciPy 2019** conference.
- **Reviewer** for Amazon's Machine Learning Conference (**AMLC**) 2019.
- Received a **financial grant** from **Python Software Foundation** to attend **PyCon 2019** in Cleveland, Ohio.
- **Technical Writer** for the **Towards Data Science** where I write blogs on **Machine Learning concepts**.
- **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.