Abhishek Sharma

Machine Learning Engineer
Oversea-Chinese Banking Corporation Ltd.,
Singapore

Web: https://abhishek9sharma.github.io/
Scholar: https://scholar.google.com.sg/citations?user=mbwvow0AAAAJ&hl=en

EDUCATION

Ph.D.	School of Information Systems, Singapore Management University, Aug 2014 - January 2019 Dissertation Title: *Social Software Development: Insights and Solutions*
B.Tech.	Computer Science & Engineering, National Institute of Technology, Hamirpur, India, 2004-2008

ACADEMIC APPOINTMENTS

May-Dec 2019	Singapore Management University, Singapore Postdoctoral Scholar, Living Analytics Research Centre
Aug 2014 -Jan 2019	Singapore Management University, Singapore PhD Candidate, School of Information Systems

INDUSTRY EXPERIENCE

2022 - Present : **OCBC Ltd.,** *Machine Learning Engineer*, **Singapore**

Major contributions:

- **Hydra**: A machine learning framework used across OCBC for running data science workloads. It manages jobs, services, and applications across various platforms, such as Ray, Cloudera, Dagster, etc.

- **Hydra Guardrails**: A Python framework + Ray Service that enables open source GUARDRAILS-AI compatible guards to be deployed in OCBC.

- **OCBC Wingman**: Provides **LLM**-based intelligent code and chat assistance to developers inside their *code-server* & *IntelliJ IDEA* IDEs. Additionally, it performs intelligent code reviews for approximately 700+ Pull Requests daily.

- OCBC's internal **LLM** Ray Serving Platform, which exposes LLM models such as *Llama3-Instruct(Nous)* across the enterprise as REST endpoints.

- Containerized CI pipelines for Data Science projects. All production-ready data science projects in OCBC AI Lab use these pipelines to test and deploy their code to production.

2019 - 2022 : **SourceClear (Veracode),** *Senior Software Engineer (Machine Learning)***, Singapore**
Wore dual hats of *Machine Learning Engineer* on some days and *Data Scientist* on others. Managed the ML life cycle from data preparation to model deployment on *AWS EKS Cluster*. Major contributions made:

- Improved the training pipeline by adding model validation on historical data as an additional decision point alongside standard accuracy metrics.

- Introduced tests and integrated them with the CI/CD pipeline increasing test coverage from 0 to 74%.

- Championed machine learning research efforts to evaluate use of deep learning models to replace existing prediction models. Integrated LightXML based multi-label text classification model in the training and prediction pipeline.

2011 - 2014 : **Accenture,** *Analyst Programmer***, Bengaluru (India); Liverpool/Horsham (UK)**
Lead BizTalk developer for a major UK insurance client. Responsibilities included:

- Requirements analysis, solution design & implementation

- On-boarding, mentoring & managing junior developers

2008 - 2011 : **Tata Consultancy Services Ltd.,** *Assistant Systems Engineer***, Gurgaon (India)**
Systems engineer for a major UK airlines client. Responsibilities included:

- Understanding requirements from business stakeholders and coding modules accordingly

- Application support & management

RESEARCH INTERESTS

Mining Software Repositories

Empirical Software Engineering

Software Analytics (Text, Socio-Technical Networks)

Applied Machine Learning

PUBLICATIONS

Peer-Reviewed Journal Articles

2021   Bowen Xu, Thong Hoang, **Abhishek Sharma**, Chengran Yang, Xin Xia, and David Lo. Post2vec: Learning distributed representations of stack overflow posts. *IEEE Transactions on Software Engineering*, pages 1–1, 2021

2021   Gede Artha Azriadi Prana, **Abhishek Sharma**, Lwin Khin Shar, Darius Foo, Andrew E Santosa, Asankhaya Sharma, and David Lo. Out of sight, out of mind? how vulnerable dependencies affect open-source projects. *Empirical Software Engineering*, 26(4):1–34, 2021

2020   Agus Sulistya, Gede Artha Azriadi Prana, **Abhishek Sharma**, David Lo, and Christoph Treude. Sieve: Helping developers sift wheat from chaff via cross-platform analysis. *Empirical Software Engineering*, 25(1):996–1030, 2020

2018   **Abhishek Sharma**, Yuan Tian, Agus Sulistya, Dinusha Wijedasa, and David Lo. Recommending who to follow in the software engineering twitter space. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 27(4):16:1–16:33, October 2018

Peer-Reviewed Conference Proceedings

2022   Stefanus A Haryono, Hong Jin Kang, **Abhishek Sharma**, Asankhaya Sharma, Andrew Santosa, Ang Ming Yi, and David Lo. Automated identification of libraries from vulnerability data: Can we do better? 2022

2022   Giang Nguyen-Truong, Hong Jin Kang, David Lo, **Abhishek Sharma**, Andrew E Santosa, Asankhaya Sharma, and Ang Ming Yi. Hermes: Using commit-issue linking to detect vulnerability-fixing commits. In *Proceedings of the 29th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 2022

2022   **Abhishek Sharma**, Gede Artha Azriadi Prana, Anamika Sawhney, Nachiappan Nagappan, and David Lo. Analyzing offline social engagements: An empirical study of meetup events related to software development. In *Proceedings of the 29th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 2022

2020   Yang Chen, Andrew E Santosa, Ang Ming Yi, **Abhishek Sharma**, Asankhaya Sharma, and David Lo. A machine learning approach for vulnerability curation. In *In IEEE/ACM 17th International Conference on Mining Software Repositories (MSR), ACM*, 2020

2017   Yuan Tian, Ferdian Thung, **Abhishek Sharma**, and David Lo. Apibot: Question answering bot for api documentation. In *Proceedings of the 32nd IEEE/ACM International Conference on Automated Software Engineering*, ASE 2017, pages 153–158. IEEE Press, 2017

2017   Xin Xia, David Lo, Lingfeng Bao, **Abhishek Sharma**, and Shanping Li. Personality and project success: Insights from a large-scale study with professionals. In *Proceedings of the International Conference on Software Maintenance and Evolution (ICSME)*, pages 318–328. IEEE, 2017

2017   **Abhishek Sharma**, Ferdian Thung, Pavneet Singh Kochhar, Agus Sulistya, and David Lo. Cataloging github repositories. In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering*, EASE'17, pages 314–319. ACM, 2017

2017   **Abhishek Sharma**, Yuan Tian, Agus Sulistya, David Lo, and Aiko Fallas Yamashita. Harnessing twitter to support serendipitous learning of developers. In *Proceedings of the 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pages 387–391. IEEE, 2017

2016   Agus Sulistya, **Abhishek Sharma**, and David Lo. Spiteful, one-off, and kind: Predicting customer feedback behavior on twitter. In *2015 IEEE International Conference on International Conference on Social Informatics (SocInfo)*, pages 368–381. Springer, 2016

2016   Doyen Sahoo, **Abhishek Sharma**, Steven CH Hoi, and Peilin Zhao. Temporal kernel descriptors for learning with time-sensitive patterns. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 540–548. SIAM, 2016

2015   **Abhishek Sharma**, Yuan Tian, and David Lo. What's hot in software engineering twitter space? In *Proceedings of the International Conference on Software Maintenance and Evolution (ICSME)*, pages 541–545. IEEE, 2015

2015   **Abhishek Sharma**, Yuan Tian, and David Lo. Nirmal: Automatic identification of software relevant tweets leveraging language model. In *Proceedings of the 22nd International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pages 449–458. IEEE, 2015

AWARDS

2020   ACM SIGSOFT Distinguished Paper Award MSR'2020

2017   Best Paper Award in the ERA track of SANER'2017

2014   SMU Ph.D. Scholarship for 4 years

2008   TCS Initial Learning Program Top Performer

TEACHING EXPERIENCE

Singapore Management University

IS111 - *Introduction to Programming* (Aug-Nov'2018 : Graduate Instructor)

IS110 - *Information Systems and Innovation* (Aug-Dec'2017: Graduate Teaching Assistant)

IS201 - *Object Oriented Application Development* (Jan-April '2016 : Graduate Teaching Assistant)

SERVICE

Program Committee Member

Foundations of Software Engineering (ESEC/FSE) 2023 Industry Forum track
International Conference on Software Engineering(ICSE) 2023 Industry Forum track
Automated Software Engineering (ASE) 2021 Industry Showcase track
Mining Software Repositories (MSR) 2018 Mining Challenge

Peer Reviewer

Transactions on Software Engineering (TSE)
Journal of Software: Evolution and Process (JSEP)

Peer Reviewer (External)

Innovations in Software Engineering Conference (ISEC)
International Symposium on Empirical Software Engineering and Measurement (ESEM)
Evaluation and Assessment in Software Engineering (EASE)
Asia-Pacific Software Engineering Conference (APSEC)
Working Conference on Source Code Analysis and Manipulation (SCAM)
International Workshop on Software Analytics (SWAN)

International Experience

I have worked professionally in India, United Kingdom and Singapore.

Updated February 2025