

# COM6012: **Scalable** Machine Learning

Spring 2026

University of Sheffield

<https://github.com/com6012/ScalableML>



Created by DALL-E

Check-in code:  
**XX-XX-XX**



*It seems like a good idea, but is it scalable?*

[https://s3.amazonaws.com/lowres.cartoonstock.com/animals-scalable-product-mice-cats-slingshot-mdbn347\\_low.jpg](https://s3.amazonaws.com/lowres.cartoonstock.com/animals-scalable-product-mice-cats-slingshot-mdbn347_low.jpg)

# Three Instructors



Shuo Zhou  
Module lead

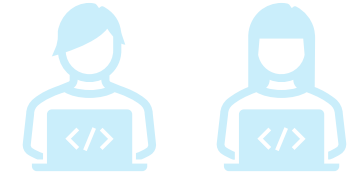


Robert Loftin



Tahsin Khan

# Contents



Week	Topic	Instructor
1	Introduction to Spark and HPC	Shuo Zhou
2	RDD, DataFrame, ML pipeline, & parallelisation	Shuo Zhou
3	Scalable logistic regression	Shuo Zhou
4	Scalable generalised linear models	Shuo Zhou
5	Scalable decision trees	Robert Loftin
6	Scalable matrix factorisation for collaborative filtering (RecSys)	Robert Loftin
7	Scalable K-means clustering	Robert Loftin
8	Scalable PCA for dimensionality reduction	Robert Loftin
9	Scalable neural networks (assessed by exam only)	Tahsin Khan
11	Apache Spark in the cloud (not assessed)	Tahsin Khan



4

# Lectures and Labs @Diamond

Week 11 lab 13:00 – 15:00  
Tuesday, 12<sup>th</sup> May

Lectures @LT5

Monday

13:00–13:52

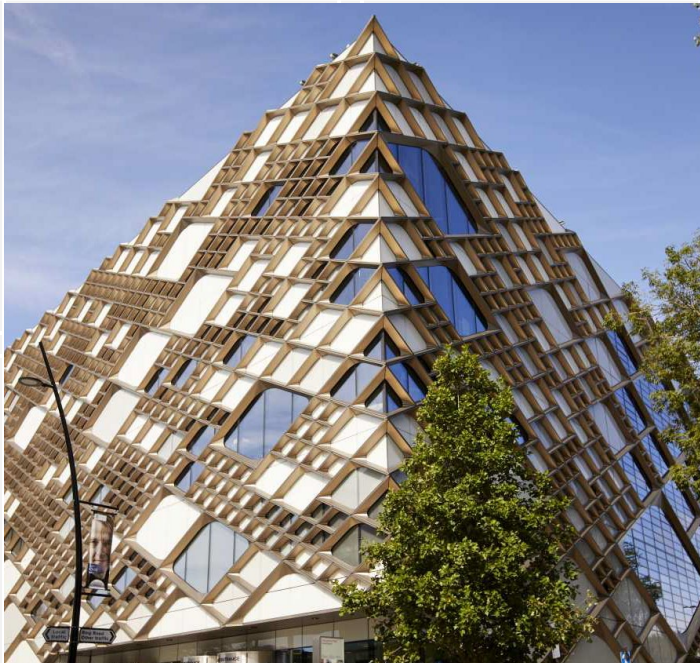
Wednesday (Week 1-9)

11:00–12:52

Labs @Computer Room 1  
(2.01)

14/02/2026

Shuo Zhou - University of Sheffield



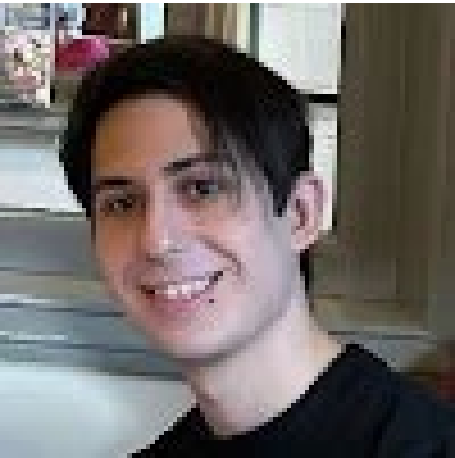


# Labs

- Connect to the HPC cluster for lab tasks
- Bring your laptop (optional)
- Finish required tasks of the labs in this slot!



# Four Demonstrators (GTAs)



Christopher J  
Noroozi (Head)



Xiaolei Xu



Alicja Szwalek



Xiaozhou Tan

# Assessment

- Lab exercises: 0% (self assessment)
  - Finish lab exercises by the following Mondays
  - Solutions to release on the following Tuesdays
- HPC driving licence (required by IT services)
- **Assignment: 40%**
  - Progressive release, to be handed out by 27th March
  - Deadline: 13:00 on Wednesday, **6th May** (end of lab 10)
  - **Avoid academic misconduct**
  - Solution release: 22nd May
  - Marking and feedback deadline: 29th May
- **Exam (on Blackboard): 60%**

# Additional Support

- Blackboard discussion board
  - Get help on lecture/lab contents
  - To ask for **clarification** on assignment **questions** (i.e. the tasks to do)
  - **NOT** to ask how to solve the problems.
  - **NOT** to ask for the correctness of a specific solution or share a possible solution.
- Direct email to instructors: **personal/private issues only**
- Instructor office hour: Wednesday 4:15-5:00 pm @G25 Regent Court
- Demonstrator online support (Week 2-9 ): Friday (Time TBC) via Discord