ALGOSOC

UOB

DATA SCIENCE & MACHINE LEARNING WORKSHOP

Week 1 - Intro

MEET THE PEOPLE

ALGOSOC UOB

We will be helping out! 😄

Louis - Focus on DS&ML workshop

Ayush - Focus on LeetCode & Chill

As well as several other committee from Algosoc will be available to provide support

STRUCTURE?

There are three main parts of this workshop:

1. Live sessions: 1 hour, concept, ideas, and case studies
2. GitHub repository: Code tutorials and examples
3. Discord: For discussions and ~~Yap~~ questions

Repo:

github.com/algosoc/Data-Science

The live sessions are sadly (mainly) not a coding tutorial.

Feel free to provide any inputs on the discord chat

AGENDA

Week 1 topic:
- Data, Data Science, Data Scientist?
- Data Science Workflow
- Problems to be solved with Data Science
- Forming a problem
- Intro to Python and Notebooks

We are also planning to have a Quantitative Finance workshop and a Datathon next semester!

Full agenda this semester:
https://bit.ly/DataScienceAlgosoc

Please look at the agenda, where we will be giving surveys at the end regarding our workshop progression.

DATA, DATA SCIENCE?

What comes to mind when you hear the word *data*?

One might be tempted to associate data with numbers and measurements, but those are not the complete example of data.

If you own a phone:

- Scrolling behaviour
- Time spent on a post
- Most visited app
- Messages you’ve sent to your friends
- Places you have visited

If you use a smartwatch:

- Walking patterns
- Biological data (heart rate and oxygen levels)

If you have been in a survey :

- Your opinions on a matter
- Your behaviour during the survey

and many more...

DATA, DATA SCIENCE?

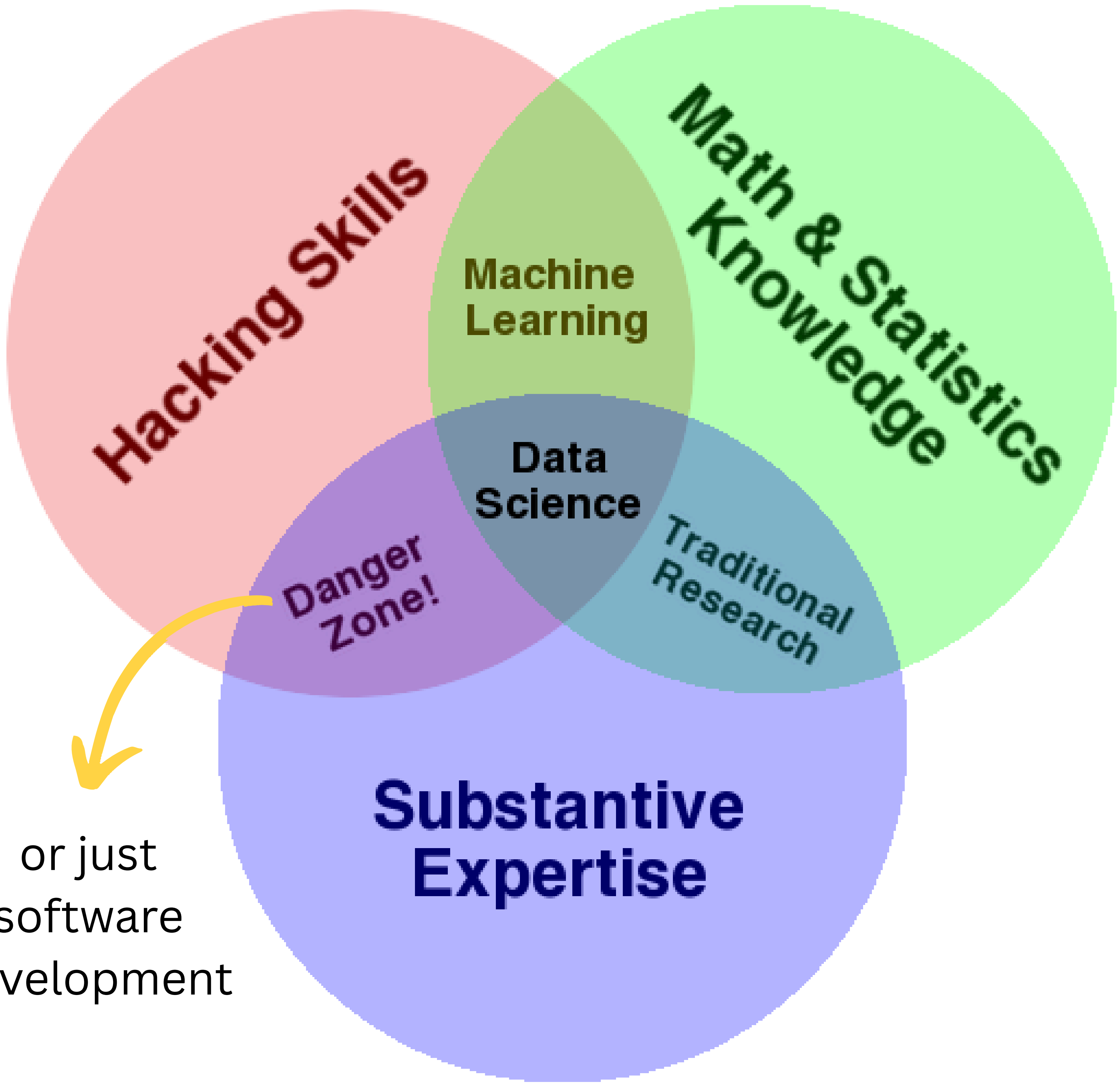
“information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer:” – **Data**, defined in Cambridge Dictionary

What can data be used for?

- Providing recommendations on instagram
- Feeding knowledge to generative AI
- Making business and marketing decisions
- Algorithmic trading
- and many more...

With the vast amount of **use** they can have, **sources** they can come from, or **changes** they can undergo, they are indeed in need of a science explaining them. Hence, ***data science***

DATA, DATA SCIENCE?



or just software development

According to Andrew Conway (2010), Data Science is a combination of hacking skills (programming), substantive expertise (domain knowledge), and maths and statistics.

What differentiate data science from machine learning is the domain knowledge being applied into solving the problem. Hence, we **need a problem to be solved.**

DATA SCIENTIST?

What is a data scientist?

If we are using Conway's definition on data science, then a data scientist is someone who is good at maths and statistics, knows how to code, and can apply both skills into a specific domain of problem that requires domain knowledge

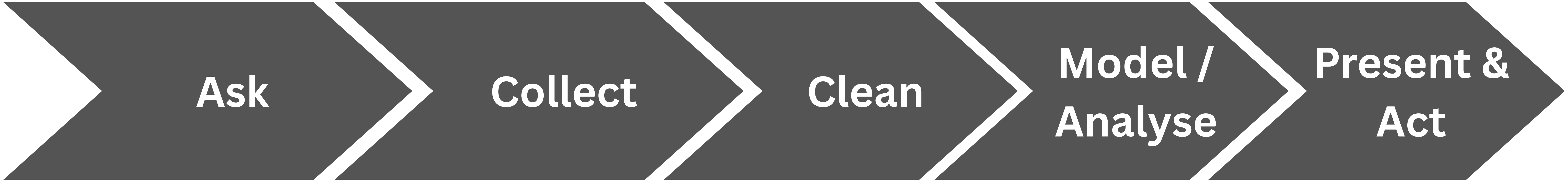
What do they actually do?

Depends on where you are...

- Research lab: analysing data, performing R&D
- Marketing company/consultant: giving presentation
- Tech company: software development, integrating technologies with DS methods

But it is not limited to these, it all depends on your current circumstances

DATA SCIENCE WORKFLOW



The general workflow of a data scientist:

1. Ask the right questions to formulate the correct problems
2. Use the formulated problems as guidance to collect the correct data
3. Process and clean the data, most of the time they will be messy
4. Make a model out of it or analyse the data, this is where you will have insights
5. Create a presentation that **concludes your finding**, where it **solves the defined problem** and act upon it for your organisation.

PROBLEMS TO BE SOLVED

Several problems to be solved by data science methods would be:

- Getting customer insights (e.g., Netflix, retail stores) → Reduce churn rate through targeted marketing
- Understanding economic, political, and social situation → Policy-making
- Getting public sentiment on certain topics → Favourable political act

When there are decisions to be made, data science can be used to support decision making.

What else...?

FRIST, DEFINE THE PROBLEM

The first important step to tackling a data science problem is to... **Define the problem!**

What does your organisation/business need?

1. Understand the business processes
2. Talk to stakeholders and find a grip on what they are/should be looking for
3. Define the core objectives

Then, turn it into a data problem, where you would need support from data to make a decision: See what types of data you would need

Finally, after getting all the points, define the problem statement!

FRIST, DEFINE THE PROBLEM

Example walkthrough on a telecommunication (sim card) company:

1. **Business need**: “Our organisation needs to reduce customer churn.”
2. **Understand the business process**: Discuss with stakeholders how products and services are marketed and delivered, (are customer behaviour, usage frequency, or complaints tracked?).
3. **Identify data sources**: Get a grip on what data are available or can be collected (demographics, usage patterns, recharge frequency, and subscription duration).
4. **Translate into a data problem**: Determine how data can support the decision-making process (e.g., identifying key factors contributing to churn and predicting which customers are at risk).
5. **Final problem statement**: “Use data-driven analysis to understand key drivers of customer churn and predict which customers are likely to leave, supporting retention efforts.”

Can you give another walkthrough example?

CODING TUTORIAL

Throughout the semester, we will be using several packages from Python, but not limited to the ones mentioned in this slide:

Link to today’s practical:
https://bit.ly/AlgosocDSWk1

Once you got in, please make a copy of the notebook

FEEDBACKS

We would need feedbacks for future progression, please access this:

go to: **https://app.sli.do** and

enter code: **1850223**