Problem solving with Python

aka MSAN 689

by Laurynas Riliskis

Hello!

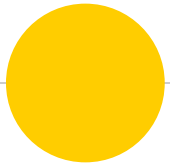
I am Laurynas Riliskis

Ph.D. in CS LTU, Sweden

Postdoc from Stanford University, USA

Security + IoT = a lot of problems

I am here because I love solving problems.



Essentials

& other stuff

Please be on time for class. It is a big distraction if you come in late.

Put your phone in silent.

This is data

- Class is held at 101 Howard in 5th floor classroom 529 until Friday 29th June.
- 5 homework's deadline 'next' Friday 9AM
- 5 quizzes: 9-9:30 am Friday's room 150
- S1: Fri 10-11:55 AM or S2: Fri 2:20-4:15PM
- Class runs for 1:50 hours, 1 day/week.
- All programming will be done in the Python 3 unless otherwise specified.

Evaluation

- 70% will come from 5 homework where you will have to solve harder problems.
- 30% will come from 5 quizzes before each class.

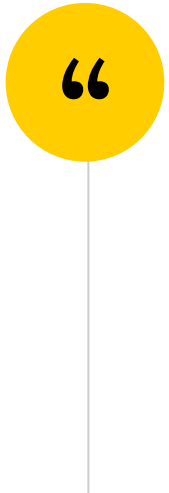
Syllabus

1. Designing solution: readable code and OO
2. Data structures and analyses
3. Problem understanding and solving paradigms
4. String and math problems
5. Geometry and graph problems
6. Guest talk: Rob Munro, CTO Figure Eight, Human-in-the-Loop platform

Lecture 1

Designing solution: readable code and OO

Design your solution



Designing solution

Designing solution

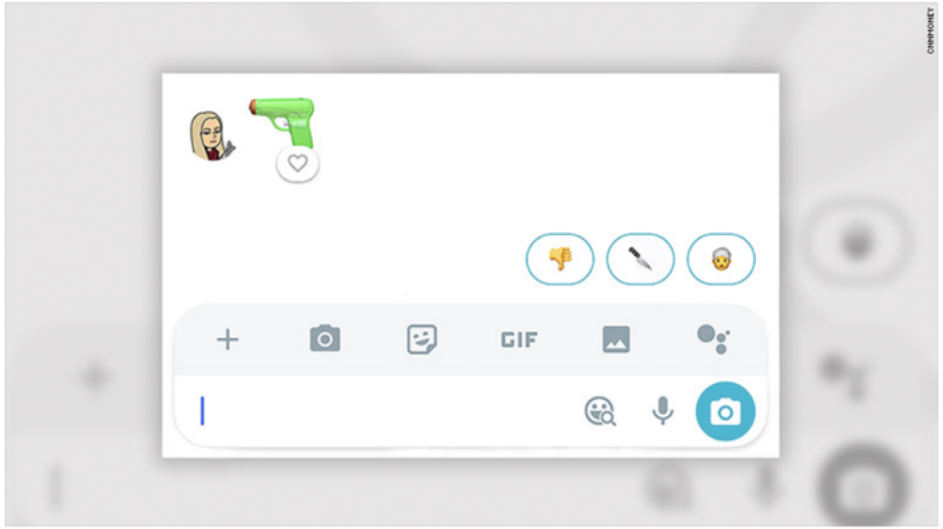
Data science in real world

- Direct business value
- Learn from less data
- Protect data and insights
- Traceable, explainable, & fair

Traceable, explainable, & fair

- Shit goes wrong!

Google Allo responds to a gun emoji with a turban emoji



A CNN staff member received an emoji suggestion of a person wearing a turban via Google Allo. This was triggered in response to an emoji that included a pistol. An embarrassed Google assured the public that it had addressed the problem and issued an apology.

Google AI looks at rifles and sees helicopters


rifle_adv.png

Helicopter	78%
Rotorcraft	66%
Aircraft	56%
Vehicle	53%

Careful what you ask Alexa for, you might get it

The Amazon Alexa virtual assistant can make online shopping easier. Maybe too easy? In January, San Diego news channel CW6 reported that a six-year-old girl had purchased a US$170 dollhouse by simply asking Alexa for one. That's not all. When the on-air TV anchor repeated the girl's words, saying, "I love the little girl saying, 'Alexa order me a dollhouse,'" Alexa devices in some viewers' homes were again triggered to order dollhouses.

Traceable, explainable, & fair

- Shit goes wrong!
- Enable human users to understand, appropriately trust, and effectively manage

Chinese facial recognition study predicts convicts but shows bias

vo researchers at China's Shanghai Jiao Tong University published a s
computer and set about analysing them." In the work, the researchers
any in the field questioned the results and the report's ethics underp

Robot injured a child

so-called "crime fighting robot," created by the platform Knightscop
cident was a " freakish accident."

AI built to predict future crime was racist

e company Northpointe built an AI system designed to predict the ch
ork by Philip K. Dick), was accused of engaging in racial bias, as black c
ftware wasn't an "effective predictor in general, regardless of race."

AI-judged beauty contest is racist

"The First International Beauty Contest Judged by Artificial Intelliger
iling to supply the AI with a diverse training set, the contest winners

Insurance company uses Facebook data to issue rates, shows bias

l, finally, this year England's largest vehicle insurer, Admiral Insurance, set out to use Fac

ile this isn't a straight AI failure, it is a misuse of AI. W
ng its policy that states companies can't "use data obt

evidenced by these examples, AI systems are deeply p
ndards for research is of utmost importance.

Fatality in Tesla Autopilot mode

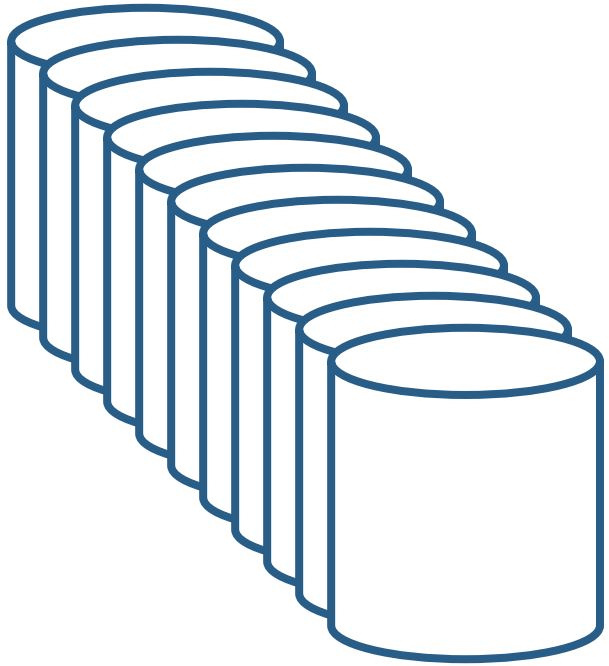
previously reported by TechRepublic, Joshua Brown was driving a Tesla
nce the accident, Telsa has announced major upgrades to its Autopilot so
ina, although none can be directly tied to a failure of the AI system.

Traceable, explainable, & fair

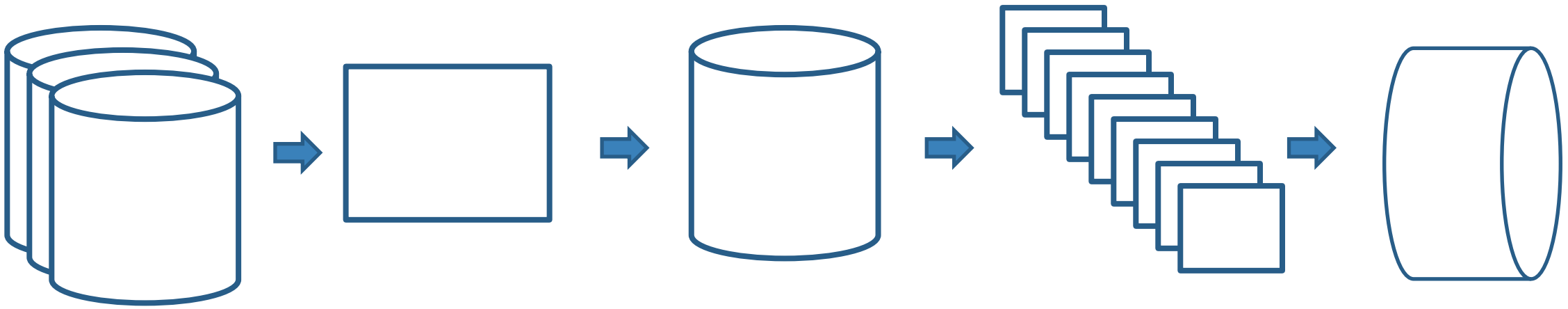
- Shit goes wrong!
- Enable human users to understand, appropriately trust, and effectively manage
- Explain rationale, characterize strengths and weaknesses, and convey an understanding of how it behave in the future

Designing solution

? +



Designing solution



Data sources

Cleaning

Clean data

Data partition

Modeling

Designing solution

Solving a data science problems goes beyond the finding the answers.

- clarity
- well-documented code
- readability of the implementation
- visibility of the of the solution design

Designing solution

- how does the data flow through the app?
- what blocks of code?
- what they will talk to?
- what is the output?

how does the data flow

- Identify data sources
  - Consolidate
  - Storage
- Provision – how are you getting? File, API,..
- Cleaning
- Sampling/features
- Training/Partitioning, etc

Designing solution

- how the data flow through the app?
- what blocks of code?

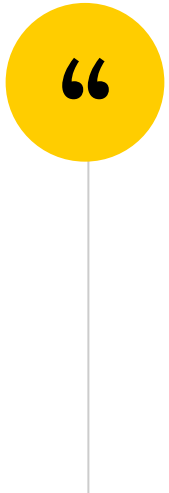
Designing solution

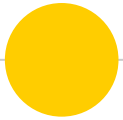
- how the data flow through the app?
- what blocks of code?
- what they will talk to?

Designing solution

- how the data flow through the app?
- what blocks of code?
- what they will talk to?
- what is the output?

Write good code

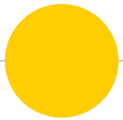




Examples

- jupyter

To OO or not to OO



OO for DS

- No, for the most parts

OO for DS

- Sequential programing
  - Typical pipelines
  - Performance > maintainability
  - Small code base

OO for DS

- BUT, you use libraries
  - pandas, numpy, and scikit-learn all rely on classes, methods and attributes == OO heavy
  - Enterprises have own code base
- And
  - DRY principle – Don't Repeat Yourself.

Object Oriented Programming

- Scalable, sharable code for data analysis
- Encapsulation
- Inheritance

Object Oriented Programming

- Why: easy to scale, test, and maintain
- Thinking of code in terms of object
- For the specific entity, object encapsulates
  - Data
  - Attributes
  - Methods

Object Oriented Programming

- Creating uniform interfaces through OOD principles can help us streamline downstream operations.

Potential opportunities

- Repeating yourself with slight tweaks to accommodate differences between data sources
  - Just different URLs, database connection strings, or file names? A well-parameterized function will probably suit you just fine.
  - Finding yourself writing a lot of ‘if’ statements in your extraction code? Probably time to refactor and consider using classes

Potential opportunities

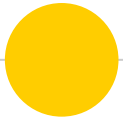
- Finding similar functions across vertical stacks
  - We looked across the process used for working with different data sources and identified similar functions for cleaning and acquiring data. If you can identify similarities like this in your code OOD may help.

Object Oriented Programming

- Creating uniform interfaces through OOD principles can help us streamline downstream operations.

Object Oriented Programming

- Why: easy to scale, test, and maintain
- Thinking of code in terms of object
- For the specific entity, object encapsulates
  - Data
  - Attributes
  - Methods



Prepare for quiz

- Next quiz: June 1, 9:30 AM!
- Good coding practice
- OO

Homework reHash

- Revisit Search Engine Implementation
- https://github.com/parrt/msan692/blob/master/hw/search.md

The end

- Design solution: data flow
  - Input, code blocks, output
- Clean code
- To OO or not to OO

That's it for today