What good data science looks like

Chris Beeley

May 23, 2023

Patient experience

- The NHS collects a lot of patient experience data
- Rate the service 1-5 (Very poor… Excellent) but also give written feedback
  - “Parking was difficult”
  - “Doctor was rude”
  - “You saved my life”
- Many organisations lack the staffing to read all of the feedback in a systematic way
- Produce an algorithm to rate theme and “criticality”

Help people to do their jobs

- Text based data is *complex* and built on *human experience*
- The tool should enhance, not replace, human understanding
- Enhancing search and filtering
  - If they read 100 comments today, which should they read?
- “A recommendation engine for feedback data”

Reflect what users want

- I have worked with this data since before it existed
- I came to realise that people were struggling to read all of their data
- Fits alongside other work happening within NHSE
  - A framework for understanding patient experience

Useful

- A fundamental principle is that everyone can use
- If you can run the code, run it
- If you can use the API, use it
- If you just want the dashboard, use it
- Credit to the growth charts API

Understandable

- Tuned to the users needs
- Not simply tuning accuracy scores
- Look at the type of mistake the model is making
- Look at the category it’s predicting
  - We can lose a few of common unimportant categories
  - We need to get every rare and important category

Iterative

- Year one
  - 10 categories
  - Moderate criticality performance
  - No deep learning
  - Weak dashboard
  - Positive evaluation

Iterative

- Year two
  - 30-50 categories
  - Strong criticality performance
  - Deep learning
  - Improved dashboard
  - WIP
- Overall five minor versions of algorithm and seven of dashboard

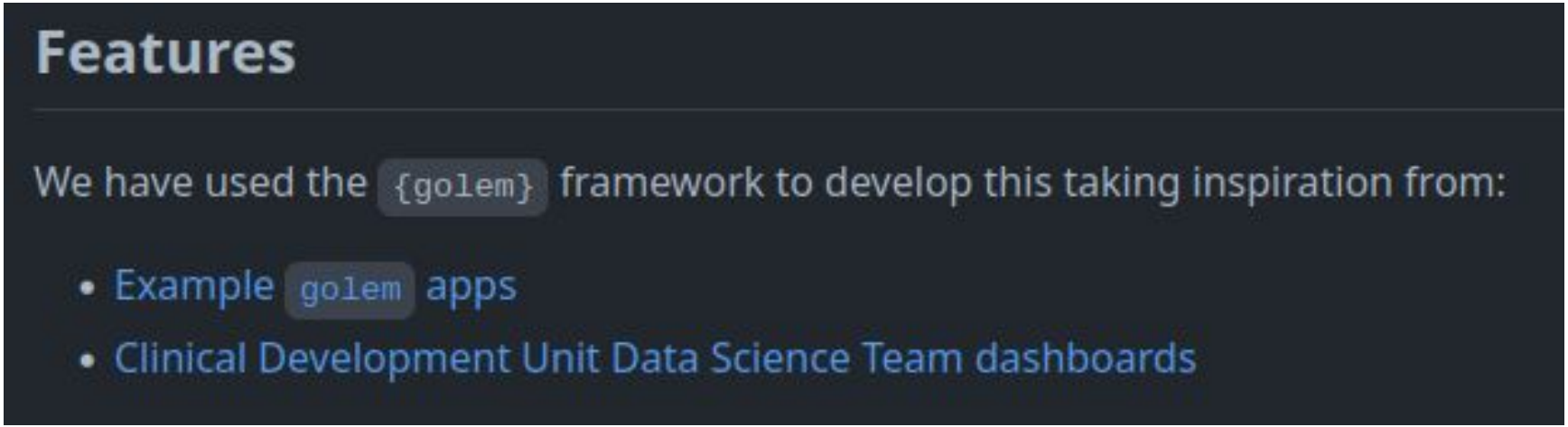
Documented

- We’ve documented in the way you usually would
- We were asked in year 1 to provide plain English documentation
- We made a website with all the product details

Develop skills of the staff, technical and otherwise

- Year one created a Python programmer
- Year two created an R/ Shiny programmer
- The team has learned:
  - Static website generation
  - Text cleaning/ searching/ mining
  - Collaborative coding practices
  - Working with and communicating with users
  - Linux, databases, APIs…

Benefits from, and benefits, the community



NHSBSA R Shiny template

Benefits from, and benefits, the community

- We benefit and benefit from
  - NHS-R
  - NHS-Pycom
  - Government Digital Service
  - Colleagues and friends

Open and reproducible

- Off the shelf, proprietary data collection systems dominate
- They often offer bundled analytic products of low quality
- The DS time can’t and doesn’t want to offer a complete data system
- How can we best contribute to improving patient experience for patients in the NHS?
  - If the patient experience data won’t come to the mountain…

Open source FTW!

- Often individuals in the NHS don’t want private companies to “benefit” from open code
- But if they make their products better with open code the patients win
- Best practice as code

Fun!

- Combing through spreadsheets looking for one comment is not fun
- Doing things the same way you did them last year is not fun
- Trying to implement a project that is too complicated is not fun
- Working with a diverse team with different skills is fun
- Accessing high quality documentation to understand a project better is fun*

Team and code

- Andreas Soteriades (Y1)
- YiWen Hon, Oluwasegun Apejoye (Y2)
- pxtextmining
- experiencesdashboard
- Documentation
- chris.beeley1@nhs.net
- https://fosstodon.org/@chrisbeeley