NHSX PhD Data Science Interns

HDR UK Seminar

October 2021

Agenda

Overview of the Scheme - Jonny

Project 1: Variational AutoEncoder for Healthcare Data - Dom

Project 2: NHS Text Data Exploration - Dan

Project 3: SynPath – Diabetes module - Tiyi

Project 4: Machine learning and shopping data - Lizzy

General Q&A - All

The Scheme

This scheme aims to connect NHS real data and issues with academic thought and research through short-term PhD internships. The benefit to the NHS is the **added value** that academia brings to evidence-based research but on timescales that allow the insights to be acted upon. For the student and university, the benefit is an avenue to conduct related **research in an industrial environment** and access to NHS data (although try to use open data where possible as learning is then sharable).

We are looking for PhD students working in a quantitative discipline but with an interest in applying their knowledge and gaining experience of creating solutions for the NHS. **Both** student and university will be funded during the duration of the internship covering between three and six months (Student - Agenda for Change Band 6 equivalent; University - £5k).

Our aim is to continually build on previous learning whilst having an avenue for including the latest research and approaches. Where possible we will work in an **open and transparent way** ensuring that learning is shared and insights made available for others to reproduce. At the end of the project the applicant will submit a final **report** suitable for publication in open literature, and **presentations** to NHSX on their results including their experience of the project. The nature of the output depends on the project.

<u>Key priorities of the internship will be:</u>

- Safe and appropriate use of NHS data
- To producing a balanced outcome for both the student and NHSX, with useable outputs
- To provide the student(s) with experience of completing a live business project
- To provide the NHS with an avenue to experience current research and ideas
- To kick-start or accelerate current projects and ideas
- To build a long term programme with developing research areas

The Role

The successful candidate will occupy a permanent role within the NHSX analytics unit for **between 3 and 5 months**. During the internship the candidate will be expected to progress their chosen project **autonomously with supervision** from colleagues in the analytics unit as well as their current academic supervisor. The candidate will be expected to be focussed on the project and self-driven during the internship period, providing regular updates on progress and issues.

Within the topic area of the chosen project, there will be **some freedom** to direct the development of the research and knowledge but this will also need to be balanced against creating a learning outcome or **tangible output**, that benefits the PhD scheme objectives and is in a shareable state for future projects to pick up and continue the development.

As the candidate will be a NHSX employee during the internship period, standard employment checks and some mandatory training are required. We are currently working entirely remotely for these projects.

<u>Key aims for the intern will be:</u>

- Progress the research project ensuring **learning/outcomes are shareable** with NHSX and where appropriate made suitable for public release of learning and code.

The Advert

The second and third waves of this internship are set to start in **January and June 2022** respectively. Each wave can accommodate up to three students working on 3 - 5 months projects.

Applications will **open early October** with interviews in late November and outcomes communicated in early December. Both the **January and June cohorts will be recruited at the same time**.

Application details will be available on our NHSX Analytics Unit site:
https://www.nhsx.nhs.uk/key-tools-and-info/nhsx-analytics-unit/nhsx-internship-scheme-innovation-and-analytics-health/

We use the “Be Applied” software for our recruitment to ensure impartiality. Applicants are sifted and invited to a competency-based interview (questions relate to broad clear topics with set scoring). Successful candidates will then need to provide two references prior to starting work.

We ask that any applicant highlights which currently available projects (see later slide) they would be interested in working on and how they would go about the project. Whilst the projects need to be agreed prior to start (so we can support appropriately and connect the work to business priorities), there is a lot of flexibility in the project direction and we are also open to discussing project proposals prior to application - please get in contact with analytics-unit@nhsx.nhs.uk to discuss project ideas or for any queries.

About Us

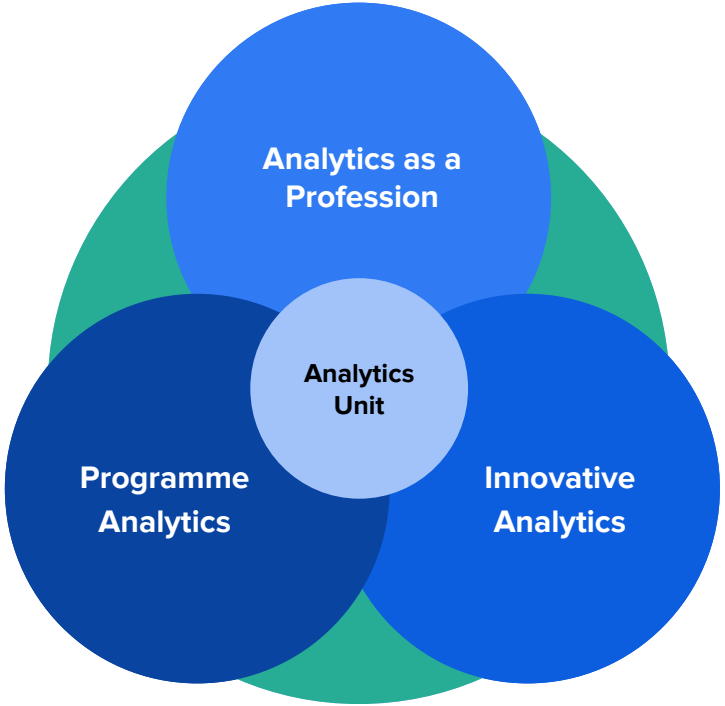
An introduction to NHSX

NHSX is a joint unit bringing together teams from NHS England and NHS Improvement, and the Department of Health and Social Care to **drive the digital transformation of health and care**

We are a diverse team with a range of skills and expertise, including clinicians, technologists, policy experts, developers, data scientists and project managers

Overview of AU

- The Unit aims to **modernise and strengthen the use of data and analytics** across the health and social care system and **support digital transformation**.
- The Unit consists of analysts, economists, data scientists, data engineers, data developers and a comms manager that sit across **three workstreams**.



Innovation Analytics

Aims to:

1. Support the development and sharing of new impactful techniques
2. Be a technical lead for collaboration and procurement of data science work
3. Lead by demonstration in working in an open and transparent way

Objectives:

- Develop and share innovative analytical projects of value to the health and care sector.

- Collect, generate and advise on best practice for conducting and procuring innovative analytics within the health and care sector in a valid, transparent and interoperable way

See our github https://github.com/nhsx/ or project website https://nhsx.github.io/AnalyticsUnit/ for open projects

“Creativity is thinking up new things, innovation is doing new things”

What makes us unique?

Our unique selling point is a combination of two parts:

Advanced Analytics

We have the remit to investigate applying novel techniques and cutting edge algorithms

Practically this means we aim to:

- Connect techniques to questions
- Connect the analysts to support
- Connect products to users
- Connect technological knowledge to systems

Transparency

We have a remit to be open regarding methods and have the aim of sharing products with the health system

This means we aim to overcome issues of:

- Sharing information
- Deploying approaches
- Distributing learning from siloed models
- Providing useable exemplar case studies

Synthetic Data Generation | **Model Interoperability** | **AI Validation and Explainability** | *Future uses of Bayesian Inference* | **Graph Structures and techniques**

Deploying Machine Learning | **System Modelling** | **Applied NLP** | **Open-source analysis**

Five year aims

Digital Transformation is the core focus for NHSX - for Innovative Analytics, what does that mean?

Enabling others through data access:

- Text data to be as common a data source for analysis as structured EHR
- All NHS analysts to be able to access necessary data through accessible platform/environment or synthetic placeholders

Enabling others through culture:

- Use of open ways of coding and sharing of code commonplace across national and local analysts
- Analysis and tools created with view on interoperability to enable reuse across different use-cases

Enabling others through example:

- Showcasing innovative techniques in areas of active interest
- Demonstrating use of frameworks and tools for sharing work

Enabling others through guidance:

- Identifying appropriate technology choices to enable modern analytical approaches to be used
- Openly discuss approaches (both benefits and limitations)

Two year outlook

Core interests:

- Natural Language Processing - what is the future of this in the NHS? Increasing use of text as evidence
- Graphs and Graph Techniques - build out community contacts? What is being done already?
- Privacy - federation, privacy preserving approaches - watch and brief?
- Emerging techniques - identify early in academia/industry and showcase…
- Synthetic data generation - shareable, open datasets and tools
- System Modelling - Intelligent representations of the reality behind the data

Open working:

- Grow and maintain momentum around open tools and analysis - links with **NHS-R** and **NHS PyCom**
- Work closely with open lead, standards and interop, and wider CTO function within NHSX
- Actively seek to engage with TREs, federated learning frameworks and off-server analysis (e.g. Open Safely work)
- Supporting and using frameworks for transferable tools and knowledge

Supporting NHSX (+ NHSEI & DHSC) Teams:

- Practical support for AI Lab (AI in Imaging and Skunkworks)
- Practical support for Innovation Lab
- Support and guidance for the Data Science communities in NHSEI
- SME for Centre for Increasing Data Collaboration

PhD Internship Programme next steps/scaling:

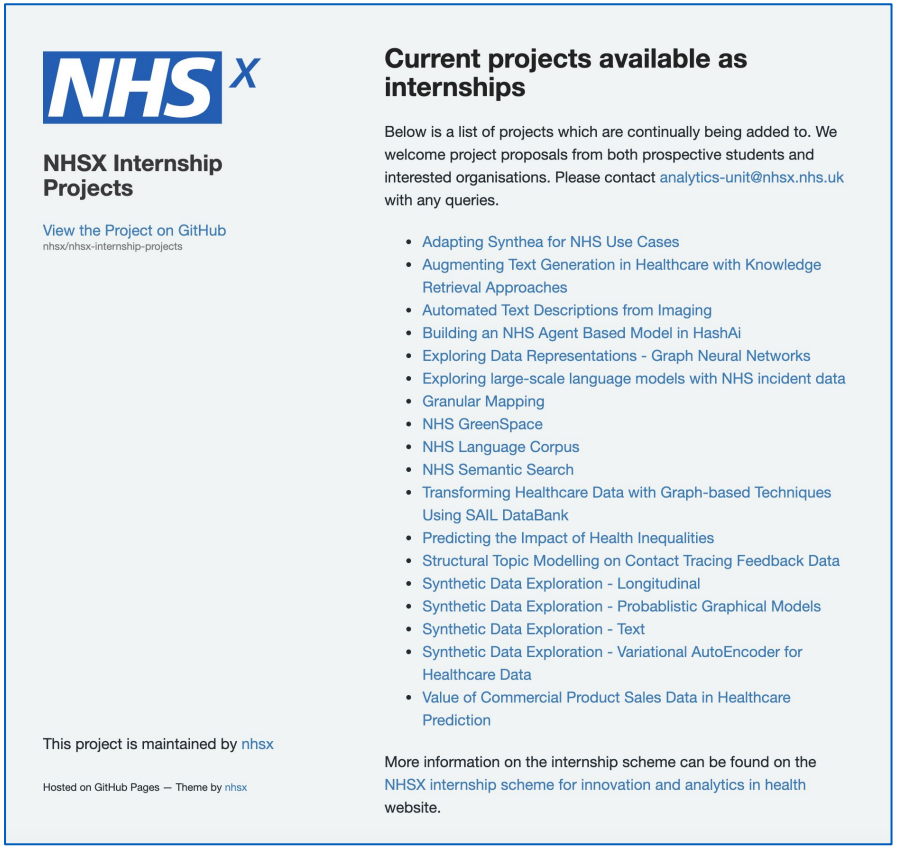
- 3-5 month internships for PhD students to work on NHS related innovation projects

Available Projects

Available Projects

To see all the current available projects visit https://nhsx.github.io/nhsx-internship-projects/

Please contact us to ask about projects, discuss amendments to align these to your area of interest, or propose new projects: analytics-unit@nhsx.nhs.uk



NHSX Internship Projects

View the Project on GitHub
nhsx/nhsx-internship-projects

This project is maintained by nhsx

Hosted on GitHub Pages — Theme by nhsx

Current projects available as internships

Below is a list of projects which are continually being added to. We welcome project proposals from both prospective students and interested organisations. Please contact analytics-unit@nhsx.nhs.uk with any queries.

- Adapting Synthea for NHS Use Cases
- Augmenting Text Generation in Healthcare with Knowledge Retrieval Approaches
- Automated Text Descriptions from Imaging
- Building an NHS Agent Based Model in HashAi
- Exploring Data Representations - Graph Neural Networks
- Exploring large-scale language models with NHS incident data
- Granular Mapping
- NHS GreenSpace
- NHS Language Corpus
- NHS Semantic Search
- Transforming Healthcare Data with Graph-based Techniques Using SAIL DataBank
- Predicting the Impact of Health Inequalities
- Structural Topic Modelling on Contact Tracing Feedback Data
- Synthetic Data Exploration - Longitudinal
- Synthetic Data Exploration - Probablistic Graphical Models
- Synthetic Data Exploration - Text
- Synthetic Data Exploration - Variational AutoEncoder for Healthcare Data
- Value of Commercial Product Sales Data in Healthcare Prediction

More information on the internship scheme can be found on the NHSX internship scheme for innovation and analytics in health website.

Available Projects

Transforming Healthcare Data with Graph-based Techniques Using SAIL DataBank

Keywords: Graph Structures, Data Representation, Hypergraphs

Need: The SAIL Databank - The Secure Anonymised Information Linkage Databank provides anonymised person-based data for research powered by the Secure e-Research Platform. Recent work using graph models to build simpler knowledge discovery systems opens up potential for increased prediction accuracies, reduced pre-processing burden and the application of models of higher complexity to our data. These models have been shown to effectively handle messy data and to learn representation of key factors from the data directly (rather than choosing a set of predictor variables).

Current Knowledge/Examples & Possible Techniques/Approaches: The recent paper entitled Ranking Sets of Morbidities using Hypergraph Centrality demonstrates applying a hypergraph analysis to a comorbidity question. Where possible, this project would seek to develop upon this work looking at including further complexity through adding demographics, direction into the graphs, or time evolution. A likely first step to this would be a comparison of hypergraphs application versus state-models and more traditional clustering approaches.

Outcome/Learning Objectives: A comparison of hypergraphs applied to SAIL data versus state-models and clustering techniques to demonstrate the value of the graph application. An extension of this would be to add a single demographic variable to each technique and show the potential of these additional data.

Datasets: SAIL Databank Datasets

Desired skill set: When applying please highlight any experience around graph databases and representations, GNNs, Multi-State models or Clustering Algorithms, coding experience, and any other data science experience you feel relevant.

Available Projects

Automated Text Descriptions from Imaging

Keywords: Synthetic, Imaging, Semantic Explainability

Need: Medical imaging is still only reaching a low amount of its potential opportunity. Image collections have large amounts of variation in both the images themselves (e.g. different machines having different angles and contrasts) and the associated reports (for the most part captured in free text). This project would aim to explore advances in machine learning and explainability to take advantage of the relationship between these two data modalities - utilising techniques to look at automating text descriptions from images.

Current Knowledge/Examples & Possible Techniques/Approaches: For a wider review of explainability in medical imaging including semantic explainability, see for example - Explainable deep learning models in medical image analysis.

Outcome/Learning Objectives: Open worked example and explanation of current state-of-the-art approaches to be built on, or used by others.

Datasets: Suitable multi-modal healthcare datasets such as MIMIC-CXR.

Desired skill set: When applying please highlight any experience around work with imaging or text data and specifically medical imaging data, tagging, explainability in ML, coding experience (including any coding in the open), and any other data science experience you feel relevant.

Previous and Ongoing Projects

Variational AutoEncoder for Healthcare Data

Project Link: https://github.com/nhsx/SynthVAE

Separate slide pack

NHS Text Data Exploration

Project Link: https://nhsx.github.io/nhsx-internship-projects/nhs-language-corpus/ (Updated outputs coming in October)

Beth Rushton-Woods

Summer Project for Masters in Data Science, Lancaster University

Supervisor: Prof. Paul Rayson

NHS Language Corpus - Vision

The **NHS Language Corpus** would look to be:

Open

Important to make this resource easily available to innovators and researchers in NLP healthcare space

Representative

Contains a range of sources and variability in language used in a given setting

Extensible

Collect a dataset that has a wide coverage as well as a large number of examples over time

~~Unique~~ Useful

Adds to the currently available resources constructively

NHS.UK Text Data

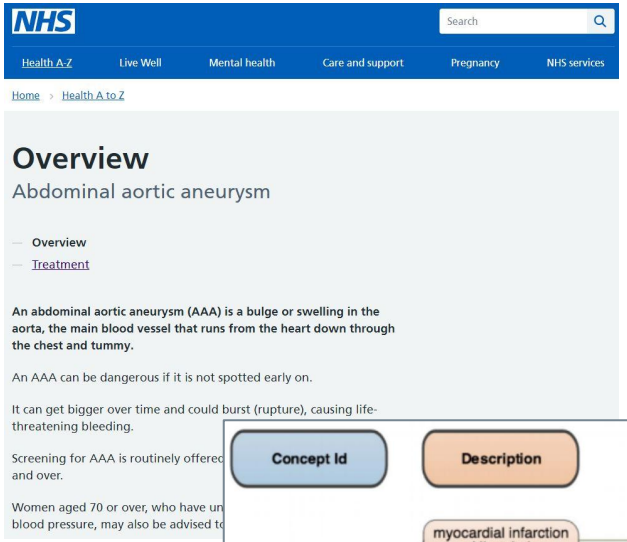
The data for the project was webpages from **NHS.UK**, in particular, the **conditions pages**:

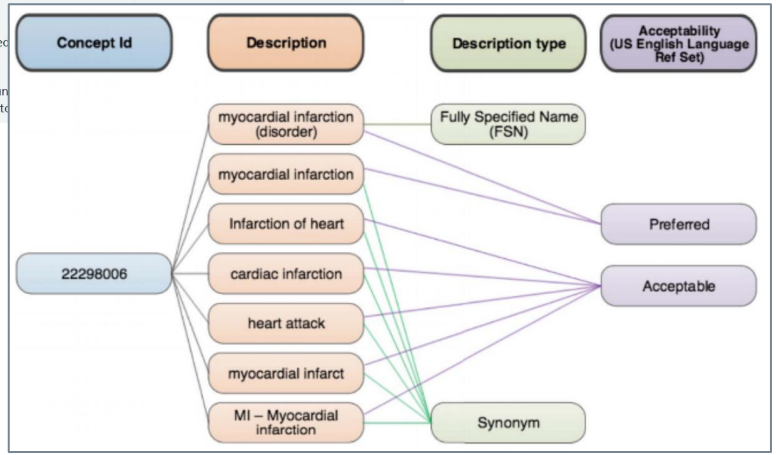
2302 pages in total, made up of ~1.43 million tokens, of which around 43k unique.

The goal was to explore frameworks and packages which would allow us to enrich the text with relevant SNOMED-CT codes*.

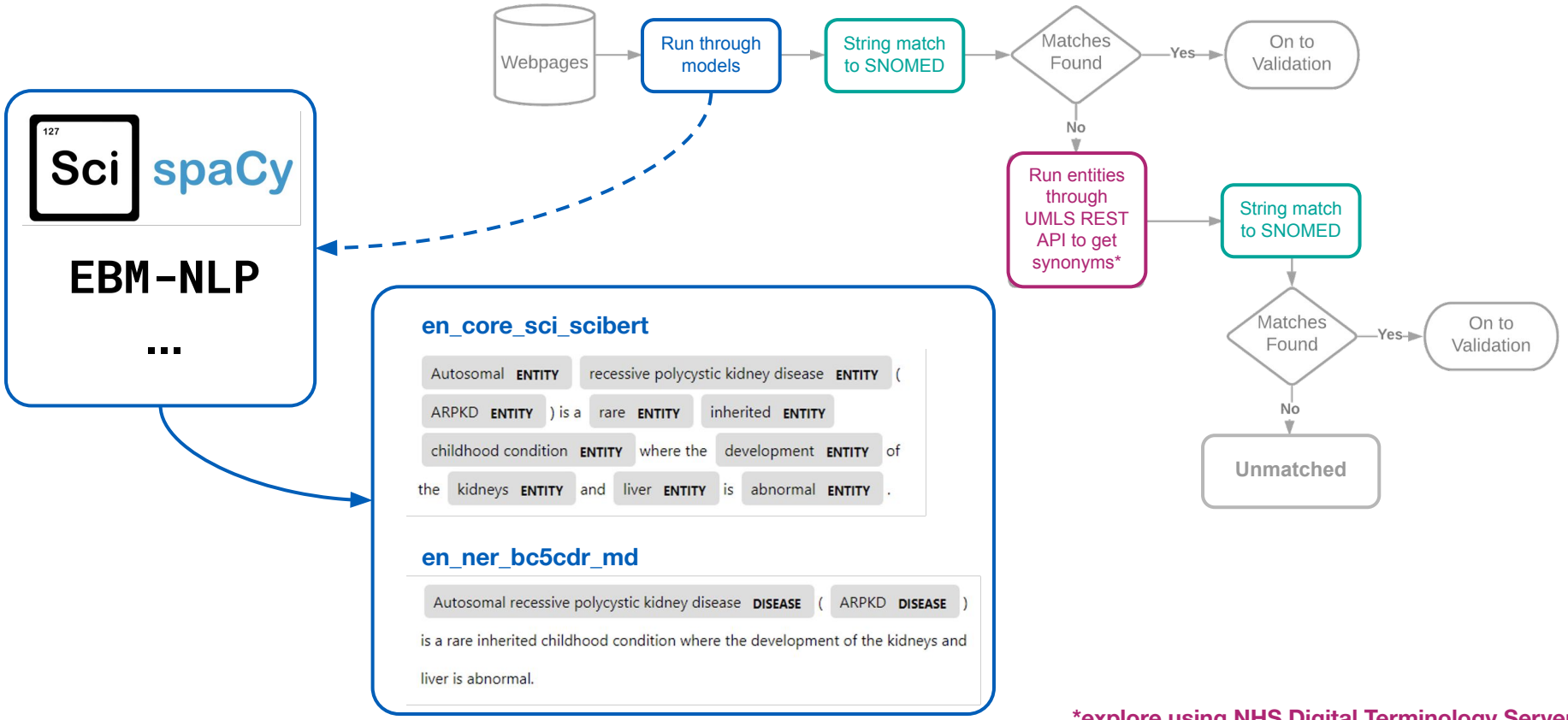
The language is purposefully written to be accessible to the general public, and so is not always technical, an issue when using tools which were trained on more technical biomedical texts.

***SNOMED-CT** is a commonly used clinical coding system, used to translate biomedical text into standardised concept ID codes, adding useful structure to the text.*



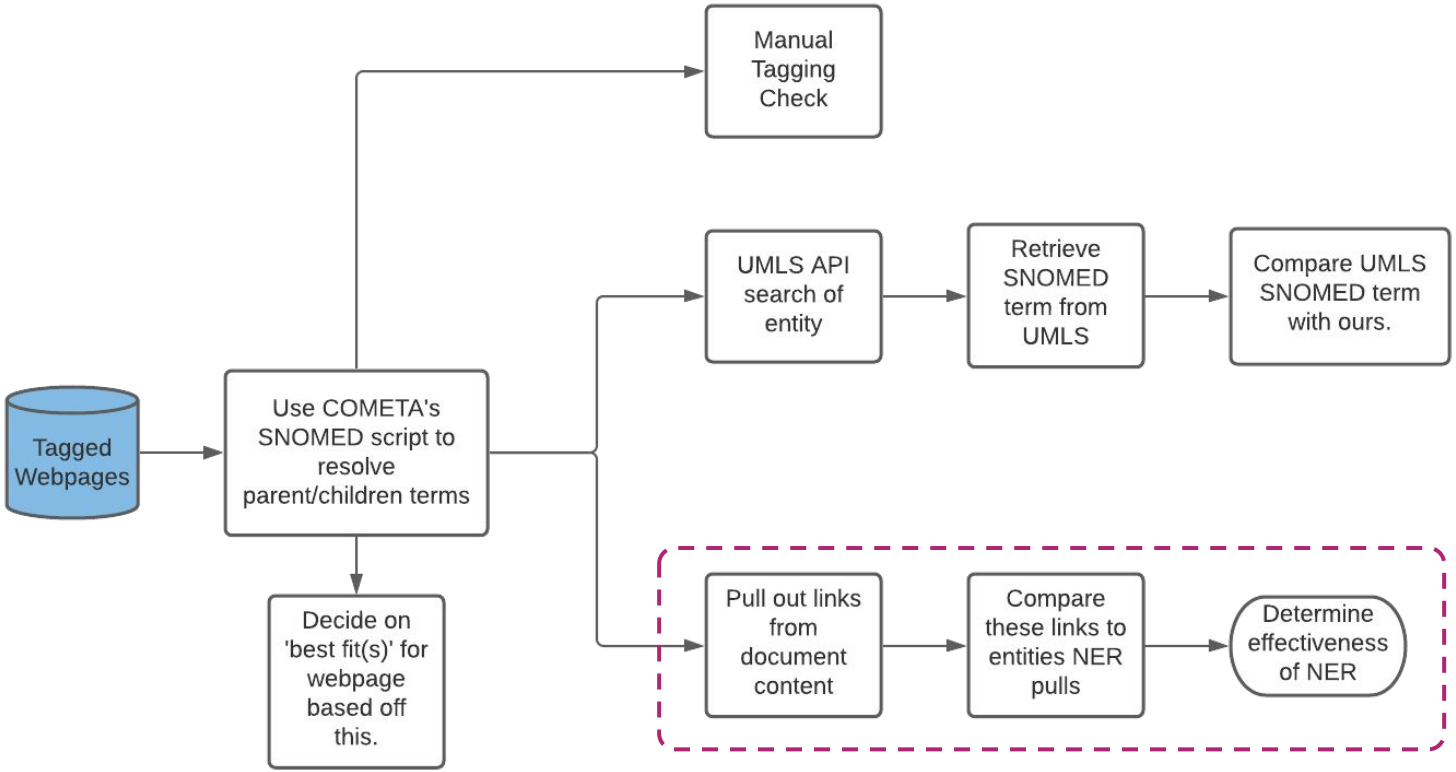


NHS Text Data Exploration - Pipeline



*explore using NHS Digital Terminology Server

NHS Text Data Exploration - Validation



NHS Text Data Exploration - Next Steps

- More indepth exploration at the content of the webpages - relationships between entities to build knowledge graphs, etc.
- Better linking/trimming of constructed entities so that they are better matched to SNOMED CT (especially a problem with EBM-NLP) and how to combine/resolve the outputs from multiple models
- ~ 70% matches of webpage ‘descriptions’, but much lower for whole webpage content - further, can we improve on our validation approaches?
- Explore/compare to other frameworks and models - stanza, HELIN, MedCat, etc.
- How does this perform with other sources of NHS text?

SynPath – Diabetes module

Project Link: https://github.com/nhsx/SynPath_Diabetes

Separate slide pack

Value of Commercial Product Sales Data in Healthcare Prediction

Project Link: https://nhsx.github.io/nhsx-internship-projects/commercial-data-healthcare-predictions/

Machine Learning

Commercial Big Data

Social Good

Shopping Data

Transactional Sales Data

Loyalty Card Data

Epidemiology

Population Health

Respiratory Disease

University of Nottingham
UK | CHINA | MALAYSIA

N/LAB

UKRI Engineering and Physical Sciences Research Council

Machine learning and shopping data

Step 1: ML Models

Create models that can predict registered deaths from respiratory disease using commercial sales data

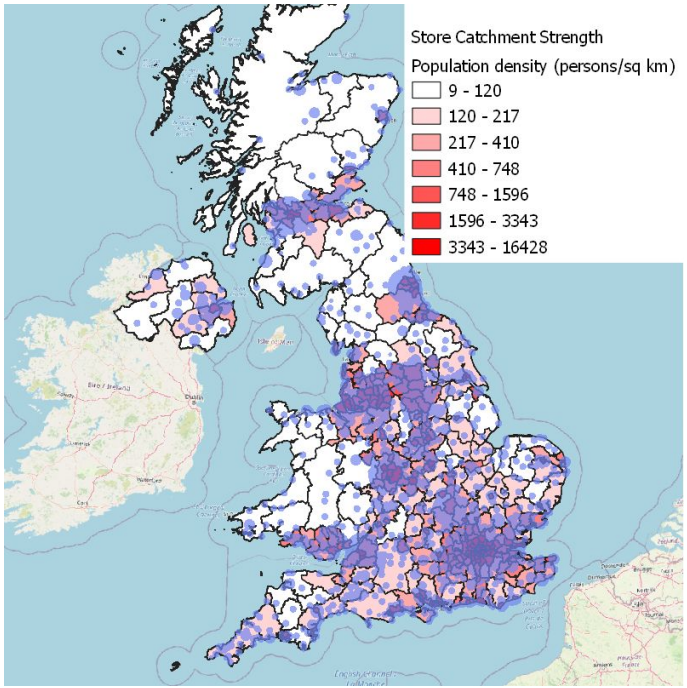
ML models predict:

- Weekly registered deaths
- 17 days in advance
- For 314 Local Authorities across England

Data from March 2016 to March 2020

Inputs on each Local Authority used to make predictions:

- Commercial sales
- Demographics
- Statistics on population density, deprivation, environmental factors incl. living environment
- Weather

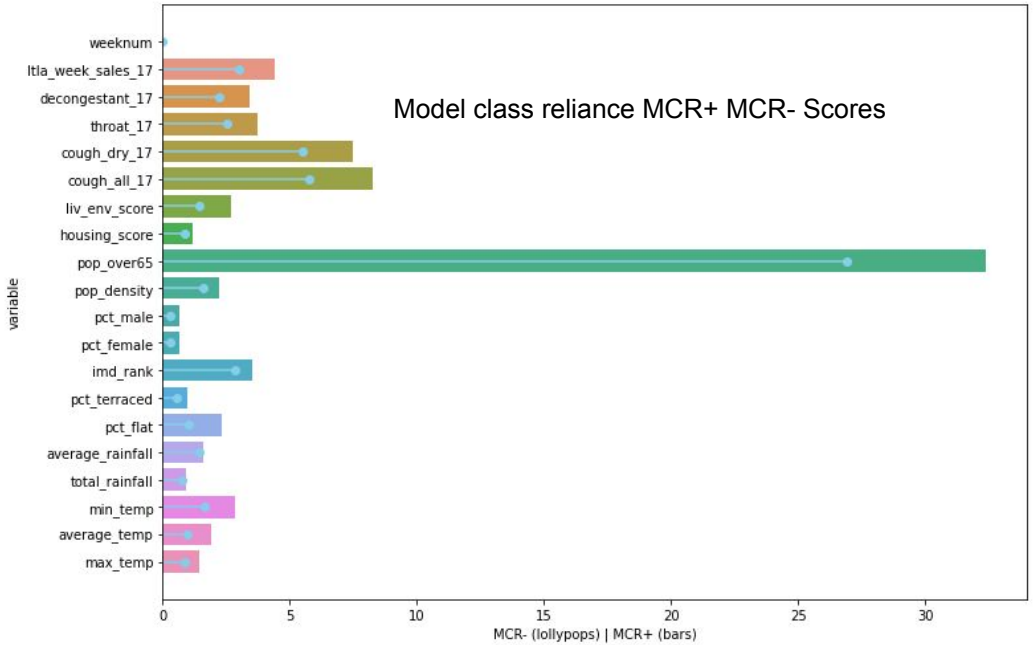


Machine learning and shopping data

Step 2: Explain the models

Use Model Class Reliance [1, 2] to explain the impact of the different variables (inputs) on the models' predictions

- Take account of the importance of the variable across all "best performing" models instead of only one instance of a model.
- Developed by N/LAB expanding MCR [1] to run on Random Forest Models [2]



[1] Fisher et. Al. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. Journal of Machine Learning Research, 2019.
[2]Smith, G., Mansilla, R. and Goulding, J. "Model Class Reliance for Random Forests". 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

Any Questions

https://nhsx.github.io/nhsx-internship-projects/

analytics-unit@nhsx.nhs.uk

NHSX PhD Data Science Internship: Variational Autoencoders for Synthetic Data Generation

Dominic Danks[1,2]

[1]The Alan Turing Institute , [2]University of Birmingham

October 20, 2021

Introduction: Who am I?

Dominic Danks

ddanks@turing.ac.uk

- Turing-funded "Doctoral Student" (2019 start)
- "Neural approaches to stochastic disease modelling"
- Supervisors: Christopher Yau (Manchester), Alastair Denniston (Birmingham), Andrew Beggs (Birmingham).
- Other collaborators: Pearse Keane (London), Vasilis Stavrinides (London)

Introduction: Who am I?

Background:

- MSci Theoretical Physics (Birmingham)
- MSc Computational Statistics and Machine Learning (UCL)

Research:

- Differential Equations x Deep Learning
- Variational Autoencoders
- Survival Analysis

Synthetic Data Exploration:

Variational AutoEncoder for Healthcare Data

Keywords: Synthetic, Variational AutoEncoder

Need: Creating high-fidelity realistic health data is not only complex but comes with multiple information governance considerations. A particularly promising technique for creating realistic synthetic data is the variational autoencoder (VAE). However, current attempts to use VAEs have struggled to put the models into practice as the confidence around appropriate usage and privacy of the ground truth data has not been sufficient. This project would seek to use a currently developed VAE to investigate and discuss its potential when implementing for healthcare data.

Current Knowledge/Examples & Possible Techniques/Approaches: Colleagues in the NHSD data science and innovation team have created a VAE to create realistic Health data.

Related Previous Internship Projects: n/a as first year of the scheme.

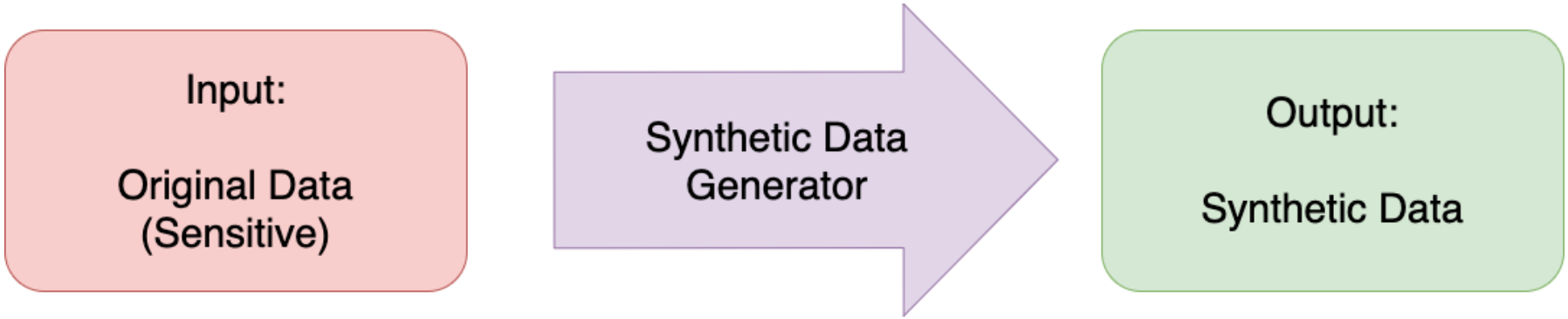
Enables Future Work: Depending on recommendation from this piece, further projects may seek to use or build upon the model.

Outcome/Learning Objectives: Application of model to open data resulting in publication of discussion around appropriate usage. Additionally, interested in coupling the current VAE with differential privacy.

Datasets: Open transactional data with rare values to simulate basic structure of health activity data

Desired skill set: When applying please highlight any experience around synthetic generation (especially variational autoencoders), differential privacy, coding experience (including any coding in the open), any other data science experience you feel relevant.

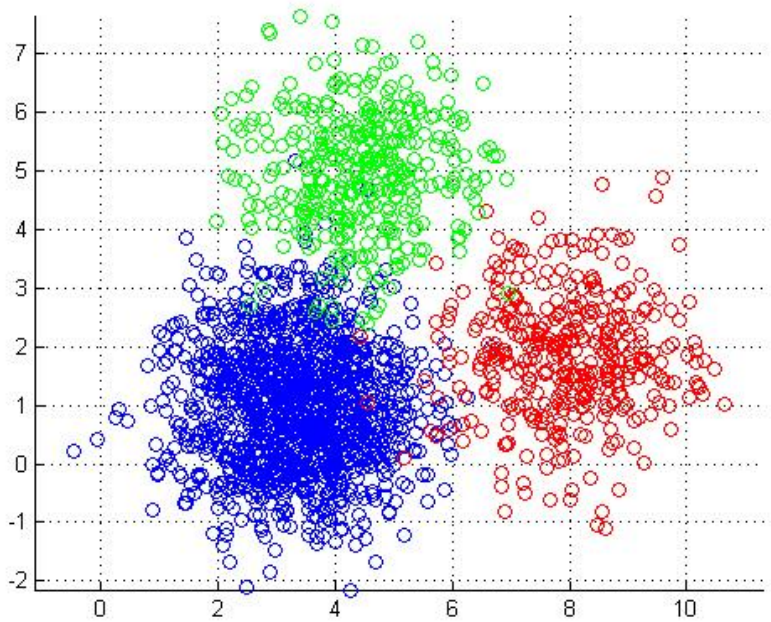
Synthetic Data Generation



- The goal of a synthetic data generation algorithm is to take a potentially sensitive dataset as input and provide a synthetic dataset as output which is **not (as) sensitive** and contains **(most of) the original data's structure**.

A Simple Example

- To understand how we might perform synthetic data generation, consider the dataset below:



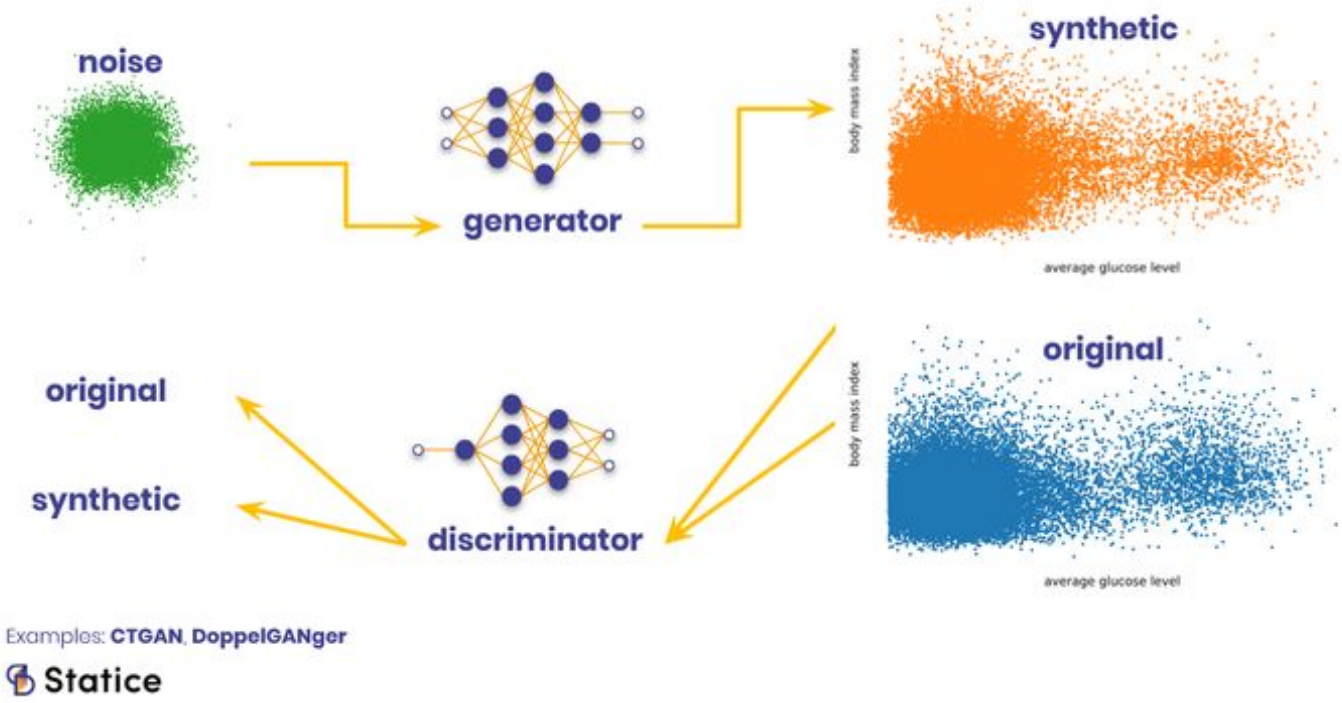
- This data comes from a "Mixture of Gaussians" distribution. If we **learn** that distribution, then we can replace the N observations with N samples from that distribution and get "good" synthetic data.

Generative Modelling

- The previous example illustrates a general point: If we can learn a good **generative model** of the data, then synthetic data can be created by just sampling from that learnt generative model.
- Modern techniques therefore tend to attempt to **learn a generative model of the data**, so that they can then sample from it to generate synthetic datasets.
- The two prevailing approaches for generative modelling in modern machine learning are: Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs).

Briefly: GANs

- GANs learn to transform noise into the observed distribution. They do this by training a **generator** (which learns to generate data) and a **discriminator** (which learns to distinguish between real and synthetic examples). Once trained we only need the **generator**.

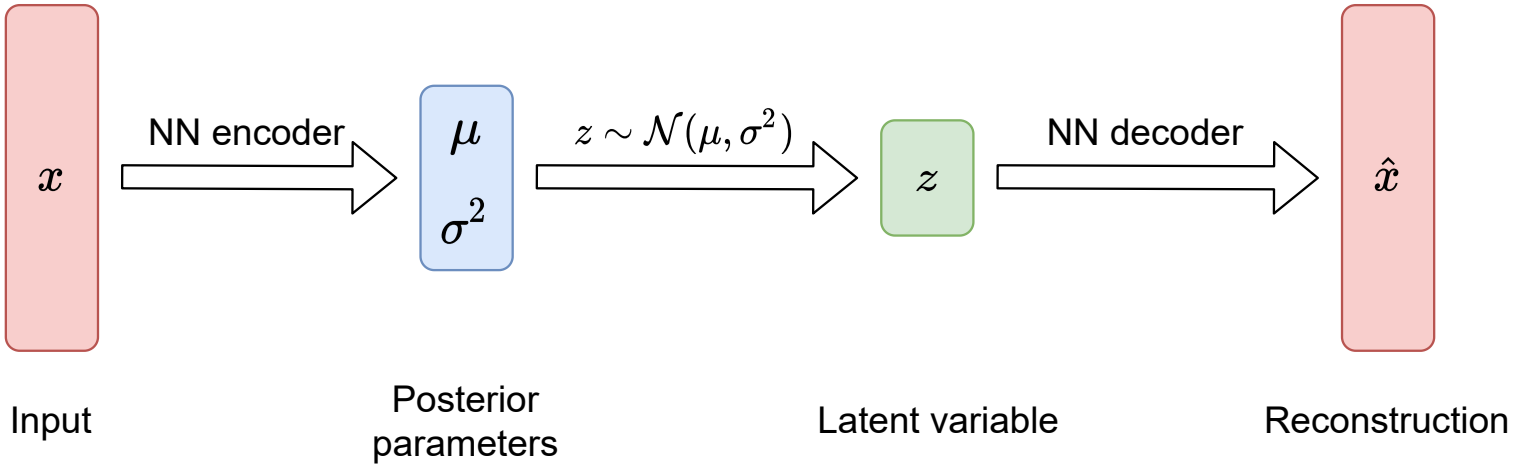


https://www.kdnuggets.com/2021/02/overview-synthetic-data-types-generation-methods.html

Briefly: GANs

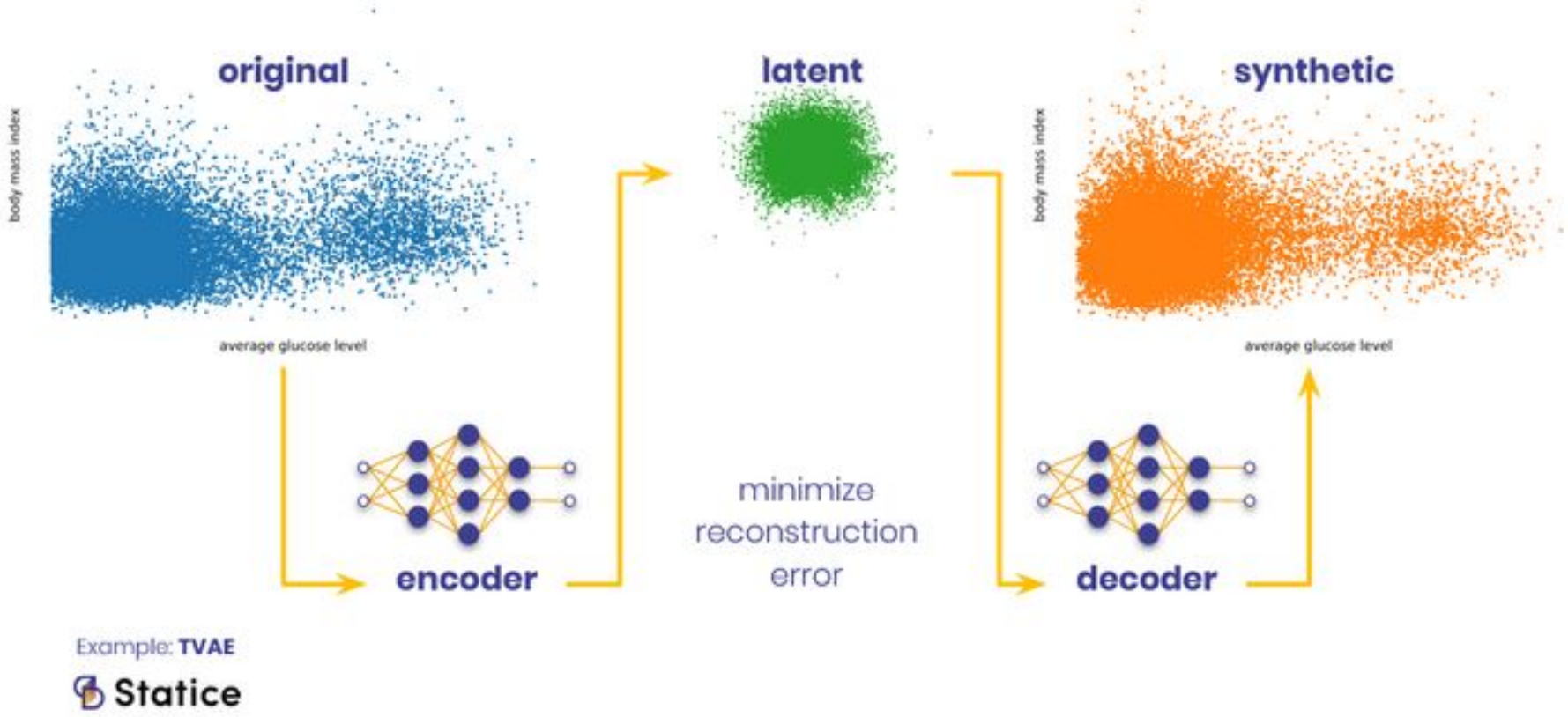
- Notable work has been done on GANs for synthetic data generation, see e.g. PATE-GAN, ADS-GAN, CTGAN.
- GAN pros: Very flexible, no need to specify a likelihood on the data.
- GAN cons: Difficult to train (instability), limited interpretability (no explicit map from observed to latent), no likelihood model frustrates applying constraints or introducing prior knowledge.

VAEs



- The VAE also works on the idea of transforming noise into the observed distribution, but it is trained differently.
- Specifically, during training, observations are passed through an **encoder** ($x \mapsto z$) and a **decoder** ($z \mapsto x$) with the aim of reconstructing the input. By doing this, the model learns to map from z to x, meaning that synthetic samples can be drawn by sampling z and passing through the decoder.

VAEs


original
latent
synthetic
body mass index
average glucose level
body mass index
average glucose level
encoder
minimize
reconstruction
error
decoder
Example: TVAE
Statice

VAEs

- VAEs are less explored than GANs in the context of synthetic data generation.
- They are less black box than GANs and tend to be less unstable.
- A likelihood model must be specified (unlike in GANs). This reduces flexibility in some sense, however generally provides greater stability and predictability and provides the practitioner with the ability to customise the model.

Introducing Differential Privacy

- Using a standard VAE for synthetic data generation does not provide a mathematical guarantee on the privacy of the original data.
- Differential Privacy (DP) is a mathematical formulation of privacy quantified by (ϵ, δ), with $(0, 0)$ corresponding to complete privacy and larger values implying less privacy (see report for details).
- DP can be introduced into the VAE model via DP-SGD. This is an edited version of gradient descent which provides an (ϵ, δ)-DP model as output.

The Project

- Title: *Synthetic Data Exploration: Variational Autoencoders for Healthcare Data*

- Primary purpose: Investigate the suitability of Variational Autoencoders (VAEs) as synthetic health data generators within an NHS context.
  - How does generated data quality compare to other methods?
  - How private is the generated synthetic data and how can this be quantified?
  - How easy is it for a typical user to apply the approach?

Shaping The Project

Shaping the project:

- Exploration
- Scope
- Data
- Timeline
- Tools

Contributions I

- Demonstrate that the Variational Autoencoder is a viable synthetic data generator with performance at least on par with other utilised generators (including copula- and GAN-based models).
- Present the notion of Differential Privacy (DP) and how it can be embedded within a model using the VAE as a case study.
  - DP provides mathematical guarantees on privacy preservation.
- Open code repository containing code to train a (DP-)VAE and benchmark synthetic data generation methods within the Synthetic Data Vault (SDV) software framework.
  - Can be readily used and built upon by others.

Contributions II

- Detailed report including:
  - Didactic introductions to the concepts relevant to the project.
  - SDV usage and functionality details beyond those available in the official documentation.
  - Motivated experiments and implications.
  - Suggestions for natural continuation.

- Project is part of a wider initiative within NHSX to investigate synthetic data as a way to share data resources in a privacy-preserving way.
  - Complements work centred around the Synthea™ package.

Potential Extensions

- Apply a wide range of adversarial attacks to the trained models in order to correlate (ϵ, δ) DP with attack-based metrics, i.e. privacy in practice.
- Introduce DP into models other than the VAE and compare performance vs privacy.
- Implement DP via PATE rather than DP-SGD.
- Apply to additional varied datasets.

Final Remarks

- Scheme offers a great opportunity to work on something slightly different than your usual PhD focus and/or to see how approaches and priorities may differ in academia vs NHS.
  - I had worked extensively on VAEs but less on synthetic data and privacy.
- Great place to intern - Jonny, Dan and the wider group are friendly and knowledgeable!
- If in doubt, apply!
- Do contact me for more info and/or an informal chat:

ddanks@turing.ac.uk

Thank you for your attention!

NHS X

NIHR | Applied Research Collaboration North Thames

NHSX internship (SynPath)

Tiyi Morris, PhD Data Science Intern
31 August 2021

Coming into the data science internship as a health economics PhD student

- Tiyi Morris (BA in Philosophy, Politics and Economics from Warwick and MSc in Public Policy and Management from King's College London)
- PhD supervisors: Dr Manuel Gomes, Dr Fiona Aspinal and Dr Jean Ledger at UCL
- PhD student in the NIHR ARC North Thames at the Department of Applied Health Research at UCL

- Internship supervisors: Dr Jonathan Pearson and Dr Daniel Schofield
- Acknowledgements: Ben Goldhaber and David Wilkinson at hash.ai, and Professor Elisabeth Murray and Dr Jamie Ross at UCL

Background

Synthetic data has potential to allow NHS analysts to analyse clinical pathways without patient data

Problem:

- Challenges around data access limit the NHS' ability to carry out analysis of clinical pathways

Proposed solution:

- Agent-based models (e.g. Synthea) for synthetic data can help us understand patient flows

Use case:

- Type 2 diabetes

Outputs:

- Technical report, project summary and GitHub repository

References:
Synthea repository (2021), available at: https://github.com/synthetichealth/synthea, accessed 18th August

SynPath for diabetes allows us to look at a key service

- Simulation was chosen to:
  - Model patient flows
  - Produce longitudinal synthetic records
- SynPath was used instead of Synthea so that we could include services for the NHS in England
- Type 2 diabetes is a great example for the first pathway to model because of its clinical relevance and resource impacts
  - Increasing prevalence and service impacts (Diabetes UK, 2021)
  - Potential for impact of digital health interventions (DHIs)
  - DHIs do this by changing individual behaviour – key to agent-based modelling

References:
Synthea repository (2021), available at: https://github.com/synthetichealth/synthea, accessed on 18th August 2021
Hex et al. (2012), Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs, Diabetic Medicine, Vol 29, Issue 7, p. 855-862
Diabetes UK (2021), Diabetes Statistics, avaliable at: https://www.diabetes.org.uk/professionals/position-statements-reports/statistics, accessed 19th August 2021

The goal was to help develop a framework for using SynPath that can be built on in future

What did I do?
Generate an outline of what is needed to build an intelligence layer that:

- works with SynPath and;
- outputs realistic synthetic records for NHS patient pathways

How did I do it?
Demonstrate and document the process of developing a use case of type 2 diabetes patient pathways to include digital technologies using the SynPath framework

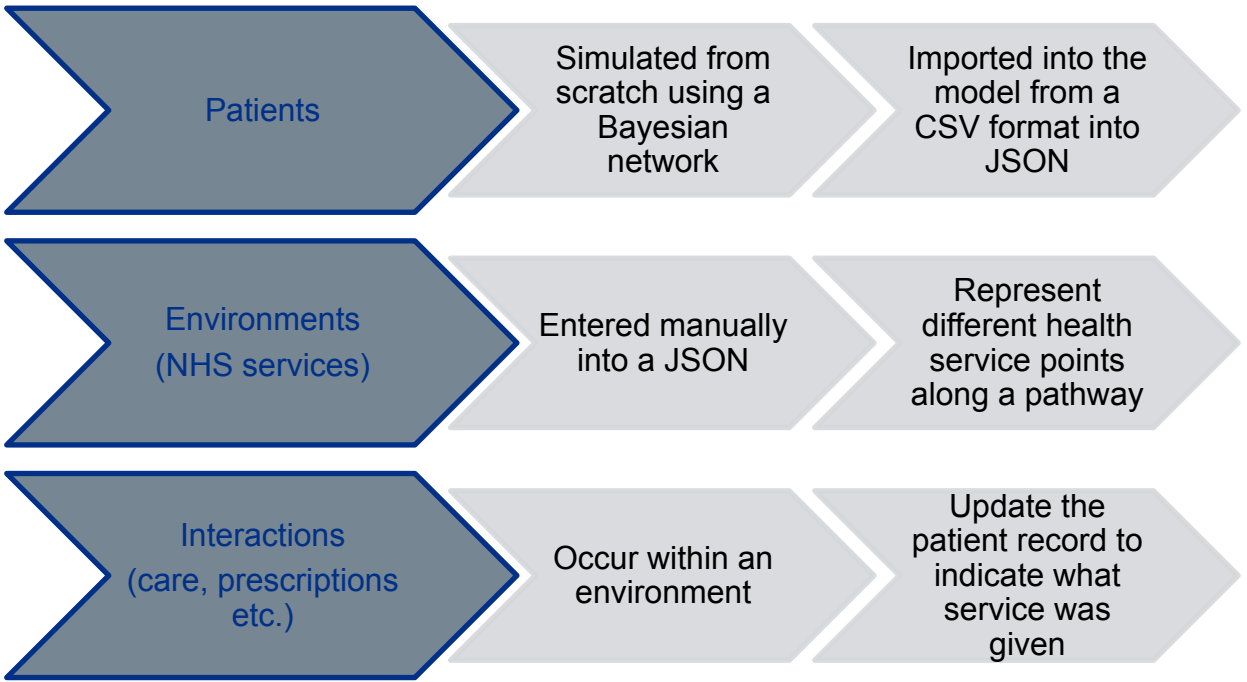
References

Faculty.ai, report available at: https://github.com/nhsx/SynPath/blob/master/REDACTED_C245%20ABM%20Patient%20Pathways_Final%20Report_V3_28042021.cleaned.pdf, accessed 18th August

SynPath repository (2021), available at: https://github.com/nhsx/SynPath, accessed 18th August

Simulation

In SynPath patients receive services from environments through interactions

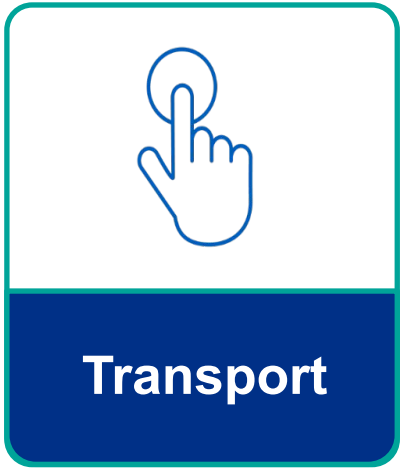


References:

Faculty.ai, report available at: https://github.com/nhsx/SynPath/blob/master/REDACTED_C245%20ABM%20Patient%20Pathways_Final%20Report_V3_28042021.cleaned.pdf, accessed 18th August

SynPath repository (2021), available at: https://github.com/nhsx/SynPath, accessed 18th August

Simulation approaches from different disciplines contributed to the approach



References:
Squires et al. (2016). A Framework for Developing the Structure of Public Health Economic Models, Value in Health, 19: 588-601.
Oxford HERC (2020) UKPDS model v.2. https://www.dtu.ox.ac.uk/outcomesmodel/.
The MITRE Corporation. (2017-2021). SyntheaTM Patient Generator, https://github.com/synthetichealth/synthea.
Hash (2021) Risk of Diabetes model, https://core.hash.ai/@hash/risk-of-diabetes/2.1.0.
Paranjape et al. (2018). The Diabetic Patient Agent: Modelling Disease in Humans and the Healthcare System Response (Springer: Berlin, Germany).
Xie et al. (2014) Agent-Based Modeling and Simulation for the Bus-Corridor Problem in a Many-to-One Mass Transit System, Discrete Dynamics in Nature and Society, 2014: 652869
Kagho et al. (2020) Agent-Based Models in Transport Planning: Current State, Issues, and Expectations, Procedia Computer Science, 170: 726-32.
Tkachuk et al.(2018) Application of artificial neural networks for agent-based simulation of emergency evacuation from buildings for various purpose, IOP Conference Series: Materials Science and Engineering, 365

Elements of the model were populated with a set of inputs to simulate the type 2 diabetes pathway

References:
Public Health England (PHE). 2018/19. Diabetes (Fingertips data profile)
https://fingertips.phe.org.uk/profile/diabetesft/data#page/0/gid/1938133136/pat/44/par/E40000003/ati/154/are/E38000004/cid/4/tbm/1
NHS England. 2017. NHS RightCare Pathway: Diabetes. https://www.england.nhs.uk/rightcare/products/pathways/diabetes-pathway/.
NICE. 2020. Type 2 diabetes in adults: management, Accessed 1st June 2021. https://www.nice.org.uk/guidance/ng28.

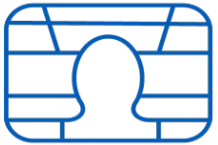
Learning

Different optimisation approaches could be used to modify outcomes in the model

Stochastic gradient descent

Reinforcement learning

Monte Carlo tree search



A* search

References:

Geron, A. 2019. Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow (O'Reilly).

hash.ai. (2021) 'Q-learning', https://core.hash.ai/@hash/qrl/1.0.0.

Hash.ai (2021) Monte Carlo tree search, https://core.hash.ai/@b/mcts/main

Baijayanta (2019) A-Star (A) search algorithm, https://towardsdatascience.com/a-star-a-search-algorithm-eb495fb156bb*

Future projects can build on this project's research to build a more robust model

During the internship, we developed insights around how to develop SynPath in future

Optimisation

- Approach
- Disease
- Demographics

Validity

- Capacity
- Summary
- Simultaneous

Connect with us

Web: **www.nhsx.nhs.uk**

Email: **feedback@nhsx.nhs.uk**

@NHSX

www.linkedin.com/company/nhsx