



Giving Structure to Unstructured Healthcare Workforce Documents Using Natural Language Processing

Jerome Wynne
jeromewynne@das-ltd.co.uk

Siôn Cave
sioncave@das-ltd.co.uk

21st October 2021



- Independent management consultancy specialising in strategy, change, programme management and data analytics.
- DAS works in collaboration with clients across Government, Local Authorities, Health, Defence, Police and private sector organisation from the UK, North America, Europe, Japan and Australasia.
- Our clients look to us for strategic and operational support for their most critical issues.
- DAS Analytics + Foresight Hub is DAS' centre for the application of innovative data science, systems modelling and futures techniques.



- Workforce dashboards using Qlik, Tableau and Power BI for NHS and Government departments.
- Systems thinking based studies to better understand workforce issues for Government Departments and national charities
- Quantitative workforce supply and demand modelling for health and social care workforces, industrial workforces and organisations.

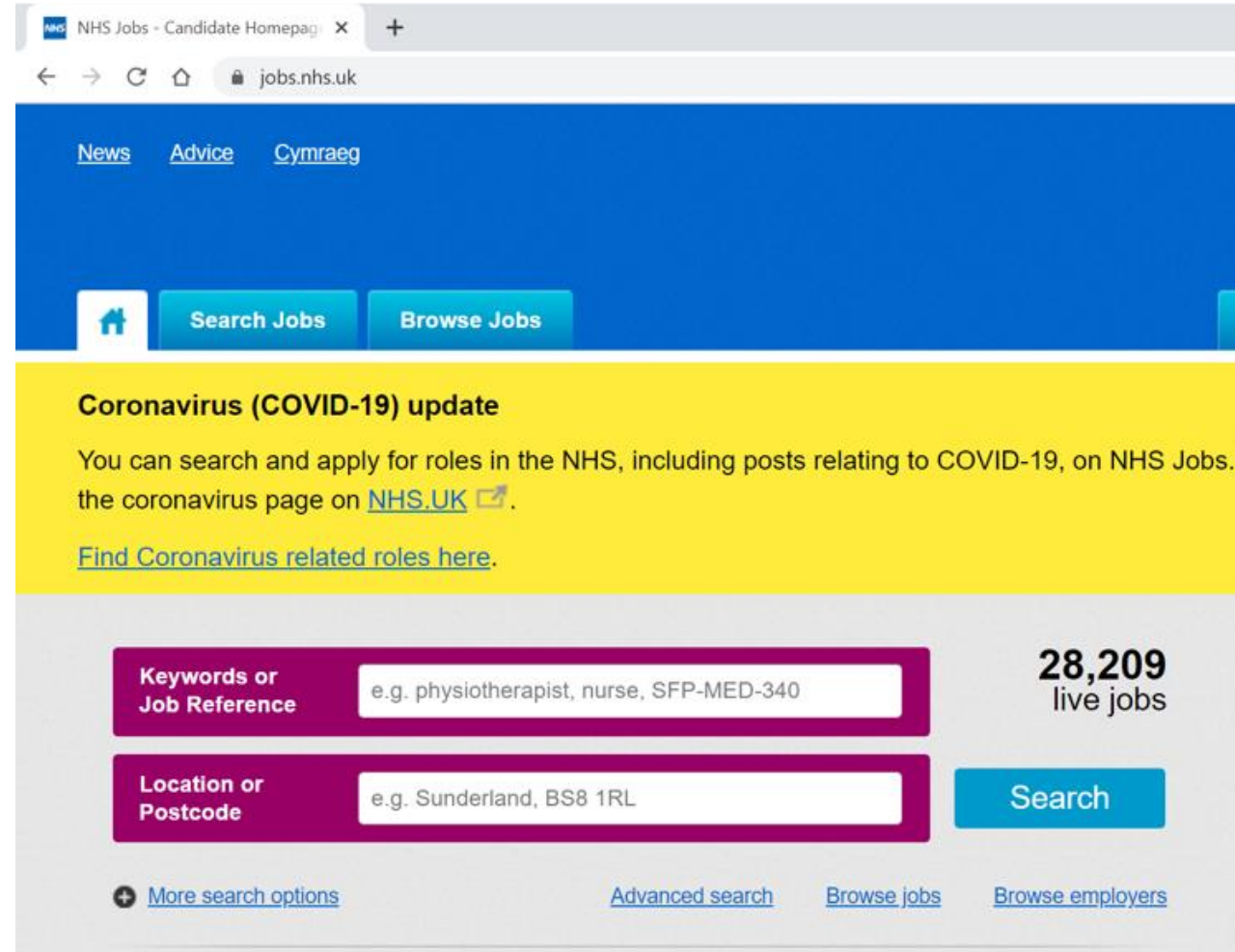
1. Context
2. Problem
3. Approach
4. Execution
5. Summary

Context

DAS Internal project initiated by NHS Wales modelling hackathon challenge.

Stakeholders from NHS workforce intelligence wanted to profile how skills are distributed across the NHS workforce.

Identified the NHS Jobs website as a possible source of useful information.



Job Reference: J265-3248

Employer: [Elysium Healthcare](#)

Location: Swindon

Salary: £40,894 - £45,953 + Benefits

If you are looking for a new opportunity to contribute to the ongoing development of a recently opened service in a growing hospital that puts service users at the heart of what they do, then join the team at The Avalon Centre.

As an experienced Clinical Psychologist, you will assist in the development and provision of neurobehavioural rehabilitation to people with acquired brain injury with complex needs at The Avalon Centre in Swindon, which opened in June 2020.

Working alongside a transdisciplinary team, you will be interested in developing or expanding your skills in clinical neuropsychology, particularly regarding the management of neurobehavioural disability.

The clinical model which underpins the service is psychological, delivered by a transdisciplinary team (TDT) enabling a 24/7 approach to rehabilitation. The Clinical Lead for the service is a Consultant Clinical Psychologist.

You will join a highly motivated and dynamic team as a sub-consultation graded psychologist, led by an established team of professionals with many years of combined experience within this specialist field led by

Job Type: Permanent

Working pattern:

- Full time
- Flexible working

Pay Scheme: Agenda for change

Pay Band: 8a

Staff Group: Allied Health Professionals

 **Read this before applying**

[Job Description & Person Specification \(21 KB\)](#)

[Apply for this job](#)

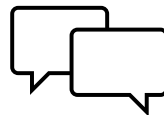
Closing Date:

This job advert will close as soon as sufficient applications have been received. Please apply for this job as soon as you can, if interested.

What is it?

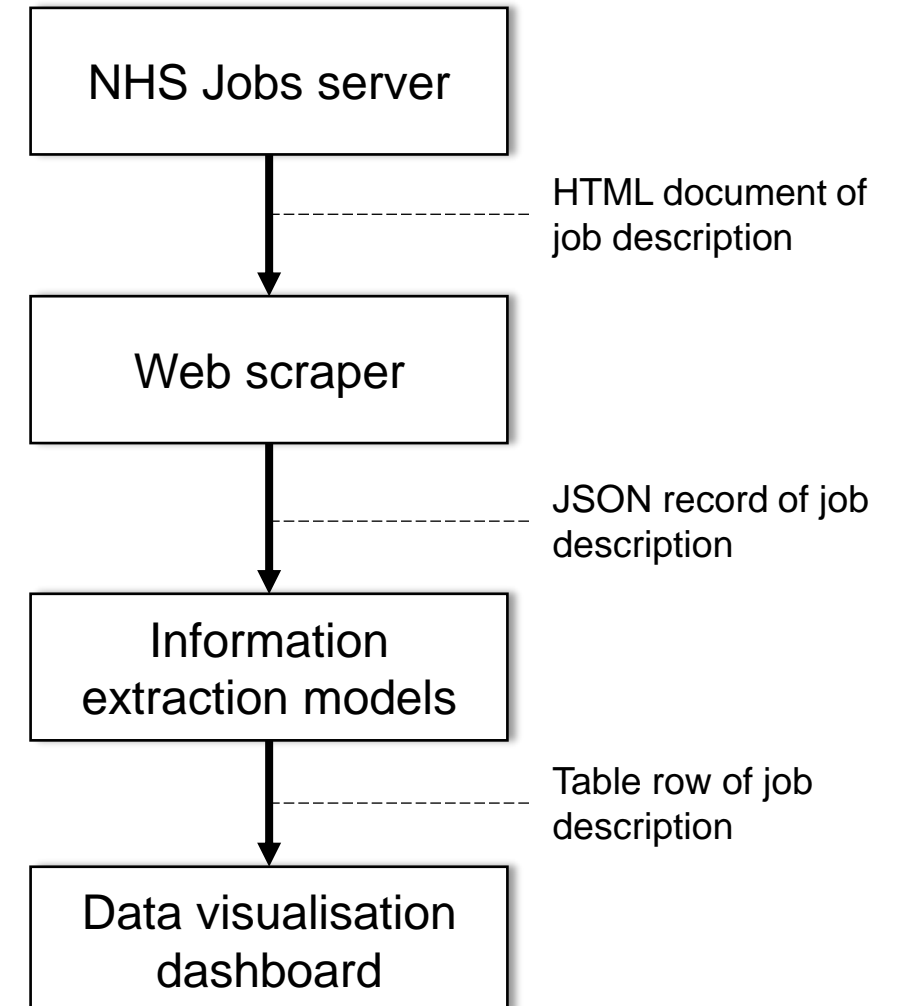
>> Computational analysis and synthesis of language data <<

>> Language data = text and speech <<



>> Use the data on NHS Jobs to profile the NHS workforce <<

1. Scrape unstructured data from NHS Jobs.
2. Apply automated information extraction techniques to map this unstructured data to structured data.
3. Visualise the structured data in a web application.



- Extracts job description from HTML document into JSON record.
- Built using Python and BeautifulSoup.
- Available at github.com/stochastictalk/nhs_jobs_data

Execution – Web Scraper

```
management of neurobehavioural disability.</p>
<p>The clinical model which underpins the service is psychological, delivered by
for the service is a Consultant Clinical Psychologist.</p>
<p>You will join a highly motivated and dynamic team as a sub-consultation grade
experience within this specialist field led by our Clinical Director Professor
<p><strong>Our Clinical Psychologist role involves:</strong></p>
<ul>
<li>Deliver Clinical Neuropsychology input as a specialist discipline under the
<li>Consultant Clinical Psychologist</li>
<li>Deliver successful strategic and clinical strategies underpinning the busin
<li>Contribute to the local and national networks and additional stakeholders to
<li>Engage in teaching, service evaluation and research as agreed, and particip
and events</li>
<li>You will provide specialist psychological assessments, interpretation and in
<li>neuropsychological tests, self-report measures, rating scales, direct and in
members and the wider team</li>
<li>Functioning as a key member of the transdisciplinary team, ensuring success
culture and Elysium's key values</li>
</ul>
<p><strong>To be successful as the Clinical Psychologist, you will need:</strong>
<ul>
<li>To be British Psychological Society accredited and HCPC approved Doctorate
<li>Registered as a Practitioner Psychologist with the HCPC</li>
<li>Chartered Member of the British Psychological Society</li>
<li>Experience in working with recipients of clinical services for people with
<li>Ambition to develop a career in Clinical Neuropsychology and an interest in
</ul>
<p><strong>What you will get:</strong></p>
<ul>
<li>Annual salary of £40,894 - £45,753 (DOE)</li>
<li>The equivalent of 33 days annual leave - plus your birthday off!</li>
<li>Free meals and parking</li>
<li>Wellbeing support and activities</li>
```

HTML document

Web
scraper

```
{
  "@context": "http://schema.org",
  "@type": "JobPosting",
  "baseSalary": {
    "@type": "MonetaryAmount",
    "currency": "GBP",
    "value": {
      "@type": "QuantitativeValue",
      "maxValue": "28333",
      "minValue": "28333",
      "unitText": "YEAR",
      "value": "£28,333.20 per annum"
    }
  },
  "datePosted": "2019-08-08T11:13:33+0100",
  "hiringOrganization": {
    "@type": "Organization",
    "name": "Priory Group",
    "url": "https://www.jobs.nhs.uk/xi/agency_info/?agency_id=133434"
  },
  "industry": "Healthcare",
  "jobLocation": {
    "@type": "Place",
    "address": {
      "@type": "PostalAddress",
      "addressCountry": "GB",
      "addressLocality": "Bartle, Preston",
      "postalCode": "PR4 0HB"
    },
    "geo": {
      "@type": "GeoCoordinates",
      "latitude": "53.8011",
      "longitude": "-2.78397"
    }
  },
  "title": "Registered Mental Nurse (RMN)",
  "url": "https://www.jobs.nhs.uk/xi/vacancy/915700878",
  "validThrough": "2020-11-27T23:59:59+0000"
}
```

JSON record

14,841 records, each record is a job description.

Fields:

- Role title
- HTML of free-text role description
- Salary
- Employment type
- Location (latitude, longitude)
- ...

```
"title": "Registered Mental Nurse (RMN)",
```

```
"description": "&lt;p&gt;&lt;span style  
=&quot;text-decoration: underline  
&gt;&lt;em&gt;&lt;strong&gt  
&gt;Are you looking for an exciting new  
challenge and want to join a service  
where you really can make a  
difference in people's lives every  
single day? If so, we have the  
perfect role for you!&lt;/strong&gt  
&lt;/em&gt;&lt;/span&gt;&lt;/p&gt  
&n&lt;p&gt;Priory Hospital Preston  
is a 38-bedded independent mental  
health hospital, specialising in the  
management and treatment of acute.  
Opened in 1998, the hospital has  
established an excellent reputation
```

Execution – Cleaning Text Fields

Challenges:

- >> HTML entities
- >> Inconsistent grammar
- >> Incorrect spelling
- >> Multiple jobs in one post
- >> Sparse descriptions

```
&lt;p&gt;&lt;strong&gt;Do you have good time management skills and  
can work accurately under pressure?&lt;/strong&gt;&lt;/p&gt;  
&lt;p&gt;&lt;strong&gt;East & North Hertfordshire Hospital  
Trust &lt;/strong&gt;would love to hear from you!&lt;/p&gt;
```

```
Do you have good time management skills and can work accurately under pressure?  
East & North Hertfordshire Hospital Trust would love to hear from you!
```

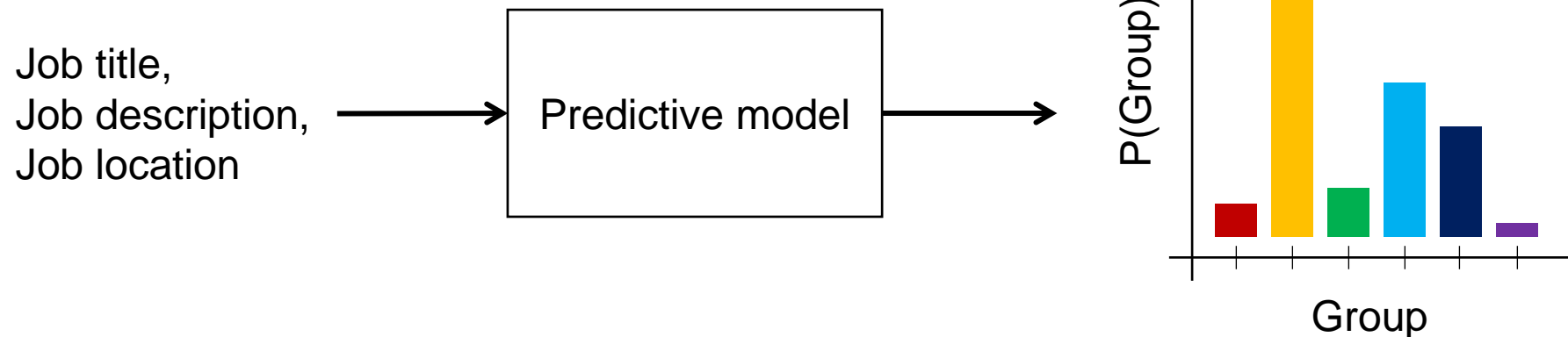
title

Children Nurse

description

Do you have good time management skills and can work accurately under pressure? East & North Hertfordshire Hospital Trust would love to hear from you! NHS Professionals run the temporary work placements at Lister Hospital, and we are looking for a Children's Nurse to join their team ASAP! Hours per week 37.5 Duration: Ongoing Working Pattern : Flexible and Rota'd Grade/ Band: 5 Job Description: To work as a skilled member of the nursing team, to provide a high standard of holistic family centred care to the children and their families. To act as a role

Explored whether a model could be built that mapped job descriptions to NHS Occupation Groups.



- 200 labelled job descriptions (2 hours of labelling time).
- 5 classes.

Occupation group	# of labelled job descriptions
Nursing, midwifery and health	63
Scientific, therapeutic and technical staff	42
Medical and dental staff	39
Administration and estates	32
Health care assistants and other support staff	21

Execution – Occupation Group Model

Dataset Labelling Utility

.\config\nhs_jobs-config.json x ▾



- 10038-Nursing Assistant
- 6105-Locum Consultant...
- 5505-Team Secretary
- 12258-Specialist Optom...
- 5702-Healthcare Scienti...
- 14354-REDS Operative
- 5411-Resilience Suppor...
- 1556-Consultant Medic...
- 10613-Associate Speci...
- 10152-Health Care Assi...
- 3380-Consultant in Obs...
- 5506-Locum Consultant...
- 7415-Team Manager- ...
- 10039-Dietician
- 6849-Band 6 Midwife
- 10986-Bank Data Entry ...
- 13527-Clinical Researc...
- 9871-Trust Specialty Re...
- 6456-Community Practi...
- ? 5462-Staff Nurse (Police...
- ? 5662-Clinical Psychologi...
- ? 4525-Project Support Off...
- ? 2629-Senior Radiograph...

title

Consultant Medical Microbiologist

description

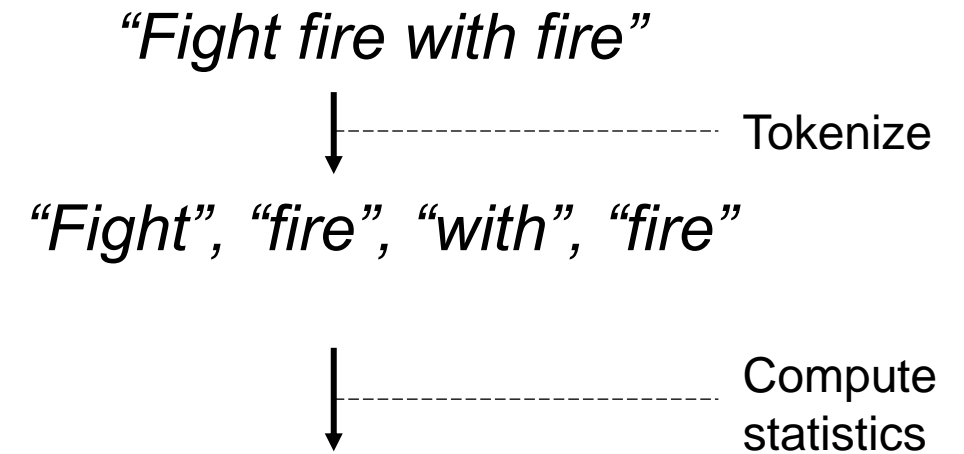
The two posts are replacement, whole time, Consultant Medical Microbiologists employed by St Helens & Knowsley Teaching Hospitals NHS Trust (STHK). The two successful candidates will join a team of four WTE Consultants working across the two Trusts, St Helens & Knowsley Teaching Hospitals NHS Trust and Southport and Ormskirk Hospital NHS Trust. The two posts shall be based at Southport and Ormskirk Hospital NHS Trust. The Pathology Directorate is within the Clinical Support Services Care Group and provides the services to STHK and Southport & Ormskirk Hospitals Trust (SOHT) as well as direct access. The Microbiology laboratory is located on a single site at Whiston Hospital. The Trust has excellent management, robust systems and innovative processes all of which have underpinned a period of sustained and improved performance. All applicants must possess FRCPath or equivalent, have training and broad experience in Microbiology.

label

Medical and dental staff x ▾

101 of 14,841 records labelled

- Free text fields were treated as ‘bags of words’.
- Stripped ‘stop’ words – words that appear in most documents.
- tf-idf statistics* were computed for remaining words.
- Yielded 38,358 features.



Token	Frequency
Fight	1
Fire	2
With	1

*Intro to applied tf-idf using R
<https://www.tidytextmining.com/tfidf.html>

- Multinomial regression model.
- Linear – classes are distinguished by comparing weighted sums of their tf-idf statistics.
- Mitigated overfitting using L1 regularisation.

Class i score assigned to example

Weighted sum of components of x

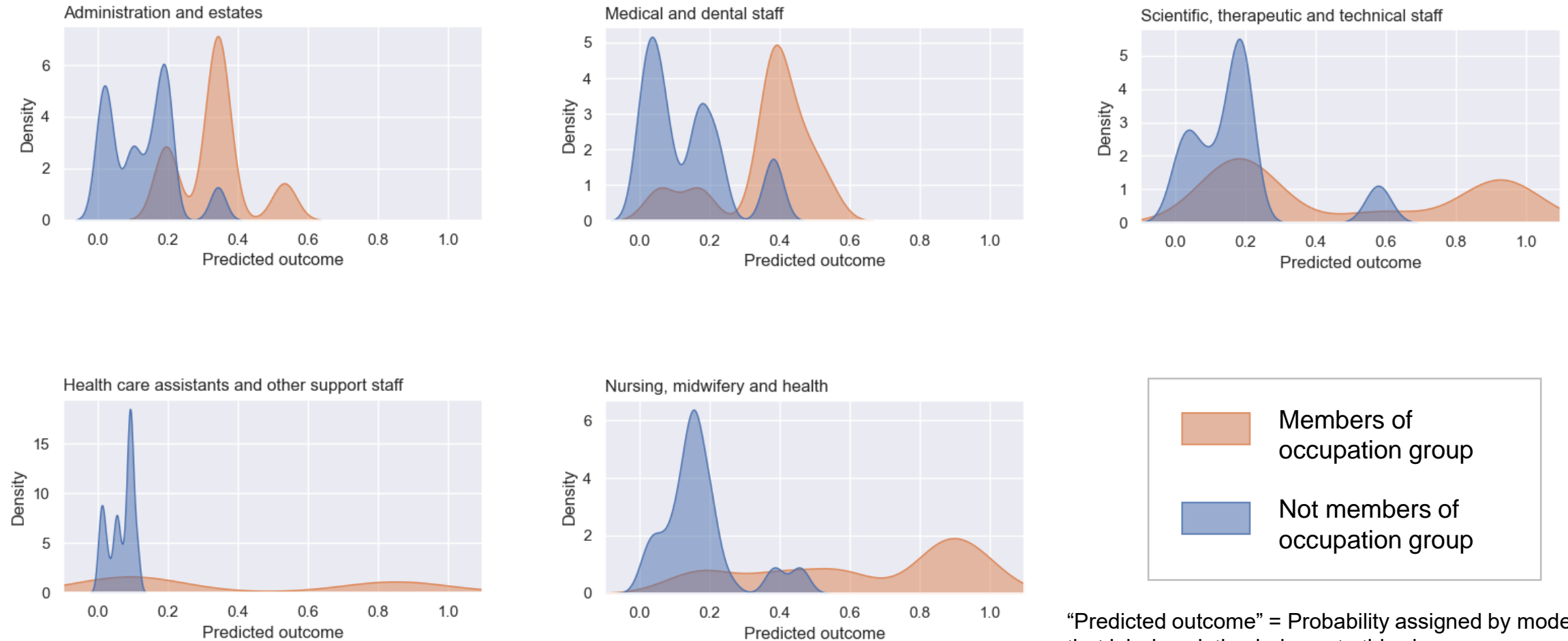
$$P(y = i|x) = \frac{e^{\beta_i^T x}}{\sum_{j=1}^k e^{\beta_j^T x}}$$

Probability example x is in class j

Sum of example x 's scores across all k classes

Execution – Occupation Group Model

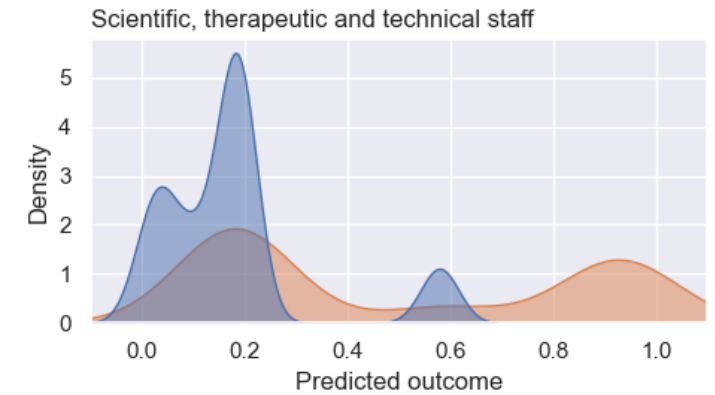
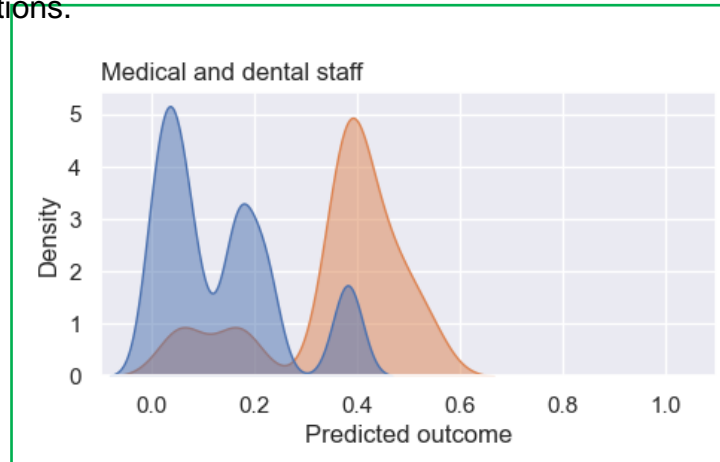
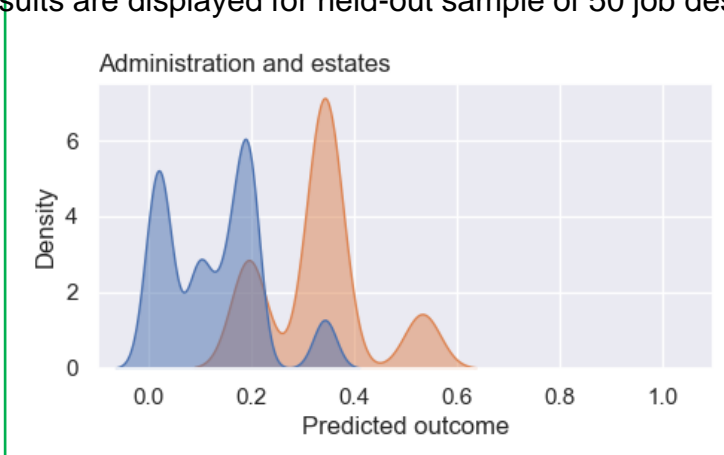
* Results are displayed for held-out sample of 50 job descriptions.



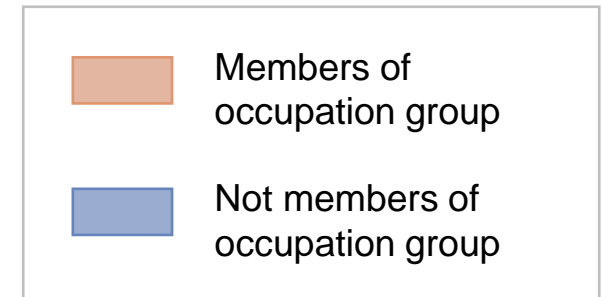
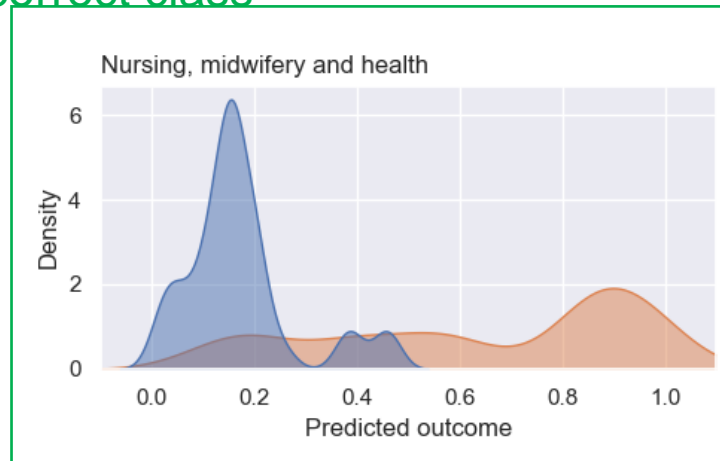
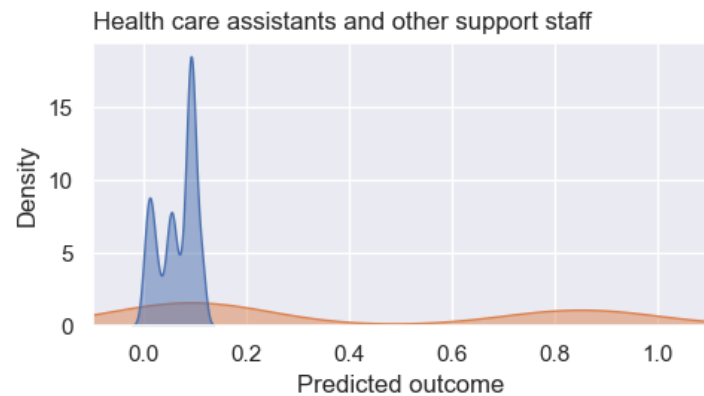
“Predicted outcome” = Probability assigned by model that job description belongs to this class

Execution – Occupation Group Model

* Results are displayed for held-out sample of 50 job descriptions.

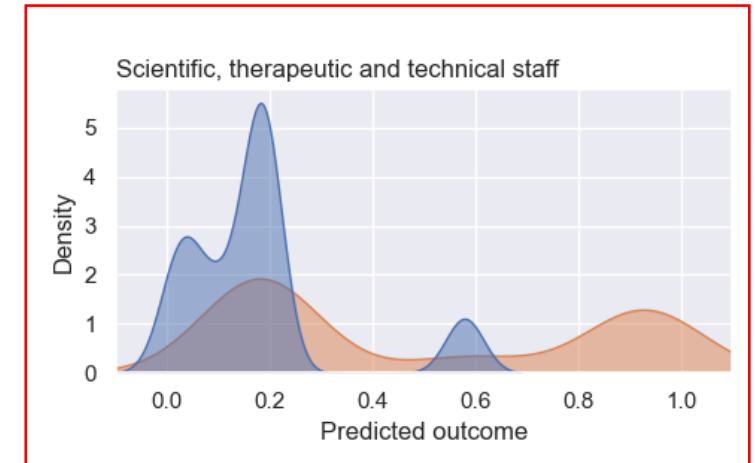
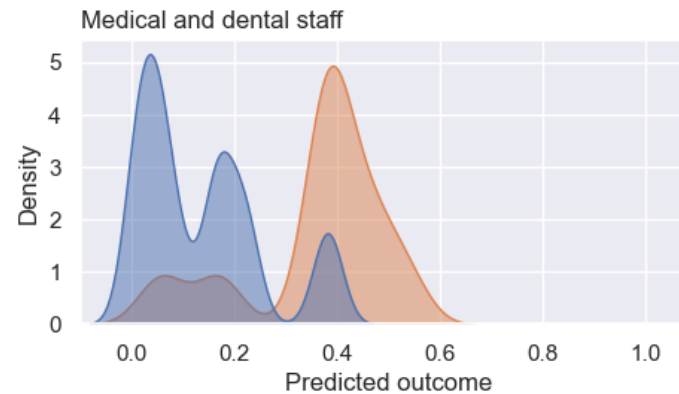
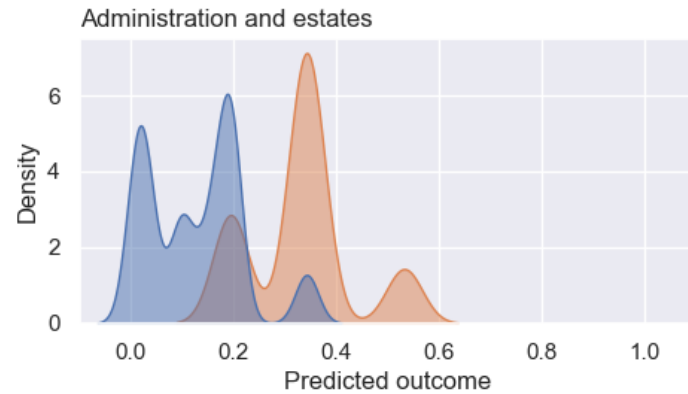


Model creates a strong signal indicating the correct class

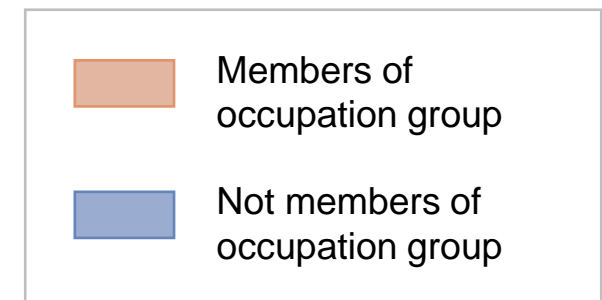
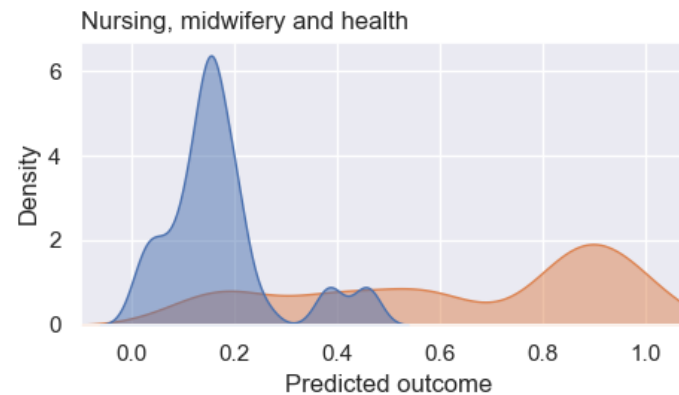
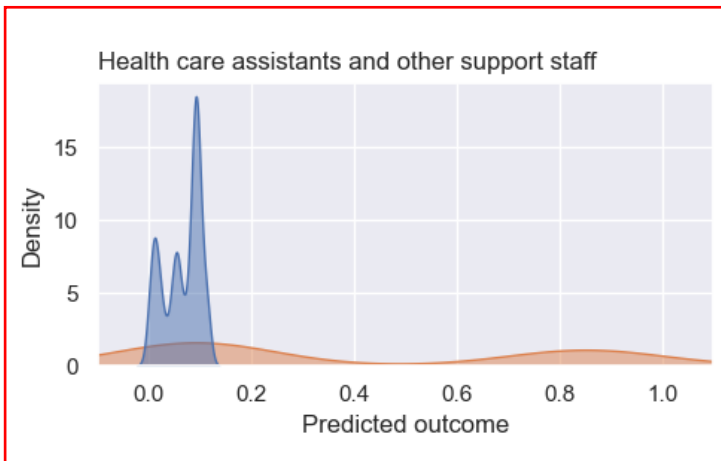


Execution – Occupation Group Model

* Results are displayed for held-out sample of 50 job descriptions.



Model creates a weak signal indicating the correct class



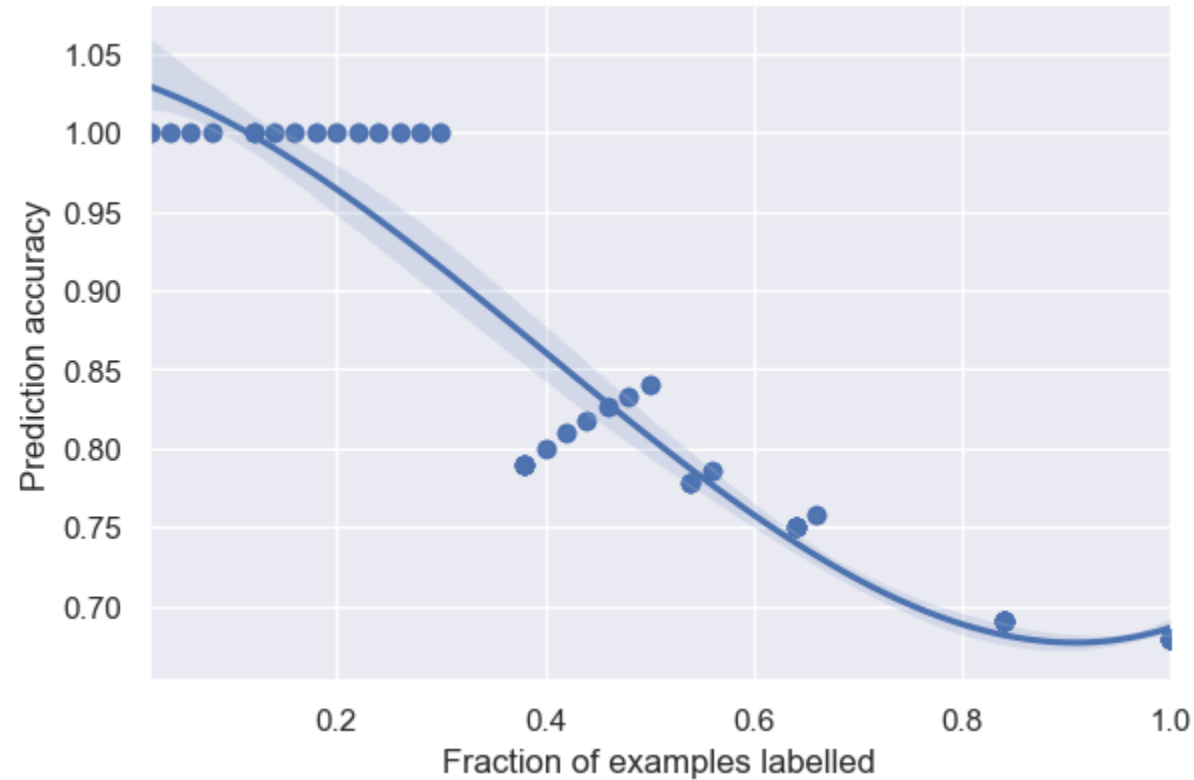
Execution – Occupation Group Model

* Results are displayed for held-out sample of 50 job descriptions.

Occupation group	% correctly labelled by model	# of training examples
Nursing, midwifery and health	81	44
Medical and dental staff	80	32
Administration and estates	75	29
Scientific, therapeutic and technical staff	46	24
Health care assistants and other support staff	40	15
All groups	70	144

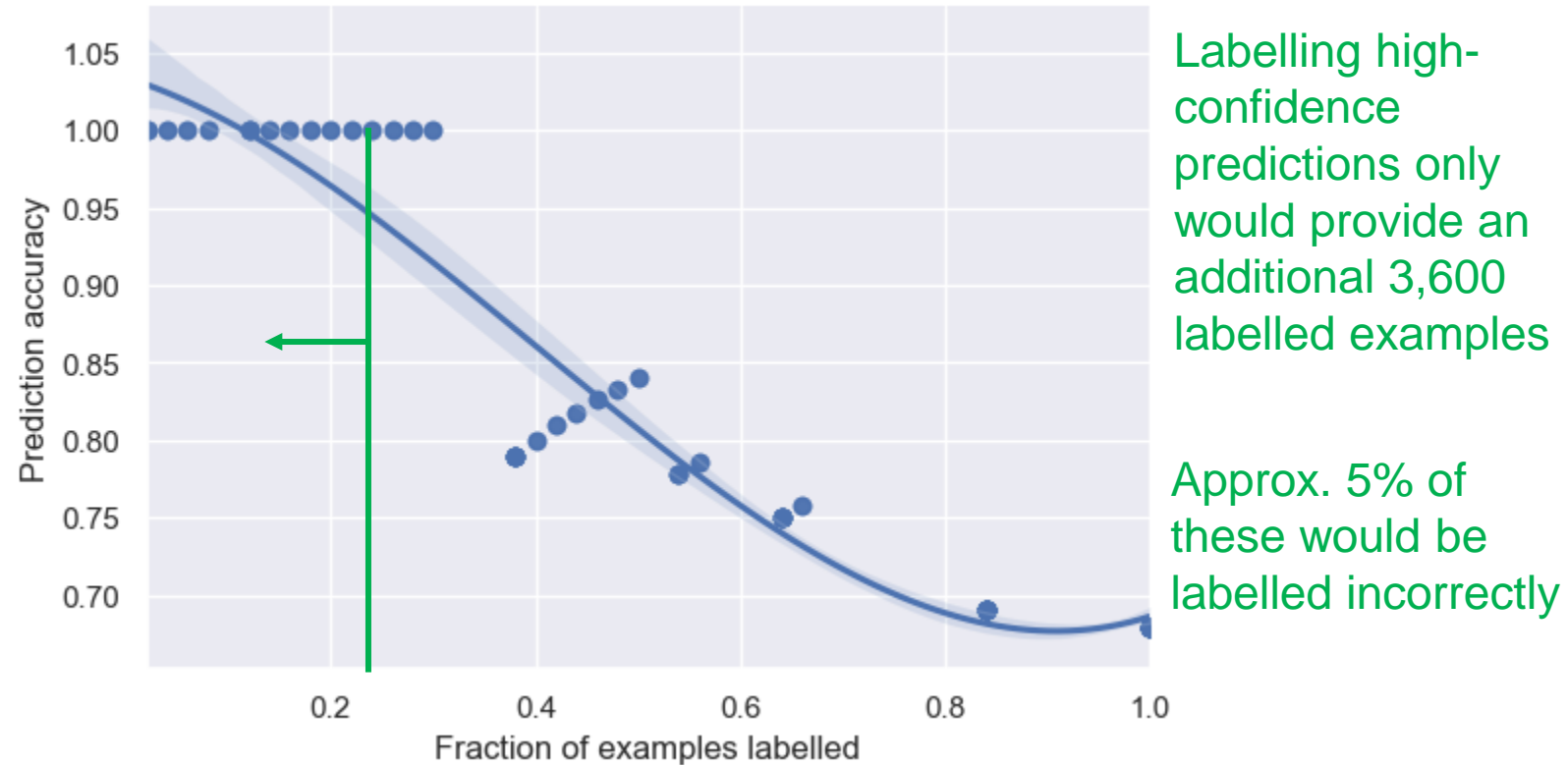
Execution – Occupation Group Model

* Results are displayed for held-out sample of 50 job descriptions.



Execution – Occupation Group Model

* Results are displayed for held-out sample of 50 job descriptions.



- Model fit to small amount of data (~150 records) acted as an entry point to a large amount of labelled data (~3,600 records).
- Job descriptions could then be analysed by staff group:
 - Mentions of specific skills
 - Experience requirements
 - Degree requirements
 - Salary
 - Location

- Where else could this be useful?
 - Anywhere that:
 - manual labelling of records is necessary but expensive,
 - and richer insights can be gleaned from more labelled examples (i.e. a larger sample is needed to profile the population).
- Limitations of approach:
 - Need to be careful that statistics analysed are not correlated with classifiability.
- How could it be developed?
 - Use a better model (e.g. extract noun phrases, use word embeddings, use a nonlinear classifier).
 - Use active learning to optimise how manual labelling time is spent.
 - Manually label using an expert.

*Survey of active learning strategies

<http://proceedings.mlr.press/v16/settles11a/settles11a.pdf>



Analytics + Foresight Hub

Giving Structure to Unstructured Healthcare Workforce Documents Using Natural Language Processing

Jerome Wynne

JeromeWynne@das-ltd.co.uk

Siôn Cave

SiônCave@das-ltd.co.uk