|  |  |
| --- | --- |
| Last updated: | December 2024 |

**JOB DESCRIPTION**

|  |  |  |  |
| --- | --- | --- | --- |
| Post title: | **Senior Data Scientist** | | |
| School/Department: | NIHR Evaluations, Trials and Studies Co-ordinating Centre (NETSCC), School for Healthcare Enterprise and Innovation | | |
| Faculty: | Medicine | | |
| Career Pathway: | Management, Specialist and Administrative (MSA) | Level: | 5 |
| Posts responsible to: | **Business Intelligence Architect – Digital Service Delivery** | | |
| Posts responsible for: | Data Scientist | | |
| Post base: | Office-based- Hybrid Working Options Available | | |

|  |
| --- |
| Job purpose |
| The senior Data Scientist will lead and be responsible for innovating with new methodologies and establishing NETSCC as a sector leader in utilising data science to more efficiently and effectively manage the research funding process, evidence the impact of investments, and communicate research outcomes. |

| Key accountabilities/primary responsibilities | | % Time |
| --- | --- | --- |
|  | **Machine Learning Development and Deployment**  Design, develop, test, evaluate, and deploy machine learning models and algorithms to solve complex business problems using structured and unstructured data.  Develop and deploy a wide range of machine learning models, including those based on natural language processing (NLP), to address business challenges such as:   * Summarising progress reports to enable efficient review and identification of potential issues. * Analysing publications and other reported outcomes to highlight and summarise key impacts. * Categorising research projects into priority areas based on criteria such as disease area (e.g., Dementia). * Projecting programme portfolio numbers and financial planning.   Provide expert Advice on developments in Artificial Intelligence such as Large Language Models (LLMs), and other cutting-edge technologies and their implications for the business | 40 % |
|  | **Data Infrastructure:**  Be responsible for the development and implementation of robust data pipelines, ensuring data integrity, quality and accessibility for analysis and model training.  Lead on the maintenance and improvement of the organisational data infrastructure, ensuring data is accessible to stakeholders  Use independent judgement and specialist knowledge to manage medium to long term projects/developments and deliver effective and efficient solutions to meet stakeholder needs. | 20 % |
|  | **Model Integration and Usability**  Develop interfaces (APIs, web applications, visualisations) to expose machine learning models and insights to applications and end-users.  Use professional knowledge to identify and establish standards and best practices for documenting code, processes, and methodologies. Be innovative in the approach to problem solving, embedding a culture of continuous improvement, and providing oversight of the development and review of policies and standard operating procedures to ensure an ethos of quality assurance and excellence. | 20 % |
|  | **Stakeholder Engagement**  Foster and build close working relationships with key and often senior internal and external stakeholders to collaborate on strategic projects and priorities Liaising with the wider management team where appropriate.  Communicating resultseffectively to technical and non-technical stakeholders at various levels of seniority, both internally and externally.  Represent/promote NETSCC as a sector leader in Data Science and effectively work with and influence colleagues across the wider NIHR. | 10 % |
|  | Deliver and participate in learning and development activities and programmes across the NIHR and externally on specialist skills and/or function. | 5 % |
|  | Any other duties as allocated by the line manager following consultation with the post holder. | 5 % |

| Internal and external relationships |
| --- |
| Collaborate with senior managers and colleagues within NETSCC, the School of Healthcare Enterprise and Innovation and the University, and the National Institute for Health and Care Research (NIHR). |

| Special Requirements |
| --- |
| Post holder may be required to undertake planned UK and International travel; to attend meetings, events or conferences with occasional overnight stays. |

**PERSON SPECIFICATION**

|  |  |  |  |
| --- | --- | --- | --- |
| Criteria | Essential | Desirable | How to be assessed |
| Qualifications, knowledge and experience | Skill level equivalent to achievement of a professional qualification or postgraduate degree in a relevant field such as mathematics, data science, computer science or related discipline  Extensive experience using Python (preferred) or R for data science projects with libraries like NumPy, SciPy, pandas, scikit-learn, transformers, PyTorch, and TensorFlow  Expertise in software development/MLOps best practices, including Git, Continuous Integration/Continuous Deployment, model tracking, model governance, and managing multiple models in different production contexts.  Experience with cloud computing platforms (GCP preferred) such as Google Cloud Storage, BigQuery, and Vertex AI.  Good understanding of deploying and serving models securely for either batch processing or invocation as a RESTful API endpoint.  Knowledge of open-source repositories of NLP models such as Hugging Face or SpaCy.  Experience in data engineering and wrangling using SQL  Proven project management skills.  Able to apply experience and awareness within specialist field.  Able to appreciate NETSCC priorities and to apply these in managing work outcomes. | Expertise in data visualisation using libraries like Matplotlib, Plotly, ggplot, or tools like Power BI, Qlik Sense, Looker Studio, or Tableau.  Use of integration tools such as SSIS or ADF.  Understanding of ‘off-the-shelf’ solutions like Alteryx.  Strong understanding of statistics and statistical learning.  Expertise in deploying or writing bespoke algorithms.  Track record of contributing to open-source projects. |  |
| Planning and organising | Able to plan and manage major new projects or significant new activities, ensuring plans complement broader organisational strategy.  Demonstrated experience delivering end-to-end machine learning projects, including natural language processing (NLP) projects, encompassing scoping, programming, testing, deploying, and monitoring.  Demonstrated ability to identify, scope, and deliver impactful data science projects. |  |  |
| Problem solving and initiative | Able to identify broad trends to assess deep-rooted and complex issues.  Able to apply originality in modifying existing approaches to solve problems. |  |  |
| Management and teamwork | Able to manage team dynamics, ensuring any potential for conflict is managed effectively.  Able to provide expert guidance and advice to colleagues to resolve complex problems.  Highly motivated self-starter with the ability to work independently and collaboratively. |  |  |
| Communicating and influencing | Able to persuade and influence in order to foster and maintain relationships.  Able to resolve tensions and difficulties as they arise.  Demonstrated ability to translate technical terms and concepts for non-technical audiences |  |  |
| Other skills and behaviours | Demonstrate behaviour that is consistent with the HEI values: Collaboration, Delivery, Knowledge and Excellence.  Demonstrate flexibility in approach to work and professional interest in the work of NETSCC |  |  |
| Special requirements | Occasional requirement to work and stay away from Southampton |  |  |

**JOB HAZARD ANALYSIS**

**Is this an office-based post?**

|  |  |
| --- | --- |
| Yes | If this post is an office-based job with routine office hazards (eg: use of VDU), no further information needs to be supplied. Do not complete the section below. |
| No | If this post is not office-based or has some hazards other than routine office (eg: more than use of VDU) please complete the analysis below.  Hiring managers are asked to complete this section as accurately as possible to ensure the safety of the post-holder. |

## - HR will send a full PEHQ to all applicants for this position. Please note, if full health clearance is required for a role, this will apply to all individuals, including existing members of staff.

|  |  |  |  |
| --- | --- | --- | --- |
| **ENVIRONMENTAL EXPOSURES** | **Occasionally**  (<30% of time) | **Frequently**  (30-60% of time) | **Constantly**  (> 60% of time) |
| Outside work |  |  |  |
| Extremes of temperature (eg: fridge/ furnace) |  |  |  |
| ## Potential for exposure to body fluids |  |  |  |
| ## Noise (greater than 80 dba - 8 hrs twa) |  |  |  |
| ## Exposure to hazardous substances (eg: solvents, liquids, dust, fumes, biohazards). Specify below: |  |  |  |
| Frequent hand washing |  |  |  |
| Ionising radiation |  |  |  |
| **EQUIPMENT/TOOLS/MACHINES USED** | | | |
| ## Food handling |  |  |  |
| ## Driving university vehicles(eg: car/van/LGV/PCV) |  |  |  |
| ## Use of latex gloves (prohibited unless specific clinical necessity) |  |  |  |
| ## Vibrating tools (eg: strimmers, hammer drill, lawnmowers) |  |  |  |
| **PHYSICAL ABILITIES** | | | |
| Load manual handling |  |  |  |
| Repetitive crouching/kneeling/stooping |  |  |  |
| Repetitive pulling/pushing |  |  |  |
| Repetitive lifting |  |  |  |
| Standing for prolonged periods |  |  |  |
| Repetitive climbing (ie: steps, stools, ladders, stairs) |  |  |  |
| Fine motor grips (eg: pipetting) |  |  |  |
| Gross motor grips |  |  |  |
| Repetitive reaching below shoulder height |  |  |  |
| Repetitive reaching at shoulder height |  |  |  |
| Repetitive reaching above shoulder height |  |  |  |
| **PSYCHOSOCIAL ISSUES** | | | |
| Face to face contact with public |  |  |  |
| Lone working |  |  |  |
| ## Shift work/night work/on call duties |  |  |  |