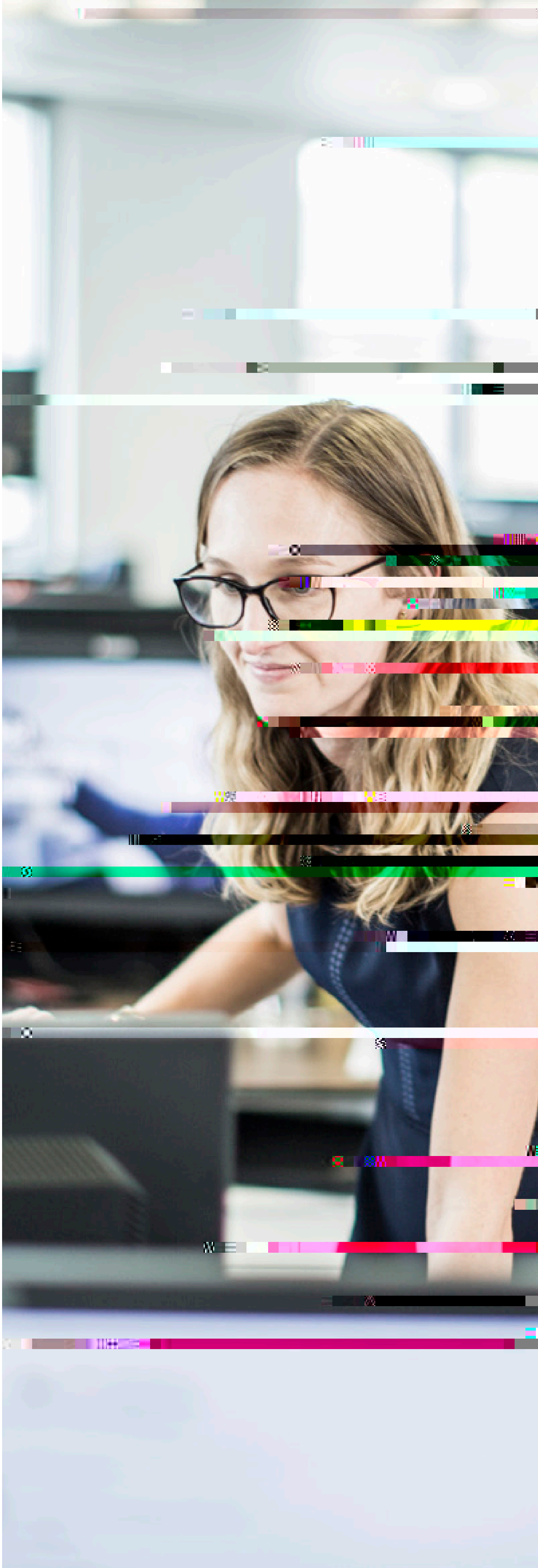
element

Digital Engineering

SUMMARY



DEFINITIONS

The following definitions are used in this document:

Term	Explanation
DS	Data science

A number of case studies are presented which were developed from interviews with industry data science experts.

CASE STUDY

Industry: Nuclear Services

Location: UK

Summary:

The organisation possess a wide variety of experience encompassing engineering (civil, structural, and aerospace), applied mathematics, physics, chemistry, and statistics. The bulk of modelling undertaken is physics-based modelling, i.e. CFD and FEA. Data science is a more niche area, and is undertaken primarily by statisticians and engineers.

No one actually has the job title of 'data scientist'. Much of what is referred to as 'data science' is just using applied statistical methods to gain insights on data. Machine learning is an area in which there is currently a lot of activity – it is thought of by some as a silver bullet that can solve every problem, but that is not often the case. A lot of the work undertaken by the Statistics team involves using statistical analysis to inform plant decisions and most of the machine learning falls into this category. Data sets are supplied to the team and they search for trends or engineering correlations that can be used to inform plant decisions.

Methods:

In general, software tools are a mixture of off-the-shelf commercial tools and internally developed tools. These tools are usually developed by engineers who are competent in coding rather than software developers.

Tools used for data science include: Python (Pydata, tensorflow), Fortran, C++, MATLAB, R (used by statisticians).

Team:

Data science projects are undertaken by cross-functional teams which may take expertise from any of the areas the organisation specialises in. Statisticians will almost always be involved where advanced statistical methods are required – there are approximately 10 people (mostly statisticians) who are interested in data science, ML, and AI. Project teams are usually small, consisting of 3 people working over short time periods in a consultancy role. The selection of team members is very much dependent on the nature of the work.

A 'virtual working group' is used to share and grow knowledge for those interested in data science.

Challenges:

A number of challenges were outlined which broadly fit into the following categories:

- **Culture:**

The nuclear sector is slow to adopt new technologies. Proven technologies with many verified examples are usually desired for confidence. Most organisations rely on robust reliable technologies. This doesn't present a major obstacle when using AI and ML as these both fall into a general category of statistics, but there is a reluctance to use newer technologies such as computer vision.

Some people are opposed to new technologies as they believe their roles will become redundant as new technologies are adopted - An internal survey on robotics demonstrated concerns by employees that they will be replaced by robots. When showing proof of concept it is possible to achieve 70% accuracy, but experience has shown customers wanting a 1 in 1 million error rate. This is totally unrealistic and exceeds the current system requirements.

- **Technology:**

The quantity and quality of data available often obstructs data science. This is sometimes a result of a lack of forward planning. When systems were installed, it was not expected that they would be continuously monitored and analysed using digital methods as there was no foresight of any additional value to be taken from data science.

There are challenges using NN as it can be difficult to see why the model has made a particular decision. Using regression models it is easier to visualize the results. The lack of clarity when using NN

CASE STUDIES

Generally, an agile sprint or Kanban approach is used for data science activities.

A number of tools are used for data science, although they are quite restrictive on software use. These tools include: R for statistics, Python, Jupyter Notebooks for sharing knowledge and growing capability, H2O.ai for machine learning (highly recommended), Orange Canvas, Apache airflow (for data pipelines).

For big data management the following tools are used: Oracle, SQL, Elastic surge.

Team:

Data engineers look after databases and undertake activities where data is structured, cleaned, and stored. Data scientists cover everything from CPU architecture to visual story telling. Corporate IT is involved to some degree with data science. A team in India is used which specialise in IT and Mathematics to undertake data engineering activities and some data science. Generally, engineers with lots of software knowledge are preferred for data science projects as they are problem orientated. It has been found that pure data scientists, while competent with maths/ IT and using algorithms, do not understand the context of the data science application, i.e. the relationship between the data and function. An example of this was an aerospace company looking at anomalies in their data caused by step climbs (changes in aeroplane altitude). This was seen as an anomaly to the data scientist but not to engineers with domain expertise.

Challenges:

The organisation as a whole is generally paranoid when dealing with IT and information security and this is usually the biggest obstacle to any data science project. For example, the use of open-source software is difficult to get approved by IT.

Another major challenge is the high risk of failure associated with data science projects– as mentioned above 33% of projects do not progress through the initial stages.

Additionally, the use of agile planning or Kanban methods for data science projects is unfamiliar to traditional corporate working, which means there is a culture difference in how things are done. This can sometimes represent a challenge.

Successful Data Science Use Cases::

R Ca A a .

Most data science is correlation and with large enough data sets insights can be taken. This principle has been applied to root cause analysis. When a failure occurs in a gas turbine engine, large data sets of engine operating data have been used to create a 'stop-motion' sequence of events that lead to the failure. This has been used to highlight previously undiscovered failure mechanisms and therefore reduce the risk of future failures. This method can be applied for almost everything, from finance to logistics, but big data is needed so records cannot be incomplete as they often are.

There are challenges using NN as it can be difficult to see why the model has made a particular decision. Using regression models it is easier to visualize the results. The lack of clarity when using NN makes it more difficult to apply in a safety critical environment.

Hardware issues, such as HPC requirements for data science and lack of access to GPU.

- **Security:**

Security poses many challenges when working with data in the

CASE STUDIES

operational cost savings through the use of data, data analytics, and machine learning in the heavy-end of production. The team initially consisted of just 3 members; process/plant engineers and technical graduates from backgrounds in engineering and pure mathematics. Initial training was given to the team to build

CASE STUDIES

This project developed a model to predict flu gas emissions produced by furnaces to better understand pollutant levels and meet statutory requirements. The model uses various process variables and sensor data as inputs. Predicted emissions from the model are supplied to a regulatory body and are accepted as accurate; this gives the plant license to operate. The advanced nature of predictions reduces operational risk as the organisation are able to plan ahead when high emissions are predicted.

Unsuccessful Data Science Use Cases:

S a a a

A works area suggested that data science could be used to track and predict the chemical composition of scrap metal to be used in the BOS plant, thus improving understanding of the operating conditions in the BOS vessel. After some review by AA, it was deemed that data science was not an appropriate tool as the data collected was of poor quality and used very small sample sizes, with data from weekly chemical samples allowing anomalies to be over represented. Another factor limiting success was the expertise in the works area – there was a lack of general knowledge/understanding about data science practices.

CASE STUDY

Industry: Technical Consultancy

Location: UK

Summary:

The organisation provides technical consultancy for development of data capture and visualization applications and has extensive expertise using data science methods in a range of industries, including nuclear and pharmaceuticals.

Challenges:

A number of challenges were outlined which broadly fit into the following categories:

C

The biggest blockers are typically in managing expectations vs reality. As people, we see some logic behind a task and infer that a machine must be capable of carrying out that task. An example of this concerned a predictive signal for a project on a gas turbine engine. It was identified that the blades were cracking due to vibration. Due to the abundance of operating data available, it was thought the data could assist in root cause analysis. However, upon review of the data, a solution was not clearly visible and a different approach was needed.

Additionally, if the person leading the project isn't experienced in data or isn't able to reformulate their approach based on the data presented, this can present a major challenge in undertaking data science. Preconceived judgment may inform solutions if new methods are not embraced or trusted.

T

A major technological challenge is when the technology being used does not support the data correctly. There is a myriad of database technologies and depending on how they are being used and architected, they either can or cannot support different kinds of machine learning. For example, a simple recursive model can be used on many types of databases but for a deep learning model the data has to be extracted and used in a different way. That can create blockers to training models and accessing data.

More challenges arise when extracting value from data. Using off-the-shelf pre-trained algorithms, it is possible to get 80% of the value of data – the last 20% of value can be disproportionately expensive to realise. If people are not aware of this, they may hit that asymptotic curve where they think they are close to their acceptance criteria but the amount of effort needed to actually reach this criteria may increase exponentially. This can sometimes be overcome by having people involved in the process, rather than using purely computational methods.

Successful Data Science Use Cases:

Nuclear:

Pa b NLP

Due to the length of time machinery is in service in the nuclear industry parts become obsolete in the sense that they are no longer manufactured. The use case for this project was to investigate how to source functionally equivalent parts (pipes, valves, etc.) across different nuclear sites. When new parts arrive, the stock keeper would receive the part and a datasheet and would type in a description of the part into a database. The data was not gathered for any other purpose other than the stock keeper knowing what they had in their database. The data

ele e t