Case Study Title: Capturing Quality Information Using Artificial Intelligence

Company Overview | SISK | johnsiskandson.com

John Sisk & Son Ltd (Sisk) is an innovative engineering and construction company employing over 1,800 people in Ireland, the UK, and Europe. Sisk has the track record, scale, and capacity to successfully undertake large, complex, multi-disciplinary programmes, and we are recognised by our global customers as world leaders in sustainability and safe delivery.

Sisk is a progressive business and Ireland's No 1 ranked provider of construction services. Operating since 1859, we have built many iconic buildings and landmark pieces of infrastructure. Our continued success is due to:

- Our ability to collaborate with customers and supply chain to provide technical and delivery solutions in an open and can-do way.
- Safety, innovation, quality, efficiency, and value are integral to everything we do.

We deliver projects in key sectors such as Data and Technology, Pharmaceutical and Life Sciences, Infrastructure, Transportation, Healthcare, Commercial, Residential, Retail, Industrial, Leisure, Education, Water, and Energy.

Author

Alex Coles

Author

Ugo Madu

Overview & Background to the Lean Initiative

Using an off-the-shelf product from Microsoft, Sisk have developed an automated Artificial Intelligence (AI) tool which captures data from a variety of forms (such as concrete cube test certificates and delivery notes). Previously this information would have been manually transcribed from forms into Excel trackers, often by engineers. By automating the process, Sisk have both freed up time for staff to undertake more productive work and have also improved the quality and timeliness of the data through the reduction of human error.

The tool works by training a machine learning model, by identifying the required data fields on a set of sample forms (generally 5-10) – after which it can automatically detect and extract the data and store it as required. This extraction process is facilitated by the Microsoft AI Builder tool alongside an automated Microsoft Flow.

The tool was developed by Ugo Madu, one of Sisk's Graduate Engineers with an interest in Data Science. After being asked to undertake the manual transcribing of hundreds of forms into Excel trackers, he undertook independent research into developing a better way of undertaking the task and developed this tool under the guidance of Alex Coles, founder of Sisk's Data Academy.

Lean Initiative Undertaken – Lean Thinking, Tools, Techniques

Project Background

Every site at Sisk receives a large quantity of documents (such as invoices, delivery tickets, testing certificates) and each of these carries valuable information that needs to be captured for further analysis. At one of Sisk's major highways projects in the UK, a study was undertaken to improve the data extract process from two types of forms. The first of these, Concrete Cube Test Certificates are required to prove that concrete on site has met the required strength levels, with a sample of concrete taken from each concrete pour on site and certificates issued for showing the strength after 7 days, 28 days and 56 days. These certificates form a vital part of our Quality Assurance, as they confirm that the concrete has reached the designed levels of structural performance. They also require coordination from an administrative perspective – as a referencing system must link the certificate to the pour and each pour will have eight certificates that are issued at different times.

The second form, Stone Delivery tickets, provide vital information from a commercial perspective and show how much gravel and stone had been delivered to a project. On highways projects, there is always a risk of wastage of stone, so capturing the amounts that have been delivered provides a level of control and the ability to compare planned vs actual quantities.

Both sets of forms would arrive electronically at the project, however there would also be a range of documents that would be issued as a hard copy, particularly in the case of delivery tickets.

Current Methodology – Manual Processing

Currently, documents processed on site come in different types of formats, which all require slightly different ways to be processed. Each document is processed by having the relevant information manually read and populated onto a Microsoft Excel spreadsheet, either through copying and pasting or re-entering the content. This process is time consuming and prone to human-error, as the information is generally "double handled". Depending on the

number of relevant data points each document has, the time taken to process a single document can range between 20 to 80 seconds, with longer durations if the form is a hard copy that needs to be scanned and processed.

The inputting process required a large amount of time from skilled individuals who could have been assigned to more productive work, with approximately 800 documents per month being processed on a single sample site. In addition, this administrative-heavy task would often result in a backlog, meaning that the information was not as timely as it should have been. Such processes are undertaken across the majority of our projects, and it was therefore apparent that providing a more lean and automated process would remove a significant amount of tedious, unrewarding work and free up thousands of hours of staff time across the business.

Improved Methodology – Automated Processing

An improved methodology was developed using a suite of data processing tools available through our Microsoft licence. The core of this was the Microsoft AI Builder tool, in which a Machine Learning model is trained to understand where on a form the information is and then an in-built "Optical Character Recognition" (OCR) tool detects the content of the fields and extracts the data.

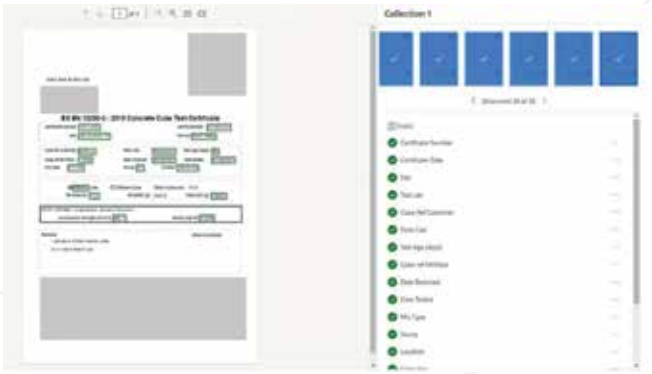

Figure 1: A screenshot from the AI Builder tool, showing a sample form on the left hand side and the designated fields on the right hand side

While the theory sounds complicated, the process of using the tool is extremely straight forward. The first step is to identify and list the required fields to be extracted from the form. In the case of a concrete cube certificate, this would be items such as "date", "strength", "location" etc. Once this list has been produced, a sample set of at least 5 PDF forms is uploaded into the AI Builder and the user manually identifies where on the form the various fields can be found. The principle of machine learning is that a model is trained to identify the information, so the more variability in the form design, the more forms that are required to train the model. In the case of concrete cube tests on a standard template, the model can achieve a high level of accuracy with relatively few forms.

Once the model has been built, a Microsoft Flow automation is developed, which facilitates an end-to-end process connecting the various tools of the Microsoft suite as follows:

1. Microsoft Outlook: Email arrives into a designated inbox. The tool recognises the email address as being from the testing company and recognises the attached cube certificate.
2. Microsoft OneDrive: The attached PDF is automatically transferred to a designated folder on OneDrive, where the system recognises that a new file has been uploaded.
3. Microsoft AI Builder: The trained machine learning tool 'reads' the content of the file and automatically extracts the data fields.
4. Microsoft Excel: The extracted data is stored in an Excel tracker as a new line item, with a column for each of the identified fields.
5. Microsoft Outlook (future functionality): Once the data is in the tracker, it should be possible to set up an automated email alert to the project engineers, if a cube test is identified as having failed. This is currently being explored by the Sisk team.

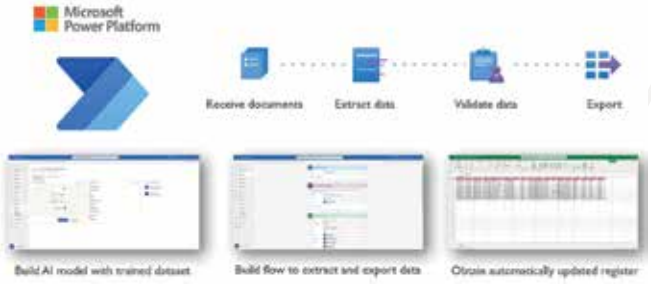

Figure 2: Overview of the data extract process

While the initial setup time for this process may take up to an hour, once the tool is set up it will be fully automated and require no user input, except to monitor the results and action any non-conformances associated with failed tests. In addition, the power of the machine learning tool and optical character recognition is that it should be possible to reach a point where, having been trained with enough variety of different certificate formats, it should be able to recognise and extract data from any certificate, as long as similar data fields are present. Sisk have not fully explored this, having only targeted forms on a single sample project. However, this is an area of proposed future functionality.

Lean Initiative Improvements & Impact

The tool that has been developed at the pilot project has been a great success and has proven a concept that we will be looking to roll-out further. The tool frees up time for junior engineers to be away from their computers and out on site where they are best positioned to learn, develop and grow. The Quality Department have also been keen to see this pilot expanded to more sites, as in the past the data in this area has often been out of date and at times unreliable, with the tool alleviating both issues.

There are a few potential risks that should be managed and identified. The most important of these is that the automation and lack of human input leads to the data being ignored, at least the manual approach will see an engineer directly interacting with the data. In addition, if the automation process were to fail (for example if there was a significant change in format that the Machine Learning model could not work with), this could result in incorrect data being extracted. As a result, it is important to embed an auditing process, which sees the results reviewed and analysed at regular intervals.

In terms of future functionality, Sisk are looking to develop additional features, which facilitate this auditing process. For example, the automatic sending of emails when cubes have not reached the desired strength. In addition to this, forms with greater variability are being explored. For example, invoices and a wider range of delivery tickets. This could also be combined with a "picture" functionality, whereby the recipient of a hard copy of the form can take a photo and automatically see the document reviewed and catalogued.

In addition to the benefits of the tool itself, this initiative has highlighted an additional consideration. The tool was developed by a graduate engineer on one of our projects who had no previous formal training in data science, but who had learnt a few key overall principles and was able to teach himself techniques through freely available online videos. This has shown the potentially transformative power of empowering more junior members of our project teams to develop solutions to problems that they see in their day-to-day roles. While a business-wide roll-out of this tool will require the rigour and project management of our IT Department, Sisk are looking to gather and implement more grassroots change initiatives through the re-establishment of a Project Management Office (PMO) and the founding of an information "Data Academy" community.

Summary and Lessons Learned

The use of AI is becoming more apparent in the construction industry and early adoption is key to closing the gap of labour productivity relative to other industries. Harnessing this technology is paramount to digitising the industry even further and this is yet another great example of how every small step counts.

The benefits that arose from the solution to this challenge have had a positive impact to many stakeholders, primarily due to the time-saving aspect of the solution. Improved quality of captured data and increased productivity are the two main benefits of this initiative, but it has also opened the door to research into more ways of applying AI and document processing in the construction industry.

Further functionality and integration with other systems using this application are currently being explored at Sisk in order to carry on streamlining more processes throughout the business. At Sisk we continue to disrupt the industry with the use of technology to improve, adapt and overcome new challenges we face, enabling Sisk to remain competitive and deliver projects faster and safer all the while maintaining quality standards.