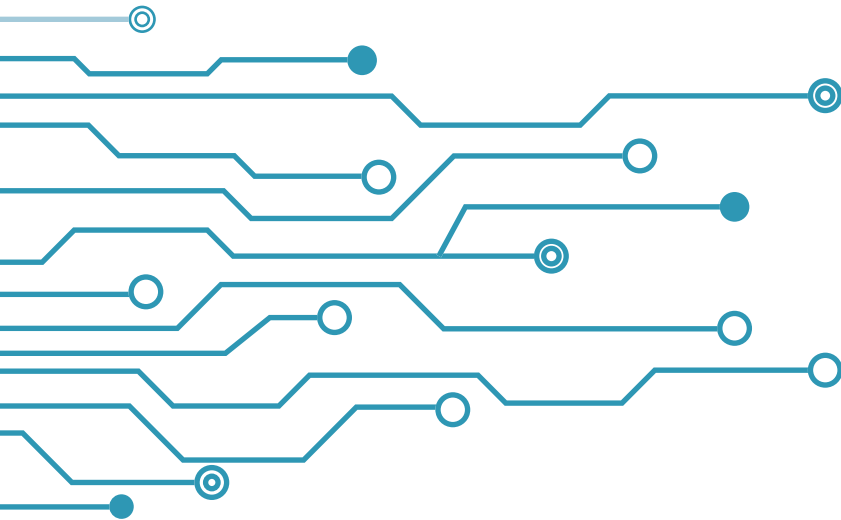# AI Drives Improvement in Business Function Identification

Be it a government body planning critical infrastructure, a business trying to understand its portfolio, or an insurer looking to identify risk, accurate business function identification is critical. Now, artificial intelligence is driving change.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

WHY ACCURATE BUSINESS FUNCTION IDENTIFICATION IS IMPORTANT

Businesses and government bodies have the need to identify accurate business functions in many areas such as market or customer segmentation, for compliance purposes, or risk management to name a few.

This white paper explores these in some detail and provides an overview as to how Dun & Bradstreet, along with its artificial intelligence (AI) Evolution AI partner, are using modern approaches to an old problem.

INDUSTRY CLASSIFICATION - A LITTLE BIT OF HISTORY

The Standard Industry Classification (SIC), is a system for classifying industries by a four to eight digit code. Established in the United States in 1937, it is still used by government agencies to classify industry areas. The SIC system is also used by registries in other countries, e.g. by the United Kingdom's Companies House (1).

For markets where registry data isn't readily available, a business classification has to be sourced.

GLOBAL BUSINESS NEEDS

It becomes difficult when reviewing data across physical markets, as SIC is NOT the global standard. The best known classification systems in addition to SIC are:

– NACE REV 2 (2006), which is a revision of ISIC Rev 1 which was developed in 1961 (2)

– NAICS North America Industry Classification System developed in 1997 (overtaking the use of the SIC in the US)

Furthermore, there isn't a one-to-one mapping from one system to another, which can result in a loss of accuracy and granularity, as well as create difficulty in the ability to analyse and utilise data effectively.

SIC STATISTICS

Dun & Bradstreet provides SIC, NAICS and NACE throughout its product portfolio. Looking at the US product data in a little more detail, its foundation is the 1987 US SIC which has 1,005 code values. However, to improve its SIC data Dun & Bradstreet carried out independent research and added a proprietary extension to the 1987 SIC. This then created a new industry code schema with 18,785 code values, the SIC 8, vastly improving granularity.

MODELLING

Where a code cannot be assigned locally, Dun & Bradstreet applies an industry code model to assign one. The industry code is inferred using company names and business trade styles to discover what a company's line of business is, so that even before AI is applied, data science has been utilised to bolster the foundation the classification is built on.

Even though this development improves business classification, it is semi-automated and keywords without context can in some cases introduce inaccuracy. For example, a business consultancy called the 'Fish Practice' was incorrectly assigned '47230 - Retail sale of fish, crustaceans and molluscs in specialised stores,' as 'Fish' was heavily populated throughout web content -- not a great improvement.

MOVING FROM THE 1930S TO MODERN TIMES
Applying AI to improve assignment and precision

Although the modelling of SIC generated improvements, Dun & Bradstreet was not content with the anomalies. So, with customer consultation, a deep review of business needs, and a desire to utilise cutting edge technology to open the aperture around business classification, a decision was made to run a pilot to see whether AI could, in fact, improve SIC assignment and precision.

The objective was to:

– Improve coverage – increase SIC coverage

– Improve accuracy – increase the level of granularity for assigned SICs and to minimise unclassified or miscellaneous business services classification

– Improve depth – increase the depth of SICs assigned to a business from one or two SICs up to five

THE PROCESS

Working with Evolution AI, Dun & Bradstreet ran a proof of concept using machine learning 'Neural Modelling' taking basic Company House ID information, SIC UK2007 descriptions, and the web.

The technology is underpinned by a new technique known as deep learning. Deep learning uses complex mathematical structures called neural networks to represent language and concepts. Whilst previous generations of Natural Language Processing (NLP) technology suffered because they viewed documents as jumbles of keywords, rather than as whole sentences, deep learning truly understands context and nuance.

The many pieces of information that Dun & Bradstreet has on individual companies, both proprietary and open source, can be read and digested by the AI system. By combining all of the information, a well-rounded, trackable decision can be made on each company in milliseconds. In contrast, a human analyst, operating under typical time constraints, relies on a small number of data points to make their decision, as the volume consumed by AI is unmanageable.

After all of the evidence has been weighed, the system produces a confidence score which reflects the confidence of each decision. Critically, the system 'knows when it doesn't know.' This is very important, since the human operators can now direct their energy to the most challenging cases, better suited to the human ability to understand complexity and nuance.

INPUT DATA FOR THE PROOF OF CONCEPT

The input data sources were:

- 3.5 million records from the United Kingdom's Companies House
- 2 billion+ websites (in index form)

PROJECT SCALE

A web page contains 440 words on average, and human beings read the text at the rate of circa 200 words per minute (4). It would, therefore, take a single person 4,560 working days to read one million web pages, or twenty years and four working months. AI systems, however, read one million websites in around 8 minutes; the whole website, not just the homepage.

THE HUMAN IN THE LOOP

Aligned with this is a continuous improvement programme and a quality assurance tool (QA), so that the process is always learning and the inputs help 'teach' the machine with 'human in the loop' processes.

For each record, the SIC codes provided can be reviewed and validated using the 'about us' section of the company website or other sources of content utilised for assignment.

So far, the process has delivered 6.7 million UK SICs, either as newly assigned or verified as below, and the programme is being extended to the US database.

When compared against the objectives, it looks like this:

- Improved coverage = 4.3 million additional SIC codes
- Improved accuracy = 1.2 million verifications
- Improved depth = 1.5 million with advancement in precision

This is clearly an improvement.

WHAT'S NEXT

As the model evolves the programme has included more 'implicit' datasets, such as transactional data and social media feeds, which using machine learning has enabled identification of industrial activity through implied insights such as relationships.

IN SUMMARY

Data science has shown that it can help improve SICs and ultimately the data Dun & Bradstreet provides to customers. This is just a start though: as big data processing and more open data become accessible, further improvements will be seen. The process will ultimately expand the view of what a business does for a far more targeted sales and marketing approach and portfolio risk segmentation.

The use of AI will only increase, as the technology improves and analysts continue to struggle to make sense of the multitude of data sources that AI can process easily to make a well-rounded decision.

WHY ACCURATE BUSINESS FUNCTION CLASSIFICATION IS IMPORTANT

Consider the following problem: you are a local government with infrastructure responsibility for a number of business parks. A developer has submitted plans to enlarge a park, purchasing an existing brownfield site that is adjacent to one of the existing parks. So far, so good. However, for the new park access, the developer plans to utilise the existing road network.

The existing park residents want to be sure that the type of business is complementary. The existing park is mainly made up of science-based organisations, and as such the science park management has lobbied the local authority to ensure it meets their demands in exchange for access rights.

What do you think is the best way to describe complementary businesses? Well, a legal document outlining the industry classifications that would sit within a science-based category; and a universally understood codification that identifies a business function that cannot be manipulated.

Does one exist? Thankfully yes (and not just one, depending on the country where the business park and planning authority is located). More to come on this in the Little Bit of History section.

The government isn't the only type of organisation that is in need of adequate understanding around business function. On a commercial level, most businesses utilise classifications in some way. For example:

GOVERNMENT	GOVERNMENT & COMMERCIAL
Disaster prevention risk prediction, identifying areas of risk, e.g. chemical plants	Cross market analysis needs an international standard
Disaster recovery for natural disaster aid, e.g. identifying and understanding impacted	Risk identification and management
Tax and licensing, e.g. which businesses are exempt or need to comply?	Logistics and planning for infrastructure projects
Identifying non-compliance, e.g. businesses that should be registered to trade	Supplier risk, where a supplier operates across business functions and an area is hit by poor trade conditions that could impact operations
Government statistics around the business universe, e.g. Department for Business and Industrial Strategy	Sales territory planning by business verticals
	Marketing portfolio segmentation and market sizing

If we break this down more specifically by business need and type of industry, the application may be slightly different, but there is still a demand for classification.

BUSINESS NEED	INDUSTRY EXAMPLE	IMPACT OF CLASSIFICATION
Market/customer segmentation	Packaging, hospitality, retail	Classification assists with business development and territory planning and/or expansion, and in real estate development.
Financial risk categories	Manufacturing, industrial services	Natural and other disasters can have a significant impact on the creditworthiness and viability of businesses.
Classification of customers	Insurance	Accurate classification of businesses allows for more exact insurance coverage and premium assessment.
Compliance	Financial services	Assists in the identification of businesses that must be registered or regulated at local/state/federal level.
Prospecting	Paper supplies/packaging	Prevents missed sales opportunities due to poor SIC classification.

Let's take a look at a commercial issue. One of the businesses most reliant on business function is the insurance sector. For underwriting, a very deep level of data is needed and Standard Industry Classification (SIC) just doesn't cut the mustard. For example, if an insurer is selling liability cover, the underwriter needs to know whether a carpenter works in shipbuilding or is a kitchen fitter, as the risks are rather different.

For commercial insurance the information collected is usually from the insured, i.e. it is self-reported, or sometimes reported via a broker. But, the loss adjuster needs to understand whether the self-reported description is accurate. So, how do they check?

Finally, the insurer also needs to be able to exclude businesses based on their function.

Dun & Bradstreet also works with less obvious examples, such as satellite navigation providers. It is easy to see where the likes of Google and Bing Maps get their content if you think of them as large online directories first. In this case, company address data is there to be utilised, but even then the business function is needed to highlight points of interest, such as the closest Indian restaurant.

INDUSTRY CLASSIFICATION - A LITTLE BIT OF HISTORY

The SIC is a system for classifying industries by a four to eight digit code. Established in the United States in 1937, it is still used by government agencies to classify industry areas. The SIC system is also used by registries in other countries, e.g. by the United Kingdom's Companies House (1).

REGISTRY DATA

When a business registers at a registry, they are asked to provide an Industry Classification Code. This is the 'as registered' SIC, and the important point to note here is that it is self-assigned, using a pre-defined list of categories and their resultant codes. With the fast pace of change, particularly in the technology sector, you can start to understand some of the complexities in maintenance and application.

> To really understand your customer base, understand business risk, or the many other business needs.

While a good starting point, the categories that exist are often not terribly useful to really understand your customer base, understand business risk, or the many other business needs we discussed earlier. For example, a common code 7299 is 'Miscellaneous personal services not elsewhere classified,' which doesn't reflect the business operation at all.

NON-REGISTRY DATA

For markets where registry data isn't readily available, a business classification has to be sourced. In these cases, Dun & Bradstreet will review a business at record creation, and the business will then be cross-referenced with other trusted sources, such as commercial telephone directories, to best determine its business sector.

Data modelling techniques are also used to refine this process, such as inferring from the name what the business may do, e.g. Swan Chinese Restaurant, to apply the relevant code. Prior to the application of artificial intelligence (AI), keywords from the business name and other sources have already been compiled and data already clustered using a multitude of sources to assume quality through plurality, and marketed by Dun & Bradstreet as SIC 8.

This is a proprietary process and Dun & Bradstreet has global coverage. However, not every record can be accurately classified using this system, as there is a reliance on the underlying data. To better understand SIC 8, the Swan Chinese Restaurant is a good example to demonstrate the need for more granularity, as the differentiation between 'restaurant' and 'take away' is somewhat blurred in the mainstream codification.
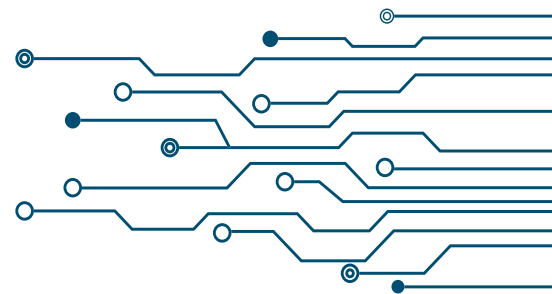
GLOBAL BUSINESS NEEDS

Waters become muddied further when companies wish to expand globally, and the need arises to categorise businesses in a number of markets across a variety of classification systems since SIC is NOT the global standard. The best known in addition to SIC are:

– NACE REV 2 (2006), which is a revision of ISIC Rev 1 which was developed in 1961 (2)

– NAICS North America Industry Classification System developed in 1997 (overtaking the use of the SIC in the US)

Furthermore, there isn't a one–to–one mapping from one system to another, which can result in a loss of accuracy and granularity, as well as difficulty in the ability to analyse and utilise data effectively.

When applying SICs, Dun & Bradstreet ranks the SIC in order of revenue generation against the company. For businesses engaged in more than one activity, industry codes are assigned for those functions that contribute, in general, to at least 10% of the company's revenue.

Things get a little more intricate where a company has subsidiaries. In this case, industry codes are assigned based on a company's own revenue generating activities as well as those performed by its direct and indirect subsidiaries. But, subsidiaries are assigned their SIC codes based solely on their own activities, unless they, in turn, have subsidiaries. Branch locations are coded based on the activity performed solely at the branch location.

To put this into a business context, consider a coffee shop that may also have another business operating at the same location, such as an ice cream shop; both affiliated with separate but well-known brands. If financials are available, then the SIC is assigned in order of the highest revenue producing activity of that business. However, when financials are absent, the assignment is done by inference.

This detailed application allows customers to understand the primary business focus. But where there are multiple SICs, the trick is to find the appropriate one for the user needs and this may not be the one that drives the most revenue as per the above explanation. This is where data planners come in.

> This detailed application allows customers to understand the primary business focus.

SIC STATISTICS

Dun & Bradstreet provide SIC, NAICS and NACE throughout the product portfolio. Looking in a little more detail at the US view, the foundation is 1987 US SIC, which has 1,005 code values. To improve SIC classification, Dun & Bradstreet carried out independent research and added a proprietary extension to the 1987 SIC. This then creates a new industry code schema which has 18,785 code values, the SIC 8, vastly improving the granularity.

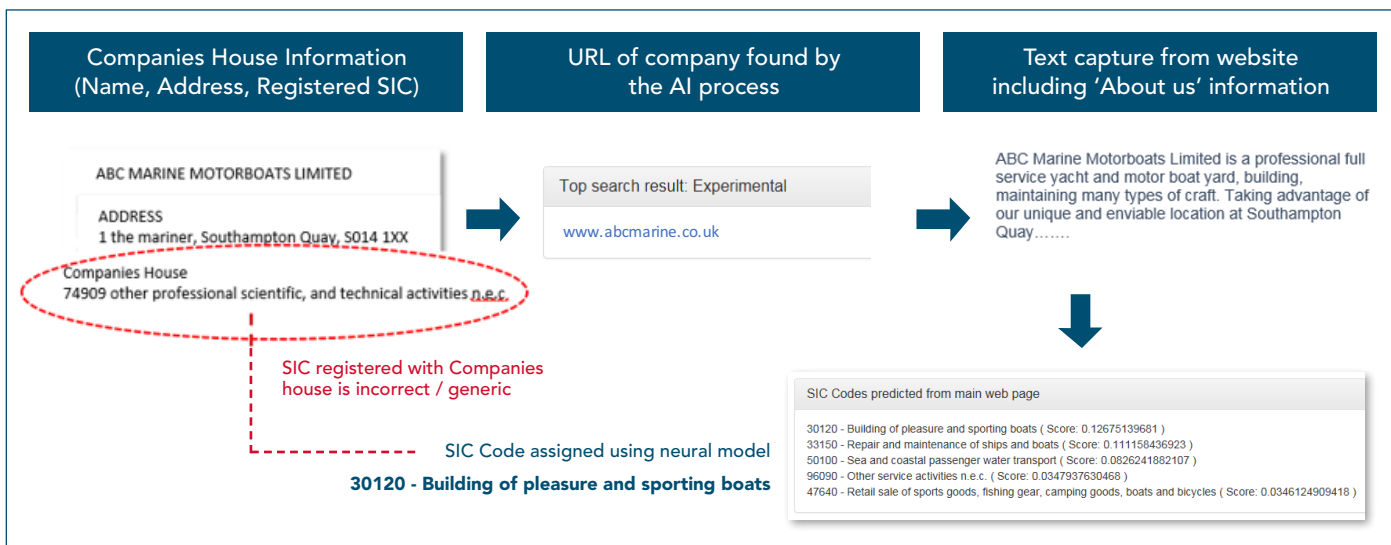e.g. SIC: 5812 = eating place SIC 8: 58120108 = Italian restaurant

– This standard allows customers to understand industry segmentation in one consistent code structure.

– NAICS and ISIC are not as detailed (NAICS has only 1,058 code values), as can be seen by the examples below: NB N.E.C = Not elsewhere classified.

ISIC3-1	ISIC Rev3-1 Description		
0111	Growing of cereals and other crops not elsewhere classified		
0111	Growing of cereals and other crops not elsewhere classified		
ISIC4	ISIC Rev4 Description		
0111	Growing of cereals (except rice), leguminous crops and oil seeds		
0111	Growing of cereals (except rice), leguminous crops and oil seeds		
NAICS	NAICS Description	SIC8	SIC Description
011130	Dry Pea and Bean Farming	01190100	Pea and bean farms (legumes)
011130	Dry Pea and Bean Farming	01190101	Bean (dry field and seed) farm
011130	Dry Pea and Bean Farming	01190102	Cowpea farm
011130	Dry Pea and Bean Farming	01190103	Lentil farm
011130	Dry Pea and Bean Farming	01190104	Mustard seed farm
011130	Dry Pea and Bean Farming	01190105	Pea (dry field and seed) farm

When you look at a large market such as the US, there are approximately 22.4 million active businesses, and Dun & Bradstreet has at least 1 SIC assigned to 93% of them. Whilst the accuracy of the SIC coverage is admirable, often there is only 1 SIC assigned to a business, and in some instances, it is challenging to assign a SIC with a strong level of granularity due to the fast pace of business evolution and the constraints of the classification standard.

The diagram below is an example of how rather generic classifications are applied when a business is registered and how the application of AI can dramatically improve the outcome.

Companies House Information (Name, Address, Registered SIC)	URL of company found by the AI process	Text capture from website including 'About us' information

ABC MARINE MOTORBOATS LIMITED

ADDRESS
1 the mariner, Southampton Quay, SO14 1XX

Companies House
74909 other professional scientific, and technical activities n.e.c.

SIC registered with Companies house is incorrect / generic

SIC Code assigned using neural model
30120 - Building of pleasure and sporting boats

Top search result: Experimental

www.abcmarine.co.uk

ABC Marine Motorboats Limited is a professional full service yacht and motor boat yard, building, maintaining many types of craft. Taking advantage of our unique and enviable location at Southampton Quay.......

SIC Codes predicted from main web page

30120 - Building of pleasure and sporting boats (Score: 0.12675139681)
33150 - Repair and maintenance of ships and boats (Score: 0.111158436923)
50100 - Sea and coastal passenger water transport (Score: 0.0826241882107)
96090 - Other service activities n.e.c. (Score: 0.0347937630468)
47640 - Retail sale of sports goods, fishing gear, camping goods, boats and bicycles (Score: 0.0346124909418)

MODELLING

Where a code cannot be assigned locally, Dun & Bradstreet applies an industry code model to assign one. The industry code that is inferred uses company names and business trade styles to determine what a company's line of business is, so even before AI, data science has been utilised to bolster the foundation the classification is built on.

These industry code models (as per the process diagram below) are used in a number of markets, including Australia, Canada, India, Israel, Ireland, Italy, Korea, Mexico, New Zealand, the United Kingdom, and the United States.
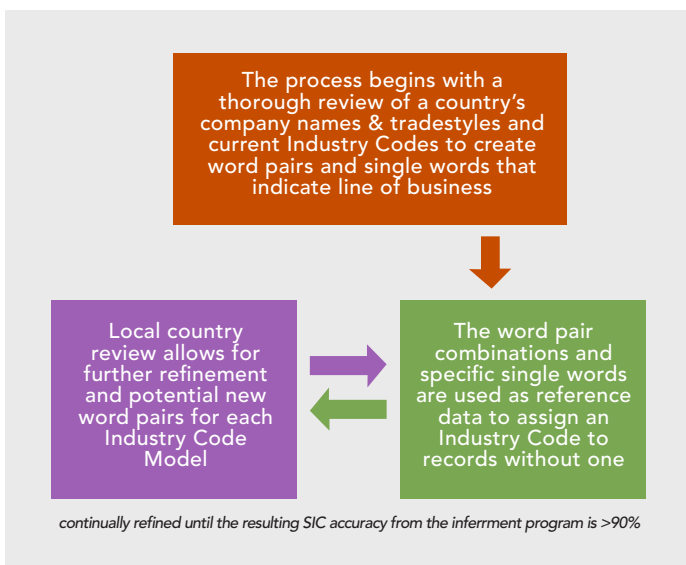
The ability to expand SIC coverage, improve its level of accuracy, and in many cases, allow for increased granularity for businesses often involved in a variety of industry verticals has allowed Dun & Bradstreet to better assist customers in the proper segmentation of their portfolio for risk management, prospecting efforts, and a general drive for new and incremental revenue.

Even though this development improves business classification, it is semi-automated and keywords without context can in some cases introduce inaccuracy, for example, a business consultancy called the 'Fish Practice' was incorrectly assigned '47230 - Retail sale of fish, crustaceans and molluscs in specialised stores,' as 'Fish' was heavily populated throughout web content -- not a great improvement.

The process begins with a thorough review of a country's company names & tradestyles and current Industry Codes to create word pairs and single words that indicate line of business

Local country review allows for further refinement and potential new word pairs for each Industry Code Model

The word pair combinations and specific single words are used as reference data to assign an Industry Code to records without one

continually refined until the resulting SIC accuracy from the inferrment program is >90%

"Keywords without context can in some cases introduce inaccuracy"

MOVING FROM THE 1930S TO MODERN TIMES

APPLYING AI TO IMPROVE ASSIGNMENT AND PRECISION

Although the modelling of SIC generated improvements, Dun & Bradstreet was not content with the anomalies. So, with customer consultation, a deep review of business needs, and a desire to utilise cutting edge technology to open the aperture around business classification, a decision was made to run a pilot to see whether AI could, in fact, improve SIC assignment with Evolution AI and precision.

The pilot approach allowed for the use of data-driven techniques to provide a confidence score around SIC coverage as well as determined where best to focus efforts for continuous improvement, closing any gaps in coverage.

The objective was to:

– Improve coverage – increase SIC coverage

– Improve accuracy - increase the level of granularity of assigned SICs and to minimise unclassified or miscellaneous business services classification

– Improve depth – increase the depth of SICs assigned to a business from one or two SICs up to five

Machine learning and natural language programming seek to improve search accuracy by understanding the searcher intent and the contextual meaning of terms as they appear in the searchable dataspace, whether on the web or within a closed system, to generate more relevant results. These systems consider various points, including the context of search, location, intent, variation of words, synonyms, generalised and specialised queries, concept matching, and natural language queries to provide relevant search results. (3)

The context could be understood as everything that surrounds a word, i.e. what gives it meaning. Context is the key component in the process.

> Context is the key componet in the process
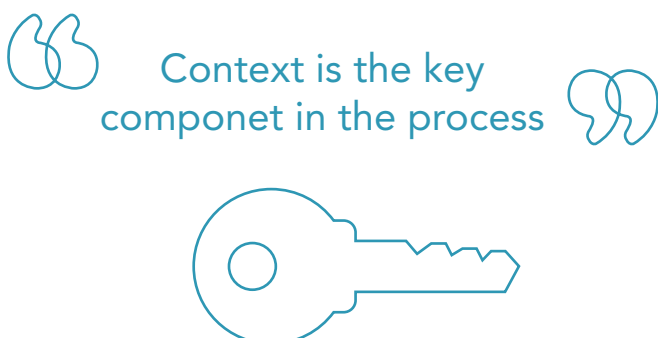
THE PROCESS

In consultation with Evolution AI, Dun & Bradstreet ran a proof of concept using machine learning 'Neural Language Modelling' using just basic Company House ID information, SIC UK2007 descriptions, and the web.

The technology is underpinned by a new technique known as deep learning. Deep learning uses complex mathematical structures called neural networks to represent language and concepts. Whilst previous generations of Natural Language Processing (NLP) technology suffered because they viewed documents as jumbles of keywords, rather than as whole sentences, deep learning truly understands context and nuance.

Learning the jargon of different industries is a critical step towards developing deep knowledge about industrial categories. By reading large quantities of text, the AI system recognises how words change their meaning in different contexts. For instance 'CNC' means 'Computer Numerical Control' in the context of engineering, but 'Civil Nuclear Constabulary' in the context of policing.

The AI system reads and learns autonomously, gaining knowledge from public domain pages (for instance, Wikipedia, company, social media, or government websites). When it identifies gaps in its knowledge in specific subjects, it actively seeks to learn more about those subjects.

Thus primed and armed with the official definition of the UK SIC Taxonomy (published by the Office for National Statistics), the system can make informed decisions about the industrial category for each company. Since the UK SIC Taxonomy runs to over one thousand pages, this problem is very well suited to an AI reading agent, rather than a human operator.

The many pieces of information that Dun & Bradstreet has on each company, both proprietary and open source, can be read and digested by the AI system. By combining all of the information, a well-rounded, trackable decision can be made on each company in milliseconds. In contrast, a human analyst, operating under typical time constraints, relies on a small number of data points to make their decision.
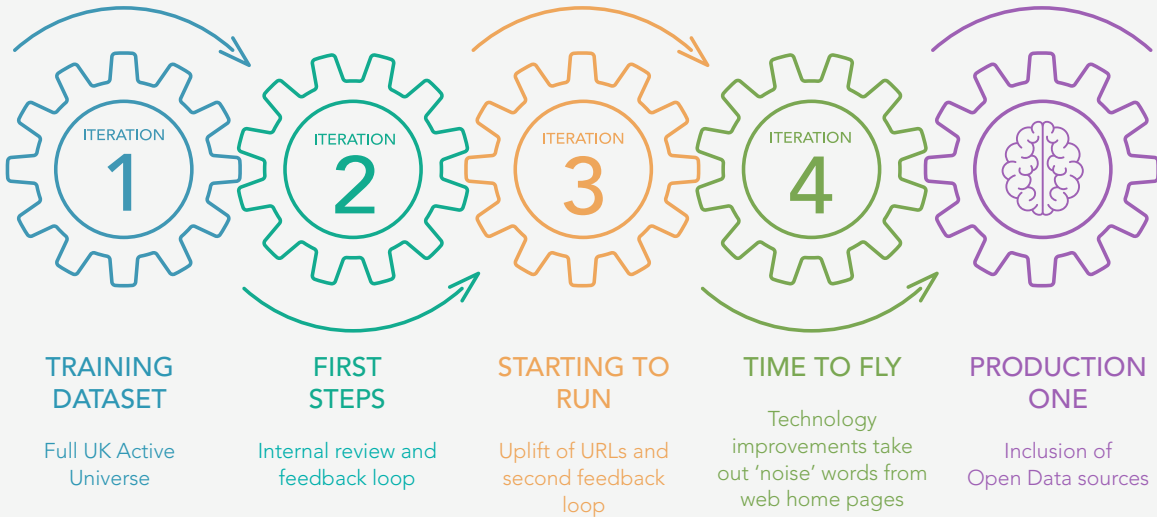
After all of the evidence has been weighed, the system produces a confidence score which reflects the confidence of each decision. Critically, the system 'knows when it doesn't know.' This is very important, since the human operators can now direct their energy to the most challenging cases, better suited to the human ability to understand complexity and nuance.

INPUT DATA FOR THE PROOF OF CONCEPT

The input data sources were:

– 3.5 million records from the United Kingdom's Companies House.

– 2 billion+ websites (in index form)

AN OVERVIEW OF THE PROCESS CAN BE SEEN IN THE DIAGRAM BELOW:

ITERATION 1

TRAINING DATASET

Full UK Active Universe

ITERATION 2

FIRST STEPS

Internal review and feedback loop

ITERATION 3

STARTING TO RUN

Uplift of URLs and second feedback loop

ITERATION 4

TIME TO FLY

Technology improvements take out 'noise' words from web home pages

PRODUCTION ONE

Inclusion of Open Data sources

The system, returned a total of 28,142,665 classifications (at least 5 for each organisation), together with a taxonomy score. The higher the taxonomy score, the more likely the SIC is correct.

A process of in-depth quality assurance followed each iteration and provided feedback to fine-tune the model and get it ready for production.

There were four iterations on average five weeks apart, increasing in reliability and accuracy.

PROJECT SCALE

A web page contains 440 words on average, and humans beings read the text at the rate of circa 200 words per minute (4). It would, therefore, take a single person 4,560 working days to read one million web pages, or twenty years and four working months. The AI systems, however, read one million websites in around 8 minutes; the whole website, not just the homepage.

In data terms, over 350TB of data was modelled for over 5.6 million UK businesses. Dun & Bradstreet began with foundational data and 'explicit' datasets, such as UK 2007 SIC codes already applied to the data, along with the broad line of business description, as per the examples on the right.
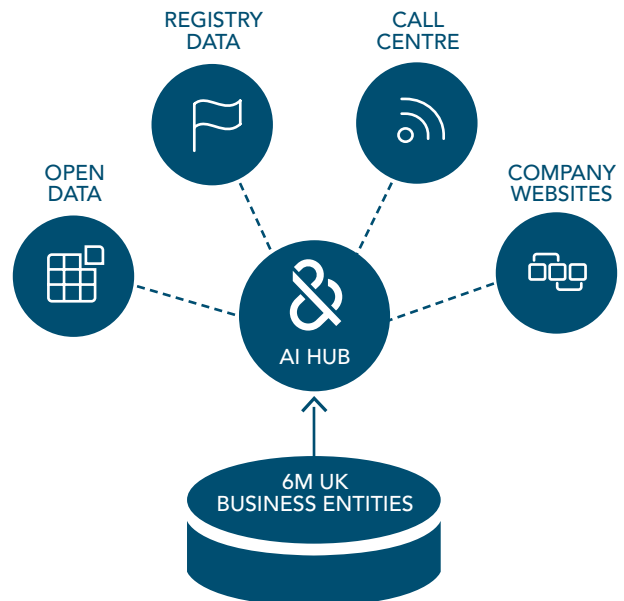
REGISTRY DATA

CALL CENTRE

OPEN DATA

COMPANY WEBSITES

AI HUB

6M UK BUSINESS ENTITIES

THE HUMAN IN THE LOOP

Aligned with this is a continuous improvement programme and a quality assurance tool (QA) so that the process is always learning and the inputs help 'teach' the machine with 'human in the loop' processes.

So, for each record, the SIC codes provided can be reviewed and validated using the 'about us' section of the company website or other sources of content utilised for assignment.

For each record the QA tool can:

– Validate the SIC, if it is correct
– Remove the SIC, if it is incorrect
– Add additional SICs
– Indicate whether a URL is incorrect or inactive
– Add in the correct URL where applicable

So far, the process has delivered 6.7 million UK SICs, either as newly assigned or verified as below, and the programme is being extended to the US database.

When compared against the objectives, it looks like this:

– Improved coverage = 4.3 million additional SIC codes
– Improved accuracy = 1.2 million verifications
– Improved depth = 1.5 m with advancement in precision

This is clearly an improvement. Consider looking at the results through the lens the business development scenario mentioned previously. Here is the data for the business park in detail after the application of the AI process:

– 40% of primary SICs were changed, providing a much clearer classification
– 6% were validated, providing a much clearer view in terms of those complementary businesses

WHAT'S NEXT

As the model evolves the programme has included more 'implicit' datasets, which using machine learning has enabled identification of industrial activity through implied insights such as relationships. Examples of these types of 'implicit' datasets are below.

TRANSACTIONAL DATA

Data sources to be used in future production runs include:

– transactional data such as payment history (e.g. understanding approximately 4 million payment interactions on over 1 million businesses)
– what other companies a company does business with and can provide insights into what the company in question does.

IMPORT & EXPORT

Understanding if a business imports or exports, along with what a company is importing or exporting through commodity codes, can also help to identify what the company in question does.
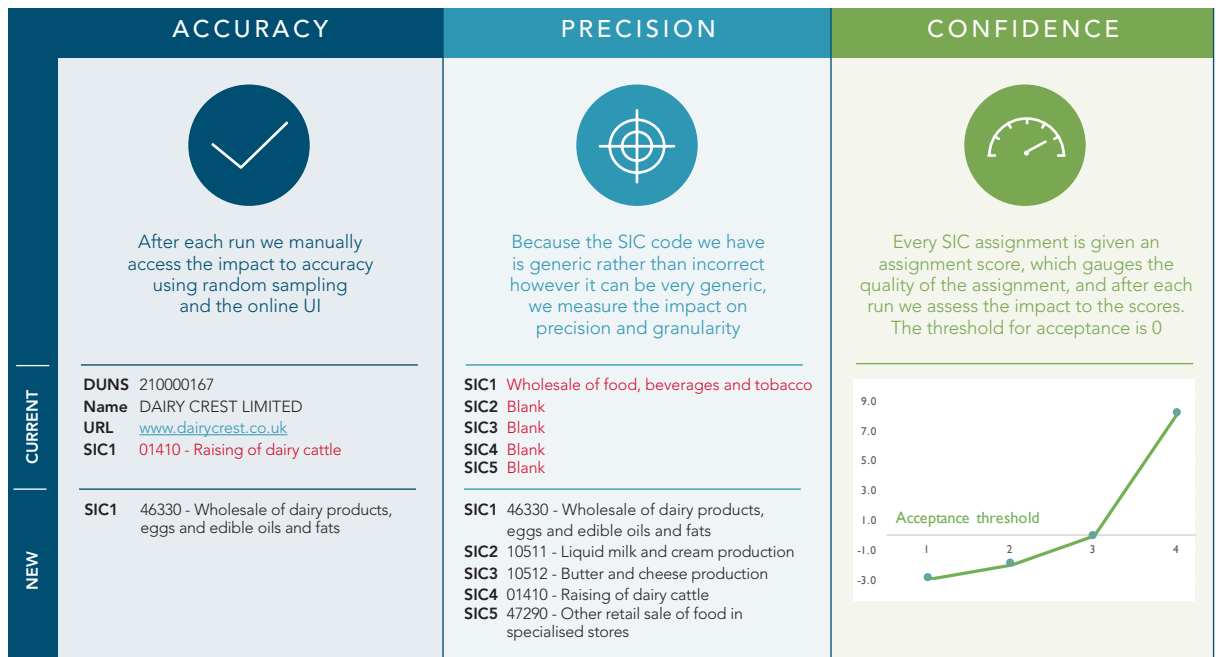
SOCIAL MEDIA

Data from digital social sources that maps business foot traffic across particular categories can provide a rich source of text for AI 'learning.'

OPEN DATA

Open data provides many opportunities to identify business activity using a growing number of resources.

Adding data sets has improved accuracy and precisions as can be seen using the examples below:

ACCURACY	PRECISION	CONFIDENCE
After each run we manually access the impact to accuracy using random sampling and the online UI	Because the SIC code we have is generic rather than incorrect however it can be very generic, we measure the impact on precision and granularity	Every SIC assignment is given an assignment score, which gauges the quality of the assignment, and after each run we assess the impact to the scores. The threshold for acceptance is 0

CURRENT

DUNS	210000167
Name	DAIRY CREST LIMITED
URL	www.dairycrest.co.uk
SIC1	01410 - Raising of dairy cattle

SIC1 Wholesale of food, beverages and tobacco
SIC2 Blank
SIC3 Blank
SIC4 Blank
SIC5 Blank

NEW

SIC1 46330 - Wholesale of dairy products, eggs and edible oils and fats

SIC1 46330 - Wholesale of dairy products, eggs and edible oils and fats
SIC2 10511 - Liquid milk and cream production
SIC3 10512 - Butter and cheese production
SIC4 01410 - Raising of dairy cattle
SIC5 47290 - Other retail sale of food in specialised stores

Acceptance threshold

IN SUMMARY

Data science has shown that it can help improve SICs and ultimately the data Dun & Bradstreet provides to customers. This is just a start though, as big data processing and more open data become accessible, further improvements will be seen. The process will ultimately expand the view of what a business does for a far more targeted sales and marketing approach and portfolio risk segmentation.

The use of AI will only increase, as the technology improves, and analysts struggle to make sense of the multitude of data sources that AI can process easily to make a well-rounded decision.

ABOUT EVOLUTION AI

Evolution AI develops AI products that read and understand human language (known as 'Natural Language Processing'). Their systems can learn to perform useful 'knowledge-work' tasks without any explicit instructions. They are built to analyse information from large volumes of text documents and integrate it into existing business processes and IT systems. They exploit recent advances in deep learning technology, running on powerful graphics processing units (GPU).

RESOURCES

1. "Standard Industrial Classification," Wikipedia. [https://en.wikipedia.org/wiki/Standard_Industrial_Classification]
2. "NACE REV 2," eurostat. [http://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07- 015- EN.PDF]
3. Ziefle, M. "Effects of Display Resolution on Visual Performance," ResearchGate. [https://www.researchgate.net/publication/13315672_Effects_of_Display_Resolution_on_Visual_Performance]

dnb.co.uk