# SME-Oriented Collaborative Customer Insight Platform for User-driven RdM (SOC-UDRdM)

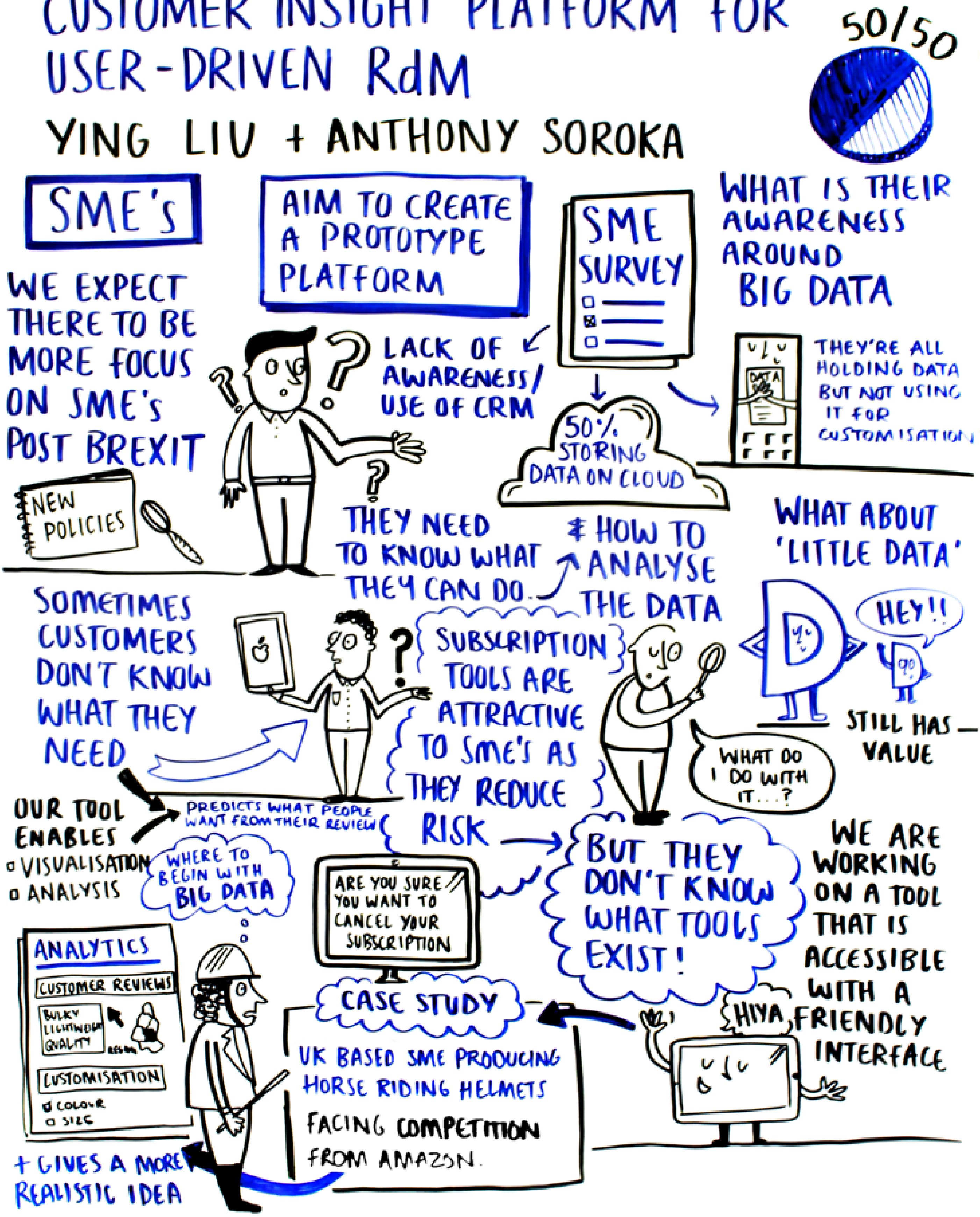
## RECODE Network

Cranfield University

RECODE-NETWORK

# SME-ORIENTATED COLLABORATIVE CUSTOMER INSIGHT PLATFORM FOR USER-DRIVEN RdM

YING LIU + ANTHONY SOROKA



**SME-Oriented Collaborative Customer Insight Platform for User-driven RdM (SOC-UDRdM)**

**Ying Liu**
Cardiff University

**Liangxiu Han**
Manchester Metropolitan University

## About Us

# RECODE Network

The EPSRC-ESRC funded Network in Consumer Goods, Big Data and Re-distributed Manufacturing (RECODE) has been created to develop an active and engaged community to identify, test and evaluate a multi-disciplinary vision and research agenda associated with the application of big data in the transition towards a Re-distributed Manufacturing model for consumer goods.

The exponential growth of available and potentially valuable data, often referred to as big data, is already facilitating transformational change across sectors and holds enormous potential to address many of the key challenges being faced by the manufacturing industry including increased scarcity of resources, diverse global markets and a trend towards mass customisation. The consumer goods industry, has remained largely unchanged and is characterised by mass manufacture through multi-national corporations and globally dispersed supply chains. The role of Re-distributed Manufacturing in this sector is often overlooked, yet there is great potential, when combined with timely advances in big data, to re-define the consumer goods industry by changing the economics and organisation of manufacturing, particularly with regard to location and scale.

The RECODE Network conducted five feasibility studies led by the academic core partners, steering group partners, and new partners who joined through the RECODE Sandpit on 02-03 March 2016. A multidisciplinary team comprised of internationally renowned experts from Cardiff University and Manchester Metropolitan University and practicing industry leaders in the fields of big data analytics, consumer insights and manufacture were involved in the delivery of this feasibility study.

RECODE has developed novel methods and undertaken innovative events to engage communities of academics, international experts, user groups, government and industrial organisations to define and scope a shared multi-disciplinary vision and research agenda. To find out more, visit our website: **http://www.recode-network.com**

# Table of Contents

## Introduction

The term Re-distributed Manufacturing (RdM), within the context of the UK Engineering and Physical Science Research Council (EPSRC) RdM Networks programme, has been defined as technology, systems and strategies that change the economics and organization of manufacturing, particularly in relation to location and scale [1]. Smaller-scale manufacturing has the potential, if applied appropriately with suitable levels of localization, to drastically reduce supply chain costs, improve sustainability and tailor products to the needs of users and consumers.

Such smaller-scale manufacturing has the potential to help tailor products to satisfy the specific needs of consumers differing in terms of geographical location, cultural roots, improve sustainability as well as drive the society towards circular economy [2].

While RdM has the great potential to improve sustainability, currently, very little has been understood on how RdM could help SMEs for gaining economic benefits due to the constraint of their business model, lack of understanding on customers, limited resource commitment on R&D, marketing and sales, supply chain integration, etc.

Micro enterprises and small and medium-sized enterprises (that for the purposes of this exploratory study will be referred to as SMEs) have an important role to play in the continuing success and growth of national economies. Studies have suggested that the contribution of SMEs in realizing the demands and improving the profitability of their supply chain partners (including large organizations) should not be understated, as such play a critical role in modern economies [3, 4].

A study conducted for the British government's department of business innovation and skills reported that in 2015 99.9% of the total number of enterprises in the UK could be classified as being SMEs. These companies contributed more than 2/3rds of private sector workforce as well as 47% of the annual turnover in the UK.

According to an extensive study of companies conducted by The Data Warehousing Institute (TDWI) in 2009, 38% of organizations surveyed reported that they utilize advanced analytics, whereas 85% said they would be practicing it within three years [5]. According to the study the respondents were spread evenly across a wide spectrum of company sizes. However, only 23% of respondents were from companies whose revenue was less than $100million [5], within the EU this is above the turnover threshold for a medium enterprise. Additionally only 4% of respondents to this study were from non-computer manufacturing companies. As such this and other industry studies may not accurately reflect the opinions of manufacturing SMEs.

This feasibility study specifically focuses upon SMEs in the UK and their potential motivations for using and knowledge of big data based customer analytics. Whilst there is much research regarding big data and collaborative design there is comparatively little that examines these from the perspective of SMEs. Two research questions are tackled:

1. What are companies doing with regards to their product and customer data?

2. What position are they in to make use of big data analytics and participate in RdM?

## SOC-UDRdM industrial survey results

### Perception of Re-distributed Manufacturing

Because RdM has been defined around smaller scale manufacturing SMEs stand to potentially gain the most from the proliferation of RdM. In order to gauge the perception of RdM amongst SMEs they were asked three questions.

1) Have they heard of RdM (awareness);

2) Would RdM be helpful in their company (usefulness);

3) Do they have any plans to get involved in RdM (involvement).

As Table 1 shows the majority of respondents had not previously heard of RdM and only a minority think it could be important for them or have any plans to become involved.

The fact that 67% of respondents were unsure how useful RdM could be suggests that the RdM research community and policymakers still have a lot of work to do to raise awareness amongst SMEs. This is concerning as theoretically it is SMEs who could stand to benefit from RdM.

The alternative to this could be large manufacturers establishing smaller, distributed manufacturing, assembly or finishing centers. The automotive sector provides an example of this, although not common in Europe it is normal for dealerships to fit optional extras in cars in emerging economies. This lowers the product variety levels offered by the car plant which in turn keeps production costs low [6].

### Availability of customer data

A company, depending on the products it makes and the IT systems it uses can carry varying types of information. The bare minimum most companies will have is sales and order information, some will carry product specification data especially when bespoke products are frequently manufactured. Other forms of data that companies may hold include CAD data from customers, returns data and potentially feedback as well.

For the purposes of the survey the different types of data were grouped into several broad categories: Supplier Data; Manufacturing Data; Product Data; Sales Data; Customer Data; Financial Data. The respondents were then asked to indicate what type(s) of data they had (they were also given a free text field to indicate any other types of data they have).

In the results shown in Table 2, it can be seen that nearly all of the respondents in the survey maintain product data and that the majority keep sales data. With hindsight the near universal maintenance of product data could be considered to be unsurprising as it is sufficiently broad that this would incorporate CAD data, product specification, product datasheets etc. Such data could for example include what is necessary for a "CE" mark within the EU that shows that the manufacturer has checked that these products meet EU safety, health or environmental requirements. Additionally one respondent indicated that they also kept research / IP related data and another kept regulatory compliance data.

| | Unsure | Yes | No |
|---|---|---|---|
| Aware | - | 40% | 60% |
| Useful | 66.7% | 6.7% | 26.7% |
| Involvement | - | 6.7% | 93.3% |

**Table 1** Perception of Re-distributed Manufacturing

| Type of data held | Percentage of respondents |
|---|---|
| Supplier data | 77.3% |
| Manufacturing data | 77.3% |
| Product data | 90.9% |
| Sales data | 81.8% |
| Customer data | 68.2% |
| Financial data | 77.3% |
| Other | 13.6% |

**Table 2** Types of data held by companies

Based on the results from this preliminary question, there is an indication that the majority of companies could be in some sort of position to apply some form of big data analysis techniques. However, this would of course be contingent on the type of data being stored and its format, for example a sales spreadsheet would be considerably easier to analyse than a scanned document.

Despite the fact that companies may maintain a lot of data those figures would not indicate how much customer specific data is kept. The companies were asked if they maintain data about their customers, from Table 3 it can be seen that 80% said that they do and 20% responded that they do not.

Based on this the respondents answering 'Yes' were asked to provide information regarding how they currently use data (in whatever way it is stored) as shown in Figure 1. Of those that responded the majority indicated that they use it to improve their products, market and customer service. This raises an interesting question, that would perhaps be worthy of future study, regarding how many companies use or keep records of their product reviews. Specifically those from websites such as Amazon where customers leave feedback and Which?, an organization that conducts formal product tests. Such websites, as mentioned earlier, can contain a great deal of information that might be of use to product manufacturers.

| | Yes | No |
|---|---|---|
| Maintain customer data | 80% | 20% |

Table 3 Is data regarding customers stored?


Figure 1 Current uses of customer data

As some of the analyses mentioned in Table 4 are provided within Customer Relationship Management (CRM) systems the respondents were asked to confirm whether the used CRM systems of not. The results if Table 5 show that the overall response was that 67% of the respondents did not use CRM and only 33% did use CRM systems. This could be a matter of concern, if a company isn't in the position to use CRM it raises a question mark regarding how able they would be to make full use of big data analytics.

## Data storage

Studies have suggested that cloud computing could be an enabler for SMEs wishing to use big data analysis [7]. However, in order to make this achievable there is some is necessity for the use of cloud storage. Therefore the respondents were asked what storage methods they used. As can be seen from Figure 2, 50% of respondents use cloud storage for some if not all of their data. This is broadly positive as it implies there is some familiarity with cloud computing at some level. The majority of respondents use Network Attached Storage which suggests a degree of familiarity with the idea of not just storing data on the hard drive of the PC they use.

| | Yes | No |
|---|---|---|
| Use customer relationship management | 33% | 67% |

Table 4 Use of Customer Relationship Management systems

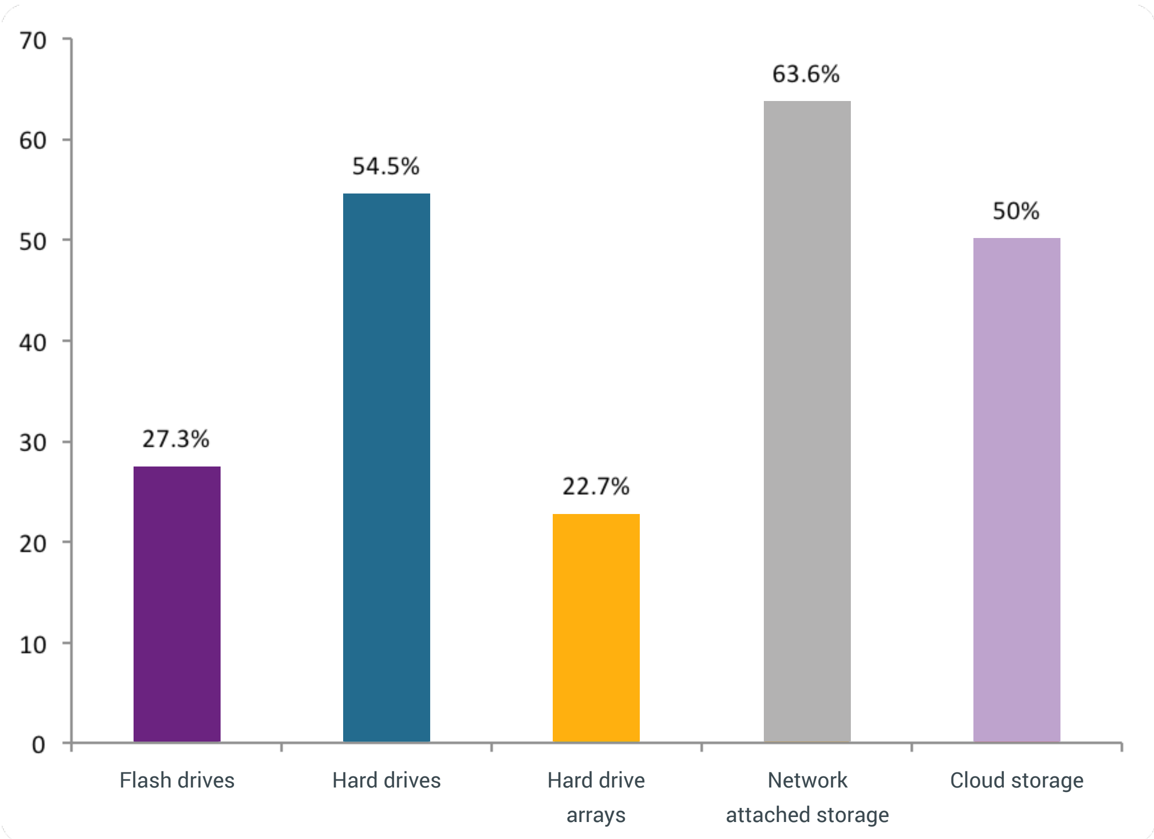



Figure 2 Storage methods used by SMEs

## Data volume

In order to gauge how much data companies actually store they were asked to estimate how much data their company actually has. The results of this are illustrated in Figure 3, as can be seen the majority of respondents have less than 2TB of data. Though this amount may seem comparatively small (within the context of big data) but taking a historical perspective 500GB of data is over 100 times the capacity of a typical hard drive used in a fileserver's RAID array in 1996 (c. 4.3GB) and would have been unimaginable for SMEs.

The respondents were asked whether they had any difficulties managing the volumes of data they have and 77% responded that they do not have a problem. However, 23% have indicated that they do have a problem and a cross-tabulation between this and the results in Figure 3 shows that it is the respondents with >2TB of data that experience problems. Problems noted by respondents included "too much [data] and not enough time to analyse" and "lack of experience".

This is broadly positive as it shows that at the very least most are able to manage the data that they already have, but it does seem to suggest that as volumes increase companies begin to encounter problems.

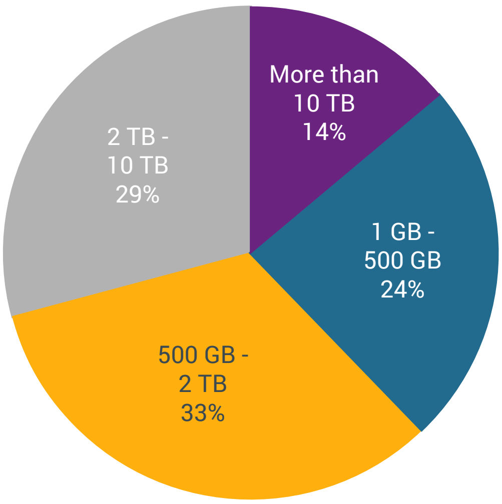



Figure 3 Amount of data stored by SMEs

## Awareness of big data

The survey respondents were asked if they were aware of big data analytics as a yes / no question (Table 5). As the table below shows less than half of them were aware of big data analytics.  When compared to the 85% figure for companies that would be using it within three year from the TDWI [5] this suggests that there is a possible discrepancy between what businesses as a whole and what manufacturing SMEs think about big data and potentially their ability to gain meaningful results.

Also of note is that 75% of companies surveyed by TDWI were non SME and only 4% were non-IT manufacturers, this also suggests that larger enterprises are more aware of big data than SMEs (especially manufacturing companies outside of computing / IT sectors).

Further analysis of the survey data showed that the respondents from the previous section who had >2TB data and were experiencing problems indicated an awareness of big data analytics.

| | Yes | No |
|---|---|---|
| Aware of big data | 50% | 50% |

Table 5 Awareness of big data analytics.

Figure 4 shows what companies would hope to garner from big data analysis, the responses when comparing to the lists provided in [8] it shows that the overall awareness of all that could be achieved through the application of big data analytics is perhaps not as good as it could. Additionally this suggests that perhaps there is a possibility that companies have heard of big data analysis but are in reality unaware of what it can offer, this would have to be investigated through further studies. In total of the companies who had heard of big data analytics three quarters of them actually had a vision of what they'd hope to achieve through big data analytics.

Based on this initial data it could be concluded that there is a need to further inform SMEs about big data analytics, what it can do and how to overcome the real or imaginary barriers to entry that SMEs might face.

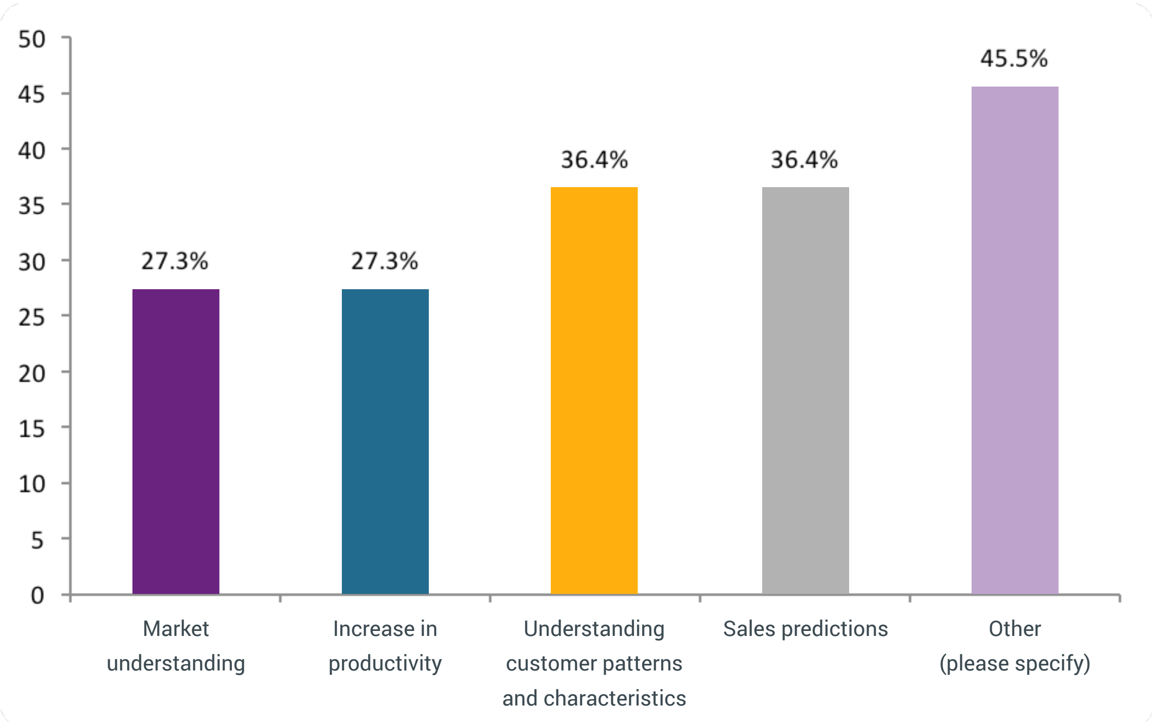



**Figure 4** Insights that companies wish to gain through analytics

## SME case studies

The case studies examined how several manufacturing SMEs dealt with their data, product design, customer engagement and what they felt big data could offer them.  In order to create a picture of the challenges SMEs face these results have been synthesised. Illustrating what small and medium companies are doing, wish to do and what challenges they face for each of the above themes.

### Customer engagement and collaboration

Amongst the SMEs interviewed it was discovered that there were essentially three models of engagement used. Direct sales from the company location, sales via company retail point and resellers (Figure 5).

The companies whose sales are direct report that it enables the companies to directly engage with the customer to obtain a good understanding of the customer needs. Even those who utilise an in-company retail point at a different geographic location (often with salespeople involved) are already reporting that they have challenges finding out what a customer desires from their product. This can include salespeople filtering out “bad news” that they think the manufacturer doesn’t want to hear – even though they do want this information in order to improve product quality. Those who have no direct or indirect contact with their customers (dealing solely with resellers) report that they find it even more difficult to obtain information, such companies acknowledge that some form of customer sentiment analysis tool would be highly beneficial.

As well as the challenges associated with obtaining feedback another challenge some companies face is that of regulations. At one level, it can be CE marking which indicates that the manufacturer claims compliance with the relevant EU legislation. A far bigger challenge is faced by those companies who face tougher regulation because of the sector they operate in and/or the products they manufacture. As an example, some safety products require type-approval, as such they are produced to a standard with little room for manoeuvre before re-approval is required.  So, colour can be changed but if the geometry is changed then new type-approval is required. This can tie the hands of the manufacturer and make customisation prohibitively expensive.

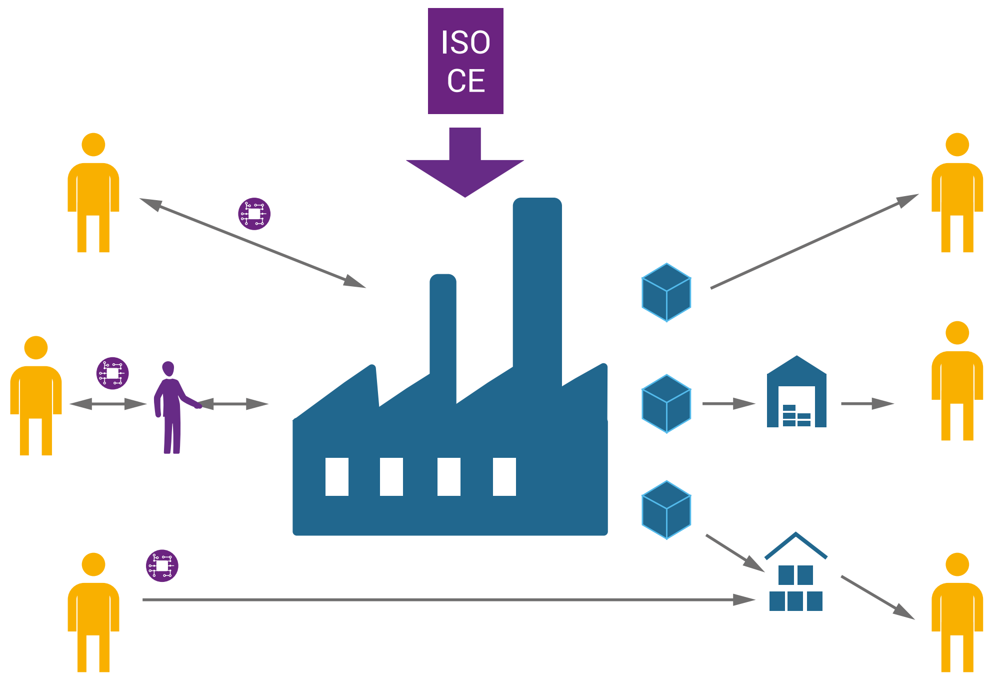



Figure 5 Captured SME-customer interactions in Re-distributed Manufacturing

### Customer and product data

The companies interviewed maintain a variety of data from and regarding their customers. All of them maintain sales data whether it's in an Excel based system or a dedicated package such as Sage 50. As mentioned in the previous section there are several different sales channels that the companies use, therefore this data could come from direct sales or via outlets that they serve (Figure 6).

The specific product data can range from the basic data relating to the design of the product, to highly detailed and specific CAD data. One company has a web-based system for its sales team that allows certain product parameters to be adjusted with costings and production data automatically produced. The availability of product feedback data from customers is also highly variable with some companies and products having very little feedback (other than perhaps data arising from product returns). Additionally, sales teams sometime filter information such that product designers do not get a full picture of what is going on. This makes the availability of some form of analysis tool where customer options and preferences could be gathered and analysed a highly attractive proposition.

The perception regarding data handling challenges by the companies is that any problems are comparatively minor, showing that the companies are comfortable handling the amount of data that they do. The challenges they have identified include not having as much feedback data as they'd like and data and IP protection (specifically stopping competitors copying products).

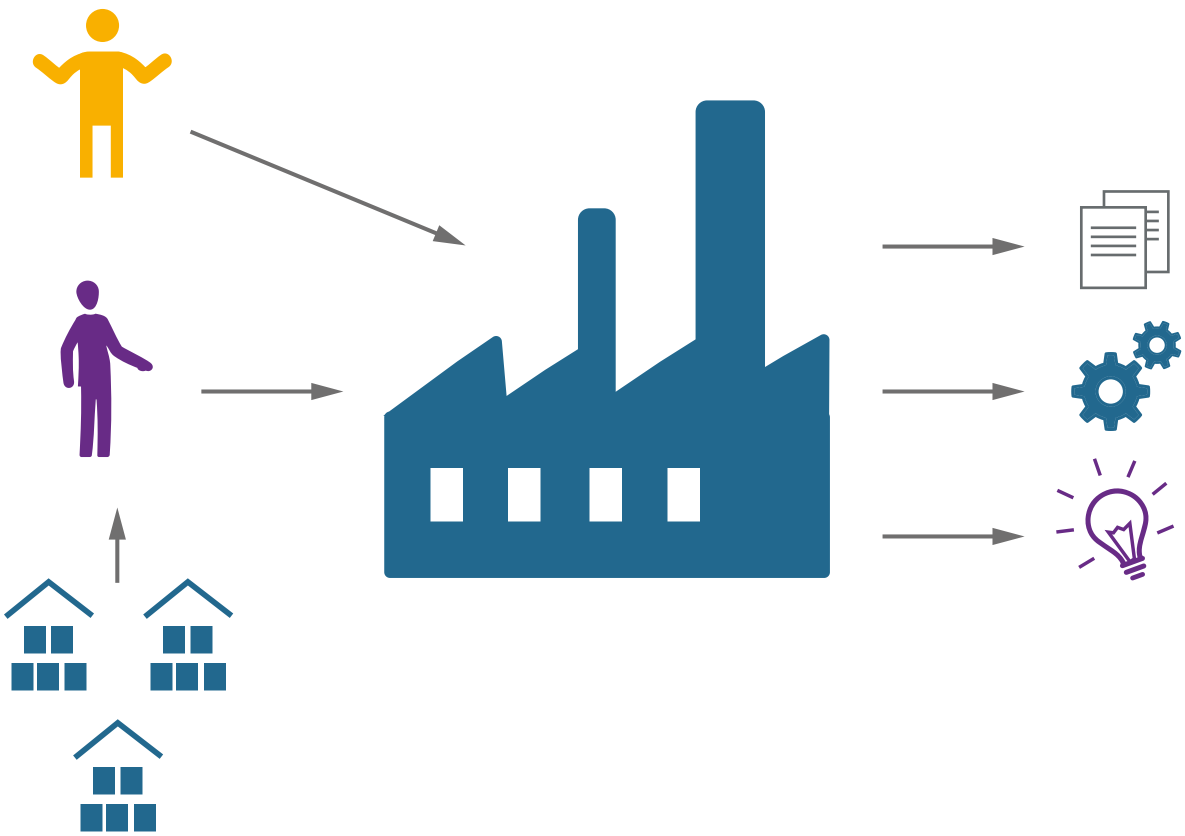


Figure 6 Product and customer data leads to new value creation and business models

## Big data enabled product and service design

Although the companies had some awareness of the idea of big data there was no real knowledge of where to start when it came to implementing it within their own companies (Figure 7).

Amongst the things companies are interested in finding out via big data are developing trends within the markets that they serve, feedback on the products they manufacture and what gaps in the market could be based. Based on this data their products could be improved and new ones potentially developed. However, some companies feel that in the short-term professional CRM systems may be of more benefit than a financial investment in big data analytics.

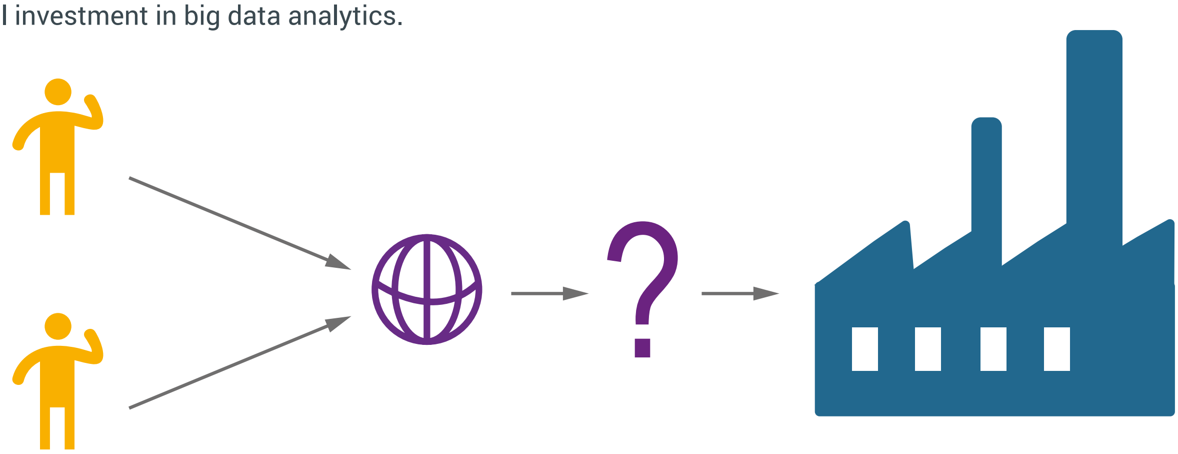


Figure 7 Big data enabled product and service design

## Summary

With regards to the two research questions posed the following conclusions can be drawn:

RQ1 - What are companies doing with regards to their product and customer data? Although companies are storing data, not all companies are actively using customer data. The survey also showed that only a few companies are actively using CRM systems, which could provide answers to some of the queries they have.

RQ2 - What position are they in to potentially make use of big data analytics and participate in RdM? Half the respondents had not heard of big data analytics, of those who did there were some ideas regarding what they hope to achieve. However, as mentioned above a small proportion of companies actively used any form CRM which could perhaps provide some of the insights they wish to obtain. The small number of respondents who had any plans to get involved in RdM is a concern as it suggests that there is a lack of awareness of what it is and how SMEs stand to benefit. The potential consequence of this is that individual SMEs may miss-out or that large manufacturers take it upon themselves to redistribute their manufacturing hence ignoring SMEs. This is something that may require actions from academia, policymakers as well as organizations involved in supporting and nurturing manufacturing.

The results of this exploratory study suggest that although there may be some demand for big data analysis it is possible that the current solutions may not be viable for SMEs and that SMEs are seem to be ill-prepared and ill-equipped to make the most of what big data analytics can offer them.

This initial exploratory survey also shows that there is a necessity to study the needs of manufacturing SMEs and big data analytics and RdM in more detail.

## SOC-UDRdM customer insight system

As the arrival of the fourth industrial revolution, data-driven digital services and products are playing a central role to drive breakthrough revenue growth, optimise customer interaction and access, and offer strategic decisions for both companies and customers on what to produce or buy [9][10].

One area of interest is in the extraction and quantification of customer's affective states and other subjective information from online reviews appearing on social media, as well as merchant and review websites, for various products and services. A product review, written by a user, refers to a customer's opinion about a product, which describes negative, positive or neutral parts of a product. Product reviews have been widely recognised to have a significant influence on customers' shopping decisions and products and services and also business strategies and manufacturing decisions [9][11][12][13]. For instance, [12] presented a comprehensive analysis on the impact on economic outcomes such as product sales and its relation to product reviews by exploring multiple aspects of review text. In [13], it described how customers read reviews to find unique information about products and reduce the risk of their buying decision. The work [10] developed an approach by mining consumer reviews to better understand what are the customer preferences and actions and how reviews impact price power of product and how to improve product sales predictions.

However, with the massive information available, manual analysis of the data is costly and time consuming as well as being subjective, prone to error. For example, if a customer wants to buy product, he or she might only be able to read a small number of reviews, which may lead to a biased purchase decision. Similarly, a business may not be able to keep track their products and understand market needs and make timely decisions on what to produce.  Despite encouraging existing research efforts on sentiment analysis in the domain of product reviews [9][14][15], most existing systems are partially automated with the main focus on sentiment analysis part, which may not be able to handle big data. In addition, these existing systems mainly provide text review summaries but not graphical summaries.  The challenges remain on how to collect data from websites, analyse them and present the results to users in a user-friendly efficient manner.

To address the challenges above, different from all existing systems, we have proposed a cloud-based data analytics platform, which can automatically scrape data form online websites, perform sentiment analysis of a given product and visualize summaries of reviews based on product feature aspects and aggregated features for customer insights in real time.

## System architecture

Based on the objectives above, the system should be able to provide the ability for a user to search for products, select them and view a number of charts that highlight the customers' affections to various features of the product. Useful information that can be gleaned includes, but is not limited to, overall satisfaction with the product, a list of features specifically chosen for comment by customers and their related sentiment, the ability to compare categories for a given polarity to see the distribution of sentiment, the distribution of sentiment for any given category and the comparison of these charts for different products. The system should also be able to show the potential for geographical maps if the data is available highlighting the countries with the most liked or disliked products or areas with the largest number or reviewers or particular sentiments. Finally, all of these can be linked directly to the text of the review allowing the user to select any section of the charts and read the filtered reviews that specifically relate. This allows the user to read why there is a negative sentiment regarding price or what exactly people like about the way it fits. All this information can be used to better improve products or services by detecting trouble areas or showing what works well.

The overview of the system is shown in Figure 8. Due to both research and commercial interests, we have chosen Amazon website as an exemplar case study. However, our system can be adapted to other websites for sentiment analysis of text data. There are several components of the system including 1) data acquisition; 2) sentiment analysis; 3) data visualization; 4) graphical user interface.

## Data acquisition

To analyse the reviews, it is important to first accurately and collect data from websites in a timely manner. However, due to high volumes of data, manual collection of web data increases the cost of labour significantly. Additionally it is known to be error prone. There is a need to develop an automated approach to scrap review data from website. In this work, we have developed web scraping service to collect review data from Amazon website, which can be effortlessly consumed by analysis components of the system. Due to restricted access to the review information in its raw form, we have combined Amazon product API with our web scraping service to search Amazon database and return the search results, which then prompts the retrieval of the review data and automatically load the product review pages from the website and copy out the text of reviews for the subsequent stage of analysis.

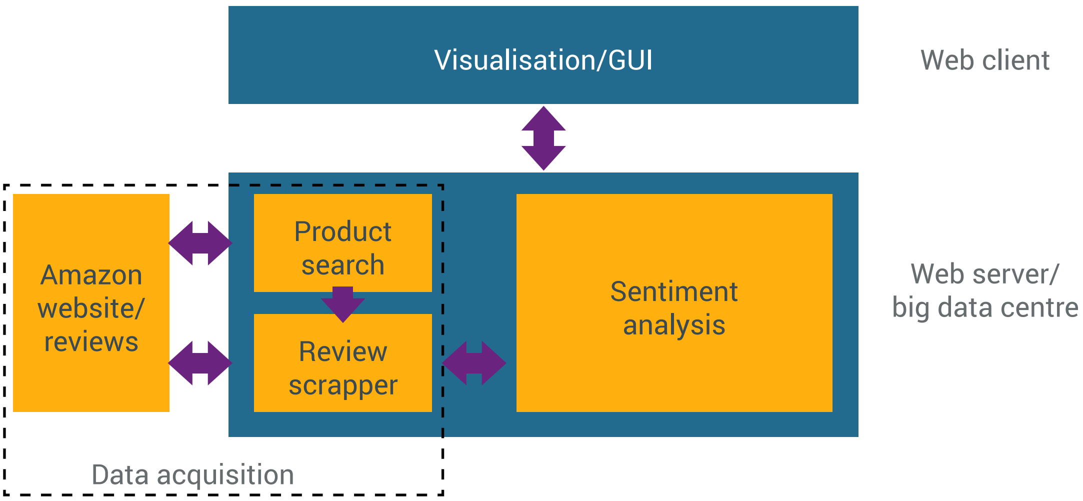



Figure 8 System overview

## Sentiment analysis

Sentiment analysis refers to detect whether a given text represents a positive or negative or neutral opinion (sentiment polarity). There are three levels of sentiment polarity classification including document, sentence and aspect levels [16]. The document level mainly focuses on classification of a whole document as expressing a positive or negative opinion. The sentence level analyses each sentence' sentiment polarity. The aspect-level sentiment analysis mainly deals with specific aspects or features of a document. In our case, we mainly target aspect-level sentiment analysis of product features. For example, a review about a mobile phone may contain the following:

"The screen was great but the speakers were terrible". An ideal sentiment analysis system should be able to identify both "screen" and "speakers" as product features with positive and negative sentiments respectively.

Aspect-level sentiment analysis is particularly useful to compare two or more products based on their feature aspects. However, most of current online websites only provide an overall score for a particular product. Consumers have to read all reviews and conduct manual analysis to make decisions.

In our work, we have applied both document-level analysis and aspect-level analysis to product reviews. The workflow of sentiment analysis module is shown in Figure 9.

As shown in Figure 3, the Natural Language Processing module (NLP) takes a review text as input and pre-processes the data by performing tokenisation on the texts for further aspect-level and document level analyses. The aspect analysis module takes the output from the NLP module and identifies features and sentiments as well as performing aspect feature categorisation, which can then be visualised through the graphical user interface. Similarly, the document level analysis performs sentiment analysis of the whole review text and the results will be displayed to the user.

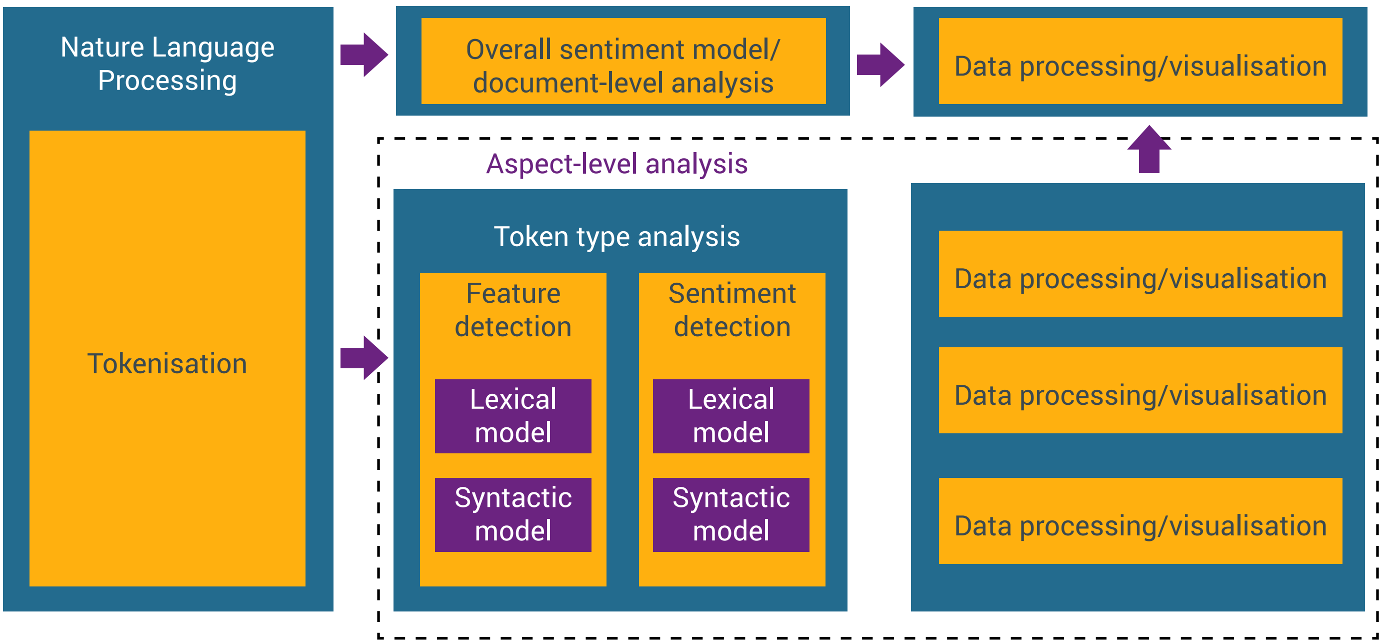



Figure 9 Sentiment analysis workflow

## Aspect-level sentiment analysis

For the aspect-level sentiment analysis of this work, we have built upon [14] and extended his work by introducing new feature categorisation and feature matching algorithm. The basic technologies in use are the Stanford Natural Language Processing library (coreNLP) for tokenisation [17][18] and Support Vector Machine (SVM) using the libSVM library implementation for feature detection and sentiment detection [19]. Essentially, it attempts to determine if any given word in the review is a 'feature' or a sentiment-bearing word by first processing the review text using coreNLP which tokenises the text and assigns features to each token based on natural language principles, these are then passed into four SVM models (both lexical and syntactic models for 'feature' and 'sentiment' detection separately). These models provide a score based on how much each token fits each role and if a token has scores above zero then the model that provided the highest score determines its type. Once the features and sentiments have been extracted, it determines which sentiment applies to which feature through the use of the matching algorithm developed for this project. This data can then be collected such that all the raw review texts are packaged up with their specific extracted features and sentiments and sent to the web application.

During this process, an important part of the analysis concerns the matching of each feature with the appropriate sentiment bearing word so that the correct sentiment value can be recorded. We have developed a matching algorithm where it follows a rule-based approach to connecting a feature with its appropriate sentiment working first on a proximity basis and then to any connected sentiment found through the natural language processing. This method consists of three steps:

- The first step checks whether the feature itself is sentiment bearing and if so matches the token with itself as the sentiment expressed by the token is certainly related to the feature, i.e. "comfortable".

- The second step iterates through the surrounding 6 tokens, 3 on each side of the feature looking for a sentiment. In this case if a sentiment is found it is matched because the closer to the feature the sentiment is the more likely it is to be related. For example many of the reviews had simple terms such as, 'good looking' or 'great price' as well as 'the colour was good' or 'it fits very well' and as such by checking the 6 tokens alternately starting with the preceding one and then the next one and continuing outwards any detected sentiment had a high probability of being connected. However a number of issues can arise from this particularly lists and comments including 'but'.

- The final step checks the semantic incoming edge of the feature token (which is set as part of the natural language processing module) to determine if it is a sentiment or a feature. If it is a sentiment then as with the previous reasoning it is likely to be connected to the feature as it is in semantic proximity. If it is a feature then the thought is that given the feature is semantically close and we have not discovered a sentiment within a 3 word radius, then we can assign the sentiment of the new feature to the current one as it is likely with a list or comment discussing the same sentiment about multiple features at the same time. To do this the matching algorithm is recursively called on the new feature and the result returned. Finally the other semantically tagged tokens are inspected.

## Document-level sentiment analysis

For the document-level analysis, we have analysed the sentiment polarity of a whole review text based on the Stanford Natural Language Processing library (coreNLP) [17][18], which takes the input of a review text and returns the polarity of each sentence and then aggregates polarity of sentences.

## User interface

One of the objectives of this work is to design a friendly easy to use web-based interface to control the system. With this in mind, we have used the different cards to delineate a number of functional areas, which contains a different aspect of the interface. We have also used a small palette of contrasting colours used to clearly emphasise the function of items such as buttons and dropdown menus. Shadows are also used to simulate depth on interactive objects. The changes of colour and shade are used to denote the state of user controls coupled with changes to the pointer standardised across browsers. The chart card is used to display the visualised data of a selected item which can be controlled using the contextual controls in the card above. This card contains a chart selection dropdown to change between different charts and a range of select inputs and buttons that will appear based on the chart chosen which can be used to specify the data displayed.

## The implementation

To enable the system to be run on the cloud computing platform and ensure scalability and efficiency, we have chosen a REST design pattern, all the system components have been implemented as web services based on Dropwizard [20], a java-based framework/library to facilitate development of high performance, restful web services. The system combines Jetty as an embedded web server, with Jersey to map Java objects directly to HTTP requests using Jackson for the JSON conversion and a variety of often used libraries in Java development. Essentially it allows us to create standard Java classes and use annotations to map the various methods and object to the HTTP request URLs without having to manually create handlers or servlets, etc. It also generates a single large jar file which combined with a configuration file is all that is needed to run the service, making distribution and deployment a simple matter.

## Data sources

The data was scraped from Amazon website in real time. Since we have used supervised machine learning approach, for the training purpose, we have first collected a total of 704 reviews containing 1073 product aspect features for the model construction. Each feature was tagged with the exact text from the review that represented. N-fold cross validation was used for the validation of the model.

## Data visualisation and system use

To provide a clearly condensed representation of the reviews about a given product that a user wants to get, we have developed visualization techniques in various of charts to provide product review summaries, which are based on features and or aggregated features of a given product. The charts have been chosen to provide insight based upon the results of the analysis but also to show the various different ways the data can be displayed. These charts include:

- A stacked bar chart to show positive, neutral and negative reviews for all features, as shown in Figure 10 a) and b). Through this figure, we can immediately understand how customers feel about specific aspects of a product. It is easily to see that most customers care about “Fit” and “Construction” with the majority positive reviews in Figure 4 a).

- A pie chart to show multiple products with positive, neutral and negative for each feature, as shown in Figure 11. This chart allows users to quickly understand which products, categories and features have strong positive or negative opinions.

- A column chart to show multiple products with positive, neutral and negative for each feature to enable comparison among the product features, as shown in Figure 12.

- A geo-location chart showing positive, neutral and negative reviews by country of manufacturers, as shown in Figure 13.

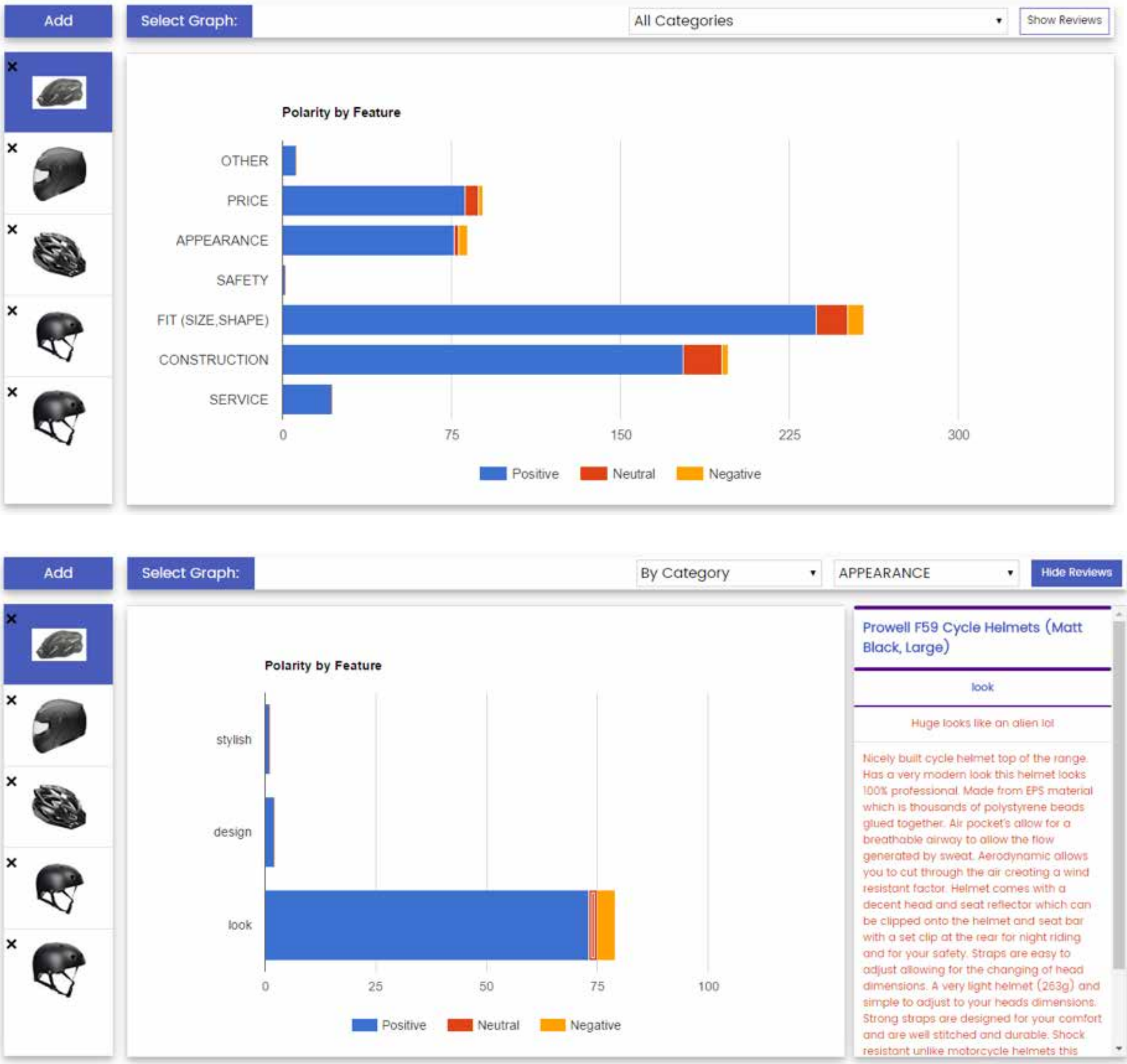



Figure 10 A stacked chart to provide summarisation of polarity for each category of a given product: polarity by category: a) without showing review texts, and b) with review texts.

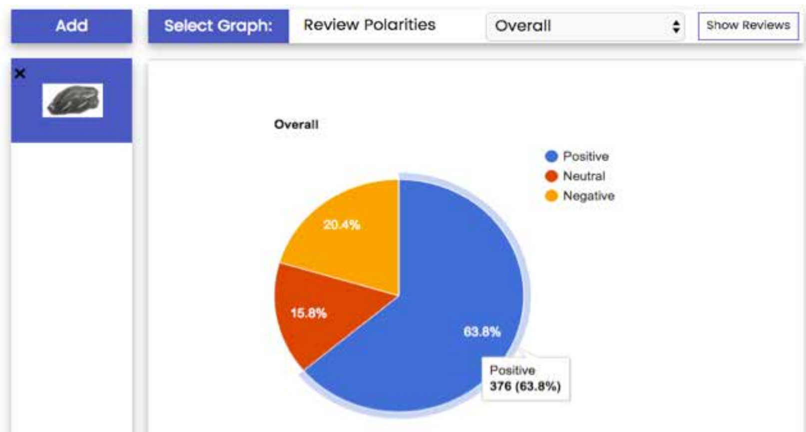



Figure 11 A pie chart to summarise the polarity of reviews

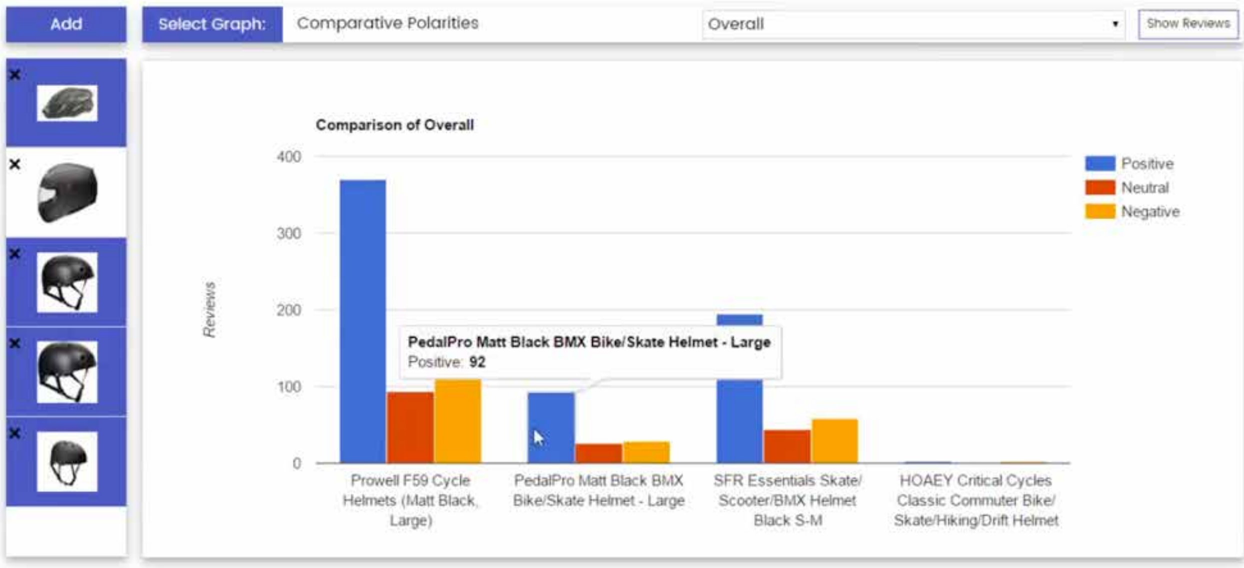



Figure 12: A column chart to show multiple product features for easy comparison

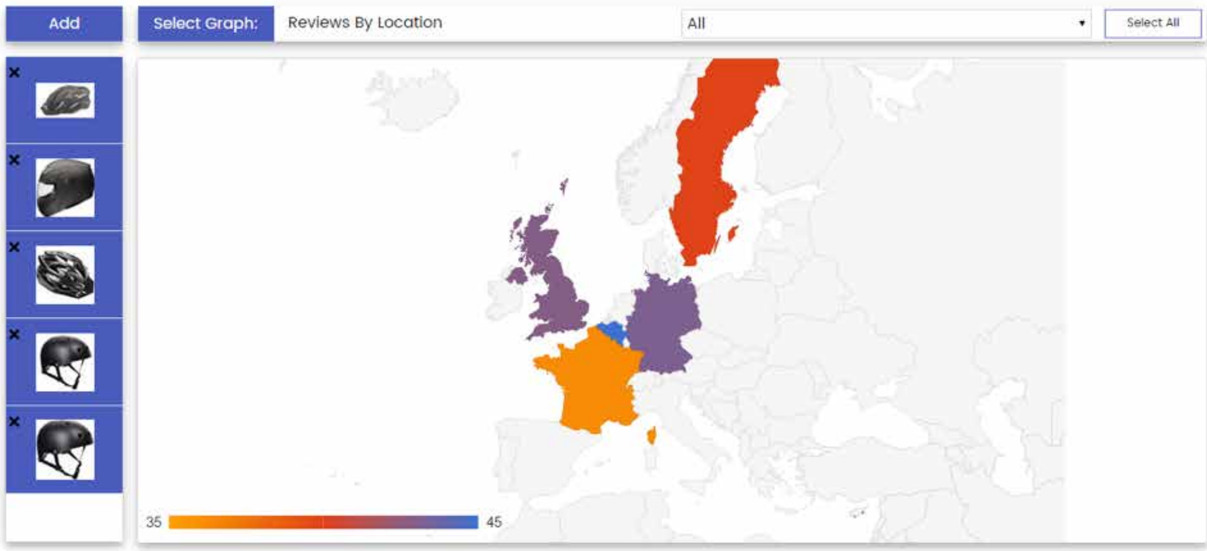



Figure 13: A geo chart to show positive, neutral and negative reviews by country of manufacturers (The colour of the country represents a normalised polarity)

## Summary

Customers are at the centre of the changes to value chains, products and services. Using smart data analytics is an efficient way to understand and meet their needs. This feasibility study has proposed a scalable, user-friendly system to enable fast automated extraction of product features and classification of polarity of product reviews for improved customer insights. The experimental evaluation shows that the proposed system has high accuracy on feature aspect-level sentiment analysis. The relationships between the number of reviews/words and the total execution time under different CPUs have been investigated. It has shown a strong correlation with a linear relationship, which demonstrates the scalability of the system.

## Workshop outputs

As part of the activities, two workshops (in Cardiff and Manchester) were organised to present the outputs of the SoC-UDRdM project, in total there were some 30 attendees from business and academia. The common theme arising from both of the workshops was that there was a large knowledge and capabilities barrier preventing SMEs from even beginning to utilise big data analytics (let alone maximising the potential benefits) to improve their products and customer service. Some of the key points arising from the workshops include:

- "Little data" might be a more appropriate term for the application of big data analysis techniques to SME data
- Big data is context sensitive e.g. an SME that uses computational modelling and CAD tools will have considerably more data than one that doesn't – yet relatively speaking they both have big amounts of data
- Even large SMEs struggle with "little data" let alone "Big"
- SMEs need to know: What they can do; Where the data is; How they can analyse the data
- Customer insights aren't guaranteed to deliver the "next big thing", they have a tendency to be reactive rather than proactive
- Insights needed into how customers use a product rather than just what they want or think of a product
- Sometimes what customers think they need from a product can actually be different to what they actually need and big data could help find this out
- Social media is becoming increasingly important for feedback (both good and bad) for both B2C and B2B customers
- Big data (plus the Internet of Things) is a precursor for moving towards a product-service system model of operating
- The customer insight system seems to be a good starting point to achieve the type of application an SME would be looking for
- SMEs are concerned about choosing the wrong tools and techniques for analysing because they have insufficient knowledge regarding big data
- Cost is a concern, SMEs are not always in the position to be able to spend large amounts of money on big data tools and even if they did they would not necessarily have the human resources to configure, run and maintain such a system
- Cloud-based tools that are low cost with access to sources of product review data such as Amazon or other internet retailers appear to be desirable

## Conclusions

This feasibility study has investigated the use of big data tools in product design and improvement from the perspective of SMEs. It has shown that although SMEs have product and customer data, and that they are aware of big data they don't know where to begin. They feel that the term big data is overwhelming and doesn't reflect the reality of SMEs and that "little data" might be a more appropriate term for SMEs and that what a large volume of data is will vary from sector to sector. Therefore in one sector 500Gb could be considered a very large volume of data in another 10Tb could be considered very large.

SMEs are particularly concerned about choosing the wrong tools and techniques for analysing data because they have insufficient knowledge and do not necessarily have the time to invest in investigating different solutions. As well as the fear of not selecting the solution that because the industry or market standard and that they have to spend money on changing systems. However, where presented with the prototype system they were very positive about the concept and could begin to see how big data could be of use for them.

Overall the idea of big data analytics for gaining customer insights was well received but SMEs will need both guidance and the suitable tools before they will be able to make the most of it.

## References

[1] Pearson H, Noble G, Hawkins J, (2013) Re-distributed Manufacturing workshop report, EPSRC.

[2] Prendeville S, Hartung G, Purvis E, Brass C, Hall A, Makespaces: (2016) From Redistributed Manufacturing to a Circular Economy, Sustainable Design and Manufacturing, pp. 577-588: Springer, 2016.

[3] Ward M, Rhodes C, (2014) Small businesses and the UK economy, Report, House.

[4] Ghobadian A, Gallear DN, (1996) Total quality management in SMEs, Omega; 24 (1): pp. 83-106.

[5] Russom P, (2009) Next generation data warehouse platforms, TDWI Best Practices Report: fourth quarter.

[6] Schaffer J, Schleich H, Scavarda LF, Parry G, (2008) Automotive production: product variety and how emerging economies can avoid problems of industrialised countries, 16th GERPISA international Colloquium: The Automobile Industry and Sustainable Development: Concepts and Doctrines, Public Policies and Company Strategies: 32.

[7] Schmidt R, Möhring M, Strategic Alignment of Cloud-Based Architectures for Big Data. pp. 136-143.

[8] Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH, (2011) Big data: The next frontier for innovation, competition, and productivity.

[9] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, (2015) Semeval-2015 task 12: Aspect based sentiment analysis, in the 9th International Workshop on Semantic Evaluation (SemEval 2015).

[10] N. Archak and A. G. P. G. Ipeirotis, (2007) Show me the money! deriving the pricing power of product features by mining consumer reviews, in KDD '07 Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 56-65.

[11] A. Duric and F. Song, (2012) Feature selection for sentiment analysis based on content and syntax models, Decision Support Systems, vol. 53, no. 4, pp. 704-711.

[12] A. Ghose and P. G. Ipeirotis, (2011) Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics, IEEE Transactions on knowledge and data engineering, 23 (10), pp. 1498-1512.

[13] M. Burton, Jamie; Khammash, (2010) Why do people read reviews posted on consumer-opinion portals, Journal of Marketing Management, 26 (3), pp. 230-255.

[14] D. H. Carter, (2015) Inferring asepect-specific opinion structure in product reviews, Master's thesis, School of Electrical Engineering and Computer Science.

[15] X. Fang and J. Zhan, (2015) Sentiment analysis using product review data, Journal of Big Data, 2 (1) pp. 1-14.

[16] B. Liu, (2012) Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan and Claypool Publishers.

[17] Stanford CoreNLP -- a suite of core NLP tools, http://stanfordnlp.github.io/CoreNLP/, Retrieved on 12, March 2017.

[18] C. D. Manning, J. B. Mihai Surdeanu, J. Finkel, S. J. Bethard, and D. McClosky., (2014) The stanford corenlp natural language processing toolkit, in the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

[19] C.-C. Chang and C.-J. Lin, LIBSVM -A Library for Support Vector Machines, retrieved on 12, March 2017.

[20] DROPWIZARD, http://www.dropwizard.io/1.0.6/docs/

To quote this report, please use the following reference: Ying L., Liangxiu H. (2017). SME-Oriented Collaborative Customer Insights Platform for user-driven RdM (SOC-UDRdM).

Doi:10.17862/cranfield.rd.5226373


USER-DRIVEN
YING LIU + ANTHONY SOROKA
SME'S
AIM TO CREATE A PROTOTYPE PLATFORM
SME SURVEY
WHAT IS THEIR AWARENESS AROUND BIG DATA
WE EXPECT THERE TO BE MORE FOCUS ON SME'S POST BREXIT
NEW POLICIES
LACK OF AWARENESS/ USE OF CRM
50% STORING DATA ON CLOUD
THEY'RE ALL HOLDING DATA BUT NOT USING IT FOR CUSTOMISATION
DATA
THEY NEED TO KNOW WHAT THEY CAN DO..
& HOW TO ANALYSE THE DATA
WHAT ABOUT 'LITTLE DATA'
HEY!!
STILL HAS VALUE
SOMETIMES CUSTOMERS DON'T KNOW WHAT THEY NEED
SUBSCRIPTION TOOLS ARE ATTRACTIVE TO SME'S AS THEY REDUCE RISK
WHAT DO I DO WITH IT...?
OUR TOOL ENABLES
PREDICTS WHAT PEOPLE WANT FROM THEIR REVIEWS
VISUALISATION
ANALYSIS
WHERE TO BEGIN WITH BIG DATA
ARE YOU SURE YOU WANT TO CANCEL YOUR SUBSCRIPTION
BUT THEY DON'T KNOW WHAT TOOLS EXIST!
WE ARE WORKING ON A TOOL THAT IS ACCESSIBLE WITH A FRIENDLY INTERFACE
ANALYTICS
CUSTOMER REVIEWS
CUSTOMISATION
CASE STUDY
UK BASED SME PRODUCING HORSE RIDING HELMETS
FACING COMPETITION FROM AMAZON.
HIYA
+ GIVES A MORE REALISTIC IDEA