**Bridging Customer Knowledge to Innovative Product Development:**

**A data mining approach**

**Abstract**

In the big data era, firms are inundated with customer data, which are valuable in improving services, developing new products, and identifying new markets. However, it is not clear how companies apply data-driven methods to facilitate customer knowledge management when developing innovative new products. Studies have investigated the specific benefits of applying data-driven methods in customer knowledge management, but failed to systematically investigate the specific mechanics of how firms realized these benefits. Accordingly, this study proposes a systematic approach to link customer knowledge with innovative product development in a data-driven environment. To mine customer needs, this study adopts the Apriori algorithm and C5.0 in addition to the association rule and decision tree methodologies for data mining. It provides a systematic and effective method for managers to extract knowledge “from” and “about” customers to identify their preferences, enabling firms to develop the right products and gain competitive advantages. The findings indicate that the knowledge-based approach is effective, and the knowledge extracted is shown as a set of rules that can be used to identify useful patterns for both innovative product development and marketing strategies.

*Keywords: product development, customer knowledge management, data mining, fast-cycle industry*

**1. Introduction**

In the big data era, firms are inundated with customer data, which are valuable for improving services, developing new products, and identifying new markets (Cooper, 2014; Zhan et al., 2016; Shimomura et al., 2018). New product development (NPD) with big data raises two critical issues. First, although the importance of customer knowledge management in product development has been widely recognized (Blazevic and Lievens, 2008; Cooper, 2014; Zhang et al., 2017), it has not been discussed in great depth, and the application of customer knowledge to product development has been under-examined (Khodakarami and Chan, 2014; Davenport, 2015; Kuo and Kusiak, 2018). Second, novel data-driven methods are required to ensure the active integration of customers’ knowledge into product development (Joshi and Sharma, 2004; Xu et al., 2016; Tan and Zhan, 2017). The specific benefits of applying data-driven methods in customer knowledge management include a greater opportunity to incorporate the latest technology, increased market share, higher profit, and more accurate forecasts of customer needs (Manyika et al., 2011; Wong, 2012; Davenport, 2015; Zhan and Tan, 2018). While providing high-level evidence of these benefits, studies have failed to systematically investigate the specific mechanics behind how firms can realize these benefits.

Recent studies have shown that fully commercialized new products have a remarkably high failure rate of 40-50%, and that this performance has not changed much over the past 20 years(Castellion and Markham, 2013; Zhan et al., 2017). Given the high costs associated with NPD, failure rate minimization has received considerable theoretical and managerial interest. Customer knowledge management, or the management of the processes a firm uses to capture, store, organize, access, and analyze data related to its customers, has been identified as a prerequisite for the success of a new product (Cooper; 1994; Joshi and Sharma, 2004; Fidel et al., 2015; Olson, 2018). Nevertheless, there has been disagreement on its importance and considerable variance in the extent to which companies engage in customer knowledge management in their product development (Cooper, 1994; Joshi and Sharma, 2004; Chang and Taylor, 2016).According to Liao et al. (2012), customer knowledge is generally concealed within the customer, and even when that knowledge has been extracted, inefficient management approaches can leave collected customer data in “data dumps.” Moreover, the potential for customer knowledge management in product development has not been studied in great depth (Su et al., 2006), and little attention has been paid to the application of customer knowledge management in innovative product development (du Plessis and Boon, 2004; Khodakarami and Chan, 2014). Although many studies have sought to introduce the concept of customer knowledge management by comparing it with other management concepts such as customer relationship management (du Plessis and Boon, 2004; Shang et al., 2011; Khodakarami and Chan, 2014), few have discussed the extraction and application of customer knowledge through the use of specific IT-based procedures and data-driven approaches for real-company projects. According to Davenport (2015), studies of customer knowledge should focus on making products or services more attractive to increase their value. Therefore, it is important to find methods for extracting and applying customer knowledge to improve product development and ensure business excellence.

More importantly, studies have increasingly recognized that effective customer knowledge management must be based on accurate identification of the preferences and requirements of customers (Joshi and Sharma, 2004; Shimomura et al., 2018). In particular, data mining techniques can help managers to extract unknown patterns and gain a better understanding of their customers, while a systematic knowledge management effort can channel information about and from customers into effective business strategies (Yee Liau and Tan, 2014). This makes the study of customer knowledge management extremely valuable for NPD (Khodakarami and Chan, 2014). According to Anand and Buchner (1998), data mining can be defined as “*the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large data sets*.” It allows companies to give accurate and effective feedback to research and development (R&D) managers, and companies can use it to provide a rapid response on fast-cycle products and identify their customers’ real preferences (Bae and Kim, 2011).

Today, new products do not remain so for long, as the technologies going into them change more quickly than they have in the past (Williamson and Yin, 2014). This is especially the case in fast-cycle industries, in which product life cycles are short—often only a few years or even less (Brexendorf et al., 2015). Most of the companies within fast-cycle industries are high-tech or electronics companies, and these are important drivers of innovation and economic growth (Brexendorf et al., 2015; Şeref et al., 2016). Such companies aim to develop innovative products faster, make decisions quicker, and convert customer needs into new product designs faster than their competitors (Şeref et al., 2016; Goswami, 2017). In particular, innovative product development requires the identification of key features or advanced technologies (Su et al., 2006). Indeed, key product features have been considered the most crucial aspect of product design and development (Bower and Hout, 1988; Goswami, 2017), and these may directly influence production costs and customer satisfaction, in addition to companies’ differentiation strategies (Bower and Hout, 1988; Su et al., 2006; Bae and Kim, 2011; Brexendorf et al., 2015). The preceding considerations lead to the following research question: how can managers extract customer knowledge and use it to enhance innovative product development in a fast-cycle industry?

The customer knowledge management model has received much attention because it can combine both the data-driven approach to customer relationship management and the people-oriented approach to knowledge management, with a view to exploiting their potential synergy (Su et al., 2006; Khodakarami and Chan, 2014; Xu et al., 2016). However, most research has focused on the interaction between customer relationship management systems and customer knowledge; that is, the concern has generally been how customer knowledge can be gained via customer relationship management systems (Iriana and Buttle, 2007; Xu et al., 2016; Khodakarami and Chan, 2014). There is no systematic approach to integrating the discovery of customer knowledge with the implementation of the knowledge, especially in the context of innovative product development in fast-cycle industries. Although data mining studies have mainly focused on the techniques used to mine the data (Anand and Büchner, 1998; Bae and Kim, 2011), customer knowledge management studies have focused on interfacing with the customer and the strategies used to manage the company’s interactions with customers (Joshi and Sharma, 2004; Zhang et al., 2017). To fill this gap, the present study proposes a systematic approach to the integration of customer knowledge management with data mining techniques for innovative product development. In particular, the study identifies approaches to customer knowledge management and examines their direct applicability to real product development projects in a fast-cycle industry. Its findings should help managers to identify customer preferences based on the knowledge extracted “from” and “about” the customers, rather than from generalizations about their characteristics.

**2. Literature review**

*2.1 Evolution of customer knowledge management in NPD*

Table 1 summarizes the progress made to date in NPD, historically categorized into four generations (Rothwell, 1994; Niosi, 1999; Cooper, 1994, 2014; Miller, 2001; Olson, 2018).

======Insert Table 1 about Here======

As shown in Table 1, the first generation spans the post-Second World War period to the mid-1960s. During this period, NPD emerged as an activity characterized by a high degree of uncertainty and serendipity. Within the corporation, scientists and engineers in the research lab studied materials, products, and processes and sometimes produced a novel idea that companies could introduce into production (Niosi, 1999). In particular, companies had no customer connections/knowledge management activities, and the industrial process was generally perceived as a linear progression from scientific discovery to technological development in companies and the marketplace. This development has been called “technology-push NPD” (Rothwell, 1994).

 Toward the second half of the 1960s, general levels of prosperity increased in the industrialized world and manufacturing productivity increased considerably (Rothwell and Soete, 1983). Organizations put more emphasis on marketing and corporate diversification to survive the intensifying competition. During this period, Polanyi (1966) defined knowledge management according to the concepts of explicit and tacit knowledge. New products were still largely based on existing technologies, and the arrival of the second generation did not change the linearity of the NPD process. However, it did introduce some order, reduce uncertainty, and increase the market focus of the corporate R&D lab by considering customer satisfaction (Niosi, 1999).

From the late 1970s to the early 1990s, third-generation NPD introduced in-house feedback and the integration of R&D within corporate strategy. During this period, it became increasingly necessary to understand the basis of successful NPD to reduce the incidence of wasteful failures, and indeed it was approximately during this period that the results of a number of detailed empirical studies of the NPD process were published (Rothwell, 1994; Cooper, 1990, 1994). Most of the models identified were logically sequential but did not necessarily represent a continuous process that could be divided into a series of functionally distinct but interacting and interdependent stages. Furthermore, companies tended to allocate resources to activities by assigning them budgets, but the accuracy of the budgetary allocation depended on knowledge of the scope and extent of the activity (Houlihan, 1985). The third-generation NPD process can be thought of as a complex set of communication paths; it represents the confluence of technological capabilities and market needs within the framework of the innovating company (Rothwell and Soete, 1983).

In the early 1990s, the fourth generation introduced two of the salient features of NPD in leading Japanese companies: integration and parallel development. These days, companies integrate suppliers into the NPD process at an early stage while at the same time integrating the activities of the different in-house departments involved, which work on projects in parallel rather than in series (Maidique and Zirger, 1985). These two factors have shortened product development life cycles and made NPD more effective. In addition, customer knowledge management has been described as an emergent process because novelty emerges at each stage of NPD, on the basis of responsiveness to customer feedback (Nonaka and Takeuchi, 1995; Joshi and Sharma, 2004). Advances in information and communication technologies (ICTs) and data-driven methods are enabling new initiatives to be explored and are materially transforming the customer knowledge management domain (Khodakarami and Chan, 2014). In particular, data from different sources can be captured to support customer-centric product development, which focuses on building and maintaining long-term customer relationships and on creating a sense of loyalty by providing customers with products and services they value, for mutual benefit.

In summary, the evolution of NPD has been characterized by (1) a shortening of product life cycles, (2) increased understanding of customer knowledge management, and (3) more widespread application of data-driven methods for product development. To gain a competitive advantage, it is imperative that the integration of customer knowledge into a company’s product development projects be institutionalized (Davenport, 2015; Zhan et al., 2017; Wynn and Eckert, 2017). Moreover, data-driven activities are strongly linked to the discovery of customer knowledge, and firms can achieve important cost benefits by learning to apply data mining methods within NPD (Fidel et al., 2015; Kuo and Kusiak, 2018).

*2.2 Data mining in customer knowledge management*

Recent research has increased understanding of the implementation of customer knowledge management and data mining methods in NPD (Khodakarami and Chan, 2014; Liau and Tan, 2014; Xu et al., 2016). With data mining techniques, unknown patterns, previously unrecognized trends, and salient and hidden information can be discovered from a massive amount of data (Anand and Büchner, 1998; Bae and Kim, 2011; Kuo and Kusiak, 2018; Aich et al., 2018). In particular, there has been a significant increase in attempts to adopt data mining techniques to support customer knowledge extraction for different aspects of NPD. Table 2 provides a summary.

======Insert Table 2 about Here======

All these of studies proved that the appropriate application of data mining techniques could not only extract salient but normally hidden customer knowledge, but also be used to produce useful representations of that knowledge to support managers in NPD. Nevertheless, little has been written about how customer knowledge management can be integrated into innovative product development in a fast-cycle industry. Moreover, most data mining studies have instead focused on the identification of different groups of customers to explore their characteristics (Tsai et al., 2003; Liao et al., 2012; Shimomura et al., 2018). The management of customer knowledge is usually associated with different phases of product development (i.e., the generation of concepts and ideas, product design and manufacturing, market testing and product launch), each of which requires different data-driven methods (Joshi and Sharma, 2004; Piri et al., 2017). Therefore, these methods need to be detailed within a systematic approach that can help managers develop innovative products that fulfill customers’ unmet needs. The present study examines two data mining techniques: association rules with the Apriori algorithm (Schuetz, 2015) and decision trees with C5.0 (Ball and Brunner, 2012). The Apriori algorithm can greatly compress the candidate item sets and the size of the frequent item sets, and performs better than other algorithms (Gu et al., 2015). Moreover, the C5.0 method boosts the accuracy of the model, and also improves the efficiency of execution and reduces the demand for internal memory (Bae and Kim, 2011). A detailed introduction to these techniques is not given here; instead, only measurements that affect their performance are illustrated.

*2.2.1 Association rules*

Association rules are a well-studied method for identifying unknown relationships between variables in large datasets (Liu, 2009; Schuetz, 2015; Guo et al., 2017). There are two parts to an association rule: an antecedent and a consequent. The antecedent is an item that can be found in the datasets, and the consequent is an item that is combined with the antecedent. In this way, the antecedent can be considered as “if” and a consequent can be taken as “then.” Therefore, association rules exist in the form “if X then Y.” Let items I = {i1, i2 … ik}. T is a set of transactions, and each t in the set of T is a set of I. Thus, T ⊆ I. The form of an association rule is X → Y, where X ⊂ I, Y ⊂ I, and X ∩ Y = ∅. Measures such as support and confidence permit evaluation of the quality of the extracted rules. The confidence is the proportion of transactions that contain X and also contain Y. Usually, the aim is to create an association rule that has the minimum support and minimum confidence (Liu, 2009; Lee et al., 2018). For instance, if a rule satisfies both minimum support and minimum confidence, it can be called a strong association rule (Schuetz, 2015). In addition, lift indicates a ratio of confidence to expected confidence, which is another parameter for an association rule. Lift reflects the relevance of X and Y. If the value of lift is greater than 1, there is a strong positive correlation between X and Y (Cariñena, 2013).

*2.2.2 Decision trees*

According to Ball and Brunner (2012), the decision tree has become one of the most common classification techniques and practical methods for data mining. It involves drawing a tree-structured diagram for knowledge discovery based on relationships within a massive amount of data (Mingers, 1989). The tree can comprise several branches and nodes, where each branch of the decision tree can be a possible outcome. Each decision tree has a large number of leaf nodes and a root node on the top. Every single non-leaf internal node refers to a test of an attribute, and each branch reveals the results of the test. Decision trees have many benefits compared with other data mining techniques. For example, the tree-shaped representation can organize knowledge as a good structure, which makes the results more comprehensive and easier to understand (Ball and Brunner, 2012). The main methods applied for decision tree analysis are (1) ID3 (Quinlan, 1979) (2) CHAID (Kass, 1980) (3) CART (Breiman et al., 1984); (4) QUEST (Loh and Shih, 1997); and (5) C4.5 and C5.0 (Quinlan, 1992).

Today, association rules and decision trees play a significant role in effective and efficient decision support management for businesses. For instance, Piri et al. (2017) analyzed data from over 1.4 million diabetics and developed a clinical decision support system based on a decision tree for diabetic retinopathy; Aich et al. (2018) proposed a decision-tree-based approach to predict Parkinson’s disease using different feature sets of voice data; Guo et al. (2017) designed an improved association rule algorithm in a mobile e-commerce recommendation system for improved customer satisfaction and data mining efficiency; and Lee et al. (2018) introduced an intelligent fuzzy decision support system based on the association rule for a flexible adjustment of dye pricing to manage customer-supplier relationships. In product development, significant advantages accrue because products are developed quickly, resources are used more creatively and efficiently, costs are reduced, customer needs are identified, and work-in-progress bottlenecks are minimized (Su et al., 2006; Xu et al., 2016; Cui and Wu, 2016).

**3. A proposed data mining approach**

Customer knowledge is a critical asset, and capturing, extracting, and understanding customer knowledge for NPD can be valuable competitive activities. This study proposes an approach based on data mining techniques to extract customers’ knowledge from customer demand data and apply it as a resource for developing innovative new products. In particular, studies such as Lee et al. (2012), Su et al. (2006), and Bae and Kim (2011) have suggested the approach of combining association rules and decision trees. However, the main purpose of this study is to extend the prior knowledge by further augmenting and recasting the conceptual basis of customer knowledge management and data-driven product development through conducting an innovative smartphone development project. This section illustrates the overall process of the proposed approach and provides a brief introduction to each phase, as shown in Figure 1. The proposed approach has three main phases. The first phase is data collection. The second phase involves data pre-processing and the application of data mining techniques to NPD. The final phase involves evaluating the results and making product development decisions.

======Insert Figure 1 about Here======

In phase 1, a pre-test with researchers, smartphone manufacturers, and relevant vendors is performed before the main test to determine the key product features that may influence the purchase of smartphones. A Web-based questionnaire is used for sample selection, survey conducting, and data collection. To reduce sampling bias and improve accuracy in predicting the target population, a simple random method is applied during data collection (McKay et al., 1979). In phase 2, the collected data are pre-processed in several stages, including data cleaning, data selection, and transformation. The collected data take the form of flat files, which are cleaned to remove discrepancies and inconsistencies and then compiled into one workable table so that different data mining methods can be applied. An association rules algorithm and decision tree modeling are used to produce sets of rules that are extracted using the Apriori algorithm and C5.0, respectively. C5.0 presents a key advantage in the test selection and evaluation process (Quinlan, 1992). It boosts model accuracy, and both its execution efficiency and internal memory have been improved (Lee et al., 2012). In the final phase, these methods are assessed based on the usability, validity, and reliability of the integrated rules for NPD. The assessment determines whether the methods can be widely applied in the future (McKay et al., 1979), and whether useful information can be extracted and turned into knowledge that helps managers to understand their customers’ preferences and make better NPD decisions.

*3.1 Data collection and analyses*

A leading Chinese electronics company initiated an innovative smartphone development project to explore opportunities for its industry and expand its smartphone market. The company aimed to serve the Chinese market initially and then expand its business internationally, ultimately to the global market, through the benefits of economies of scope and scale. The project showed that innovative product development required not only the integration of complex technology, but also an understanding of customers’ knowledge. First, to identify the key product features that met customers’ preferences, pre-test analysis was performed before the main test to determine those key features that influenced the purchase of smartphones. The participating managers, researchers, and vendors were experts and had many years’ experience in the smartphone industry. Based on their suggestions (as shown in Table 3), nine features of smartphones that could influence customer purchases were identified.

======Insert Table 3 about Here======

The main test was conducted with a Web-based questionnaire developed based on the pre-test results. Advanced ICTs can facilitate effective and efficient communication by sending out an e-mail to potential respondents who play a key role in the purchase decision and asking them to answer a questionnaire posted on a website. Most of the respondents had already purchased or wanted to purchase new smartphones. In addition, most of the respondents were postgraduate/Master of Business Administration (MBA) students. Due to the difficulties involved in obtaining data, studies have often used executive MBA audience samples (Zirger and Maidique, 1990; Zhu and Sarkis, 2004; Chen and Tan, 2015). In this study, after data cleaning, 167 usable and unique questionnaires were available for analysis. Hutcheson and Sofraniou (1999) suggested that research based on questionnaires required a sample of at least 150 subjects. Thus, the present sample of 167 was considered large enough to produce statistically significant results. To test for non-response bias, the first one third and last one third of the respondents were compared based on their personal characteristics (such as age, employment status, and Internet access per day), and no statistically significant differences were found. This provided evidence that there was no non-response bias in the data.

 The webpage with the study questionnaire first briefly introduced the study and thanked respondents for taking the time to answer the questionnaire; the respondents were then asked to complete the questionnaire, which was divided into two main parts. The first part collected personal information such as age, marital status, educational level, employment status, monthly income, and Internet access per day. The second part related to the ideal innovative features of smartphones. The results from the first part of the questionnaire showed that 92 of the respondents (55.0%) were male and 75 (44.9%) were female. Most of the respondents were in their 20s (93, 55.7%), 55 (32.9%) were in their 30s, and 19 (11.4%) were aged over 40. Most of the respondents were students at either the undergraduate (31, 18.6%) or postgraduate/MBA (106, 63.5%) level; the remainder were employees (30, 18.0%). The details of the respondents’ personal information are shown in Table 4.

======Insert Table 4 about Here======

The second part of the questionnaire listed the nine innovative features of smartphones identified in the pre-test. The respondents were asked to rank these features from 1 to 9 according to their importance, with 1 being the most important feature (i.e., the feature that respondents were most concerned with when purchasing a smartphone) and 9 the least important. The respondents were also asked to pick out the top three important features and select a favorite type of smartphone from three popular types (Figure 2). In particular, the three smartphone types were proposed by a Chinese electronic company focusing on different market segments: Type A was a sophisticated-looking smartphone that included a physical keyboard, and so was designed for customers who still preferred using a physical keyboard on the phone; Type B was a palm-sized and easy-to-use smartphone (with a relatively lower price) targeting elderly people and customers who might not have cared much about the advanced functions of the smartphone they purchased; and Type C followed one of the most popular smartphone styles (with great processing power and an excellent multi-touch user interface) for attracting young adults and high-level customers.

The relationship between the nine features and three types of smartphone was then analyzed by running an association rule algorithm and applying decision tree modeling. Table 5 shows the dataset to which the association rule analysis was applied, while Table 6 displays the dataset to which the decision tree modeling was applied.

The association rule algorithm was used to identify the relationships between different types of smartphone and their key features. Table 5 details the dataset used; it includes 167 groups of data, with each comprising 4 rows. The first column shows the ID of each respondent. The second column shows the three most important features and the type of smartphone the respondent selected. To facilitate the reading of the data in the analysis, the third column gives the abbreviation used for each feature. SAS Enterprise Miner was used to perform the analysis. SAS Enterprise Miner is an open-source integration with R and can run accurate descriptive and predictive analyses on massive datasets. Thus, managers find the generated results easy to read and can use them to support their decision making.

======Insert Figure 2 and Tables 5-6 about Here======

Table 6 shows the dataset to which decision tree modeling was applied. It includes 167 samples reflecting the respondents’ preferences for different smartphone features and types. For instance, the feature “color” had five different options: (1) white (denoted CL1), (2) gold (CL2), (3) silver (CL3), (4) black (CL4), and (5) does not matter (CL5). In particular, the smartphone features could be seen as “If,” and the style was defined as “Then.” Decision tree modeling aimed to generate rules in a format, such as the kinds of features that led the respondents to choose the relevant type of smartphone product. This was achieved by splitting the data source into subsets according to attribute-value analysis. The C5.0 package was applied to analyze whether the respondents’ selections of particular features were related to their choice of smartphone type (i.e., purchase decision). IBM SPSS Modeler 18.0 was used to conduct the decision tree modeling. IBM SPSS Modeler 18.0 is a data mining software package that features visual programming and a data flow interface. It can work with various types of data and so provides organizations with extensive data mining capabilities to deal with various specific problems.

**4. Findings**

Tables 7 and 8 respectively present the findings from the association rules algorithm and decision tree modeling for the smartphone development project. The results are produced by identifying different decision variables and applying them to customers’ purchasing preferences to acquire knowledge of consumers’ product expectations. In the association rule analysis, the lift value is set at greater than 1, while the minimum support and initial confidence values are set at 5% and 30%, respectively, and then adjusted as necessary during analysis.

======Insert Tables 7-8 about Here======

*4.1 NPD: Type A smartphone*

As Table 7 shows, the association rules for the type A smartphone have support (5.39-16.77%) and confidence (30.03-47.18%). All of the support and confidence values for these rules are above the threshold, and therefore these association rules are valid and relevant. All of the lift values are above 1, and therefore the attributes within each association rule are positively correlated. When companies design and develop a type A smartphone, these rules are as follows: (1) screen resolution → type A; (2) material → type A; (3) weight → type A; (4) battery → type A; (5) color → type A; (6) screen resolution and material → type A; (7) size and screen resolution → type A; and (8) screen resolution and camera resolution → type A. Four of the eight association rules include screen resolution, which indicates that the strongest association exists between the feature screen resolution and the type A smartphone. Therefore, high screen resolution is the key feature to consider when designing and developing a type A smartphone. In addition, “material” appears twice in these eight rules, which is the second highest frequency, and therefore it can be seen as the second most important feature. The confidence and support values for the screen resolution and material rules are in the top two positions among all of the association rules, and so these features should be considered first when developing a type A smartphone.

 Additionally, the rules identified from decision tree modeling, as shown in Table 8, are as follows: (1) IF screen resolution = 1920p or above, and material = plastics and operating system = Android THEN type A; (2) IF material = does not matter THEN type A; and (3) IF operating system = does not matter THEN type A. There are 33 instances of rule (1) in the sample, and the percentage is 78.8%. There are 21 instances of rule (2) in the sample, and the percentage is 57.1%. There are nine instances of rule (3) in the sample, and the percentage is 44.4%. These three rules provide clear detailed specifications for further type A smartphone development. For instance, rule (1) suggests companies adopt screen resolutions of 1920p or above, use a plastic smartphone body, and adopt the Android operating system for their type A smartphone, as this should meet the preferences of most customers. Rule (1) has the highest probability and therefore is of most significance in helping companies to narrow down the specifications and functional features and enhance the effectiveness of their product development.

*4.2 NPD: Type B smartphone*

As Table 7 shows, association rules analysis produced three rules that should be considered when companies develop a type B smartphone: (1) price → type B; (2) weight → type B; and (3) battery → type B. The support values for price, weight, and battery are 11.98%, 10.18%, and 6.59%, respectively. The confidence values of the rules are 36.16%, 34.17%, and 32.04%, respectively. The confidence and support results indicate that price is the most important feature in choosing a type B smartphone, meaning companies should highly prioritize price in the development of these products. Moreover, the “weight” and “battery” features should be considered, as they have higher relevancy than other features in the total transaction dataset in the development of type B smartphones.

Only one valid rule was generated from the decision tree modeling for the type B smartphone (Table 8): IF weight = does not matter, and material = plastics THEN type B. There are 11 observations in the transaction set, and the percentage of the rule is 72.7%. The rule shows that most of the respondents who chose type B did not care about the smartphone’s weight, although they preferred a plastic body. The high probability of this rule clearly indicates what companies should focus on in their further development of type B smartphones and should help them to avoid unnecessary work on other features.

*4.3 NPD: Type C smartphone*

The association rules for the type C smartphone have high levels of both support (8.98-23.95%) and confidence (51.23-87.86%) compared with the type A and B smartphones, as shown in Table 7. The high support and confidence ratio indicates a significant association between the selected features and type C smartphone. In particular, the following rules must be considered in NPD: (1) size → type C; (2) operating system → type C; (3) camera resolution → type C; (4) price → type C; (5) operating system and camera resolution → type C; (6) size and material → type C; (7) size and operating system → type C; (8) material and camera resolution → type C; (9) operating system and material → type C; and (10) weight and screen resolution → type C. “Operating system” has the highest occurrence rate in these 10 rules, appearing 4 times. This indicates that the feature most strongly associated with the type C smartphone is the operating system, making it the key point to consider in developing a type C smartphone. In addition, companies should pay attention to the “material,” “size,” and “camera resolution” features, as these appear three times each, the second highest frequency in the findings. These four identified features should satisfy most customer preferences; companies can accordingly narrow the range of features to optimize their development of new type C smartphones.

 Decision tree modeling identifies the following rules (Table 8): (1) IF material = metal and operating system = IOS THEN type C; (2) IF screen resolution= 1920p or above, and material = metal THEN type C; and (3) IF material = does not matter THEN type C. These three rules provide a clear, detailed model for the further development and improvement of type C smartphones. For example, rule (1) suggests that companies use metal materials for the body and use the IOS operating system for their type C smartphones, as these are favored by customers. Rule (1) has the greatest probability level of the three, which indicates that it has the greatest potential to help companies in developing a new type C smartphone.

**5. Discussion and implications**

*5.1. Discussion*

According to the results shown in the previous section, customers focus on the screen resolution of type A smartphones, as most of the association rules are associated with “screen resolution.” Therefore, screen resolution should be considered as a critical feature in product development. Correspondingly, it can be assumed that these customers want good image editing functions, and these could also be the focus of attention in NPD. In addition, the frequency of “material” is relatively high in these rules; therefore, companies can consider this as a secondary feature. The combination of association rules and decision tree analysis produces a clear model for companies to develop or improve their products. Accordingly, the screen resolution can be seen as the dominant feature, and the development of any new type A smartphone should include the following: a screen resolution of at least 1920p, a plastic body, and an Android operating system. Beyond that, the decision tree results show that customers who do not care about the materials and operating systems used in their smartphones mostly choose type A; therefore, the appearance of Type A is highly competitive out of the three smartphones. In product improvement and development, companies do not need to make any significant changes to the exterior of a phone. Rather, companies must make high screen resolution their highest priority and try to use advanced plastics for the smartphone body to satisfy customer needs.

Compared with the two other types of smartphone, type B has the smallest number of association rules and decision tree results, which to some extent reduces the ability to apply these results to the development and improvement of these products. Even so, the association rules identified three critical features for type B smartphone development, namely price, weight, and battery. This indicates that customers who choose type B phones may chiefly pay attention to the cost-effectiveness of the phone and have no particular preferences in relation to its functionality. They prefer a smartphone with the basic attributes of a large-capacity battery and a light weight. Based on these features, some assumptions can be made about the customers. For example, most may have a relatively low income or work outdoors for extended periods. Assumptions of this kind may help companies to identify their customers and accordingly target their product development strategies. In addition, the decision tree results offer some guidance for product design, as they indicate that customers do not care about the weight of their phones, although they do prefer those phones to have a plastic body. Consequently, plastics are the best choice for the development of type B smartphones that meet the needs of most consumers, as they are affordable and lightweight.

Normally, when more rules are produced in the analyses, this indicates that customers have more diversified needs. Thus, product development for the type C smartphone is likely to be both difficult and complex. The results nevertheless suggest some key features—those items that appear most often in the rules—that companies should prioritize in their NPD: “operating system,” “size,” “camera resolution,” and “material.” “Size” and “material” are features of the smartphone’s appearance, while “operating system” and “camera resolution” can be defined as features of the smartphone’s function and performance. Therefore, consumers who favor type C smartphones may have comprehensive consumption demands, relating to both functionality and appearance. This in turn may lead to an assumption that most consumers of type C smartphones are teenagers and young adults. After determining the target customers in this way, the company can refine its development strategies and target its products more effectively. Furthermore, the design details of a type C smartphone can be specified based on the decision tree results. For instance, customers who focus on the material are likely also to consider the operating system in their choice of smartphone. In this study, most of the respondents selected metal as the ideal material for the type C smartphone. The rules from the decision tree analysis can be applied to the development of the type C smartphone as follows: (1) IOS operating system and metal body and (2) screen resolution of at least 1920p and metal body. In addition, a relatively small number of consumers would choose a smartphone with a plastic body and not care about the other features. In this way, if companies cannot focus on several different product features at the same time or do not have enough R&D investment to balance improvements in all of the different features, the development of a type C smartphone with a plastic body would also be a good choice.

The preceding results show, as expected, that by applying the proposed data mining approach, companies can use the knowledge extracted “from” and “about” customers, and identify the right knowledge “for” innovative product development. Table 9 summarizes the proposed three-phase approach to knowledge management and innovative product development. Potential tasks and roles that can be transferred to customer knowledge management are demonstrated in each of the phases.

======Insert Table 9 about Here======

Overall, by delineating customers’ preferences for particular product features, the different characteristics of each product type can be identified and analyzed. Therefore, the tacit customer knowledge, dispersed among the individual customers, is extracted, and companies can codify it into their desired explicit customer knowledge. The results indicate that paying more attention to customer demands and adopting data-driven methods to turn customer knowledge into innovative product ideas should help companies to compete with their rivals, especially in fast-cycle industries. The approach proposed in this study can produce effective suggestions at different stages of innovative product development. To be specific: Web-based surveys can be used to collect huge amounts of customer data (i.e., information and knowledge) that managers can use in their development of product concepts and ideas; association rules can identify key product features extracted from customer knowledge; and decision tree modeling can help companies to identify those details of product design that will have a bearing on fact-based decision making. Therefore, the best solution can be chosen from a massive range of permutations and combinations. Furthermore, the proposed approach not only increases the capacity to manage customer knowledge and accurately position products for particular target markets, but also increases the capacity for innovation by enhancing and extending company product development lines.

*5.2 Implications for practices and research*

The proposed data mining approach has several managerial implications. On the one hand, the results reveal that data mining techniques play an irreplaceable role in innovative product development. Rather than merely making data-driven decisions, companies should use data mining techniques and make knowledge-driven decisions. Therefore, turning data into knowledge is the key for companies to gain competitive advantages. On the other hand, this study proposes a data mining approach to harvest customer knowledge for innovative product development. Although it applies a smartphone case as a research object, this approach is developed with a general purpose. It can be further extended to other industries such as the digital product industry, which includes household appliances, digital cameras, and so on. Additionally, this study provides a practical procedure for managers to conduct their own innovative product development effectively. The findings are useful for manufacturers and product managers to extract critical product features and identify specific design details. The approach can also help managers to facilitate their product design by enhancing customer engagement during the development process. As product development is a process associated with significant fuzziness and uncertainties, managers must be aware of several critical issues (Zhan et al., 2016; Wynn and Eckert, 2017). For example, the extraction of customer knowledge may cause intellectual property issues; the new approach may disrupt a company’s existing organizational structure or NPD processes; or certain data mining expertise and knowledge are required to meet the company’s objectives through adopting the proposed approach. Thus, to balance the expected benefits and potential costs, it is important for managers to understand their actual needs and adopt the data mining approach carefully for their specific purposes.

 This study contributes to the knowledge management literature by developing an approach for innovative product development in a fast-cycle industry. The approach can be applied through all phases of NPD. In terms of theoretical contributions, this study extends understanding of the integration of customer knowledge management with product development (Joshi and Sharma, 2004; Su et al., 2006; Zhan and Tan, 2018), and proposes a systematic approach for integrating the customer knowledge extraction process with the management and application of that knowledge for innovative product development. A point of interest is its application of IT-based procedures and data mining techniques in a real innovative NPD project, and its attempt to diminish the “knowing–doing gap” that has been criticized in major knowledge management research (Pfeffer and Sutton, 1999; Choi et al., 2010). Moreover, the study identifies the data mining techniques, procedures, and processes of the approach and showcases the outcomes. The approach represents a concrete methodology for successful NPD in fast-cycle industries, enabling companies to move away from product-focused NPD, turn their attention to customer knowledge management, and apply data-driven methods to identify customers’ actual needs. Therefore, it extends the traditional NPD boundaries and provides evidence of the vital role of customer knowledge management.

**6. Conclusion, limitations and future research**

This study proposes a systematic approach for innovative product development based on the association rule and decision tree techniques. The findings show that the approach is effective. Customer knowledge was extracted using data mining techniques, and sets of rules were produced that could offer ideas and suggestions for innovative product development and possible strategic marketing solutions. In addition, the study indicates that, as companies in a fast-cycle industry, smartphone companies should pay more attention to not only advanced technologies but also the management of customer knowledge (i.e., customer demands and preferences) and use it as a resource to support better decision making in NPD.

Despite its main contributions, this study has limitations. First, using the questionnaire as a survey tool may increase the time and cost of data collection. Therefore, the data collection process should be improved such as by using product-service systems to avoid these problems (Carreira et al., 2013; Mourtzis et al., 2018). Second, the data used in this study mainly came from full-time postgraduate/MBA students, which might have biased the results. Furthermore, as discussed in the previous section, the analysis for the type B smartphone produced a relatively small decision tree and relatively few association rules than that for the other two smartphone types. Therefore, future studies should increase the sample size and include a wider range of respondents to improve research accuracy. Third, this study examines three popular types of smartphone (A, B, and C) to capture customer needs, and its findings should provide managers with detailed suggestions for product development. However, in real life, some customers may not care much about the type of smartphone they purchase. A future study should therefore add a “Type D” product whose features are all satisfactory. Finally, the findings of this study may not be generalized to other industries. This study related to the smartphone industry and did not investigate other industries with different types of products; this may make the proposed approach difficult to popularize or spread. Therefore, the suggested approach needs more testing in a wide range of product areas; indeed, this is essential for the approach to gain external validity and to be proved feasible in other areas.

**References**

Agard, B. and Kusiak, A., 2004. Data-mining-based methodology for the design of product families. *International Journal of Production Research*, *42*(15), pp. 2955-2969.

Aich, S., Younga, K., Hui, K.L., Al-Absi, A.A. and Sain, M., 2018. A nonlinear decision tree based classification approach to predict Parkinson’s disease using different feature sets of voice data. In *20th International Conference* *on Advanced Communication Technology (ICACT)* (pp. 638-642), IEEE.

Anand, S. and Büchner, A. (1998). *Decision support using data mining.* London: Financial Times Management.

Bae, J. and Kim, J. (2011). Product development with data mining techniques: A case on design of digital camera. *Expert Systems with Applications*, *38*(8), pp. 9274-9280.

Ball, M. N. and Brunner, J. R. (2012). Data Mining and Machine Learning in Astronomy. *International Journal of Modern Physics*, *19*, pp. 1049-1106.

Blazevic, V., and Lievens, A., 2008. Managing innovation through customer co-produced knowledge in electronic services: An exploratory study. *Journal of the Academy of Marketing Science*, *36*(1): *138-151*.

Bower, J.L. and Hout, T.M., 1988. Fast-cycle capability for competitive power. *Harvard Business Review*, *66*(6), pp. 110-118.

Breiman, L., Friedman, J., Olsen, R. and Stone, C. (1984). *Classification and regression trees*. New York: Chapman and Hall.

Brexendorf, T.O., Bayus, B. and Keller, K.L., 2015. Understanding the interplay between brand and innovation management: Findings and future research directions. *Journal of the Academy of Marketing Science*, *43*(5), pp. 548-557.

Cariñena, P. (2013). Fuzzy temporal association rules: Combining temporal and quantitative data to increase rule expressiveness. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *4*(1), pp. 64-70.

Carreira, R., Patrício, L., Jorge, R.N. and Magee, C.L., 2013. Development of an extended Kansei engineering method to incorporate experience requirements in product–service system design. *Journal of Engineering Design*, 24(10), pp.738-764.

Castellion, G. and Markham, S.K., 2013. Perspective: New product failure rates: Influence of argumentum ad populum and self-interest. *Journal of Product Innovation Management*, *30*(5), pp. 976-979.

Chang, W. and Taylor, S.A., 2016. The effectiveness of customer participation in new product development: A meta-analysis. *Journal of Marketing*, *80*(1), pp. 47-64.

Chen, R.Y., 2009. A problem-solving approach to product design using decision tree induction based on intuitionistic fuzzy. *European Journal of Operational Research*, *196*(1), pp. 266-272.

Chen, Z. and Tan, K., 2013. The impact of organization ownership structure on JIT implementation and production operations performance. *International Journal of Operations & Production Management*, *33*(9), pp. 1202-1229.

Chesbrough, H.W., 2006. *Open innovation: A new paradigm for understanding industrial innovation*, Berkeley: Oxford University Press.

Choi, S.Y., Lee, H. and Yoo, Y., 2010. The impact of information technology and transactive memory systems on knowledge sharing, application, and team performance: A field study. *MIS Quarterly*, pp. 855-870.

Cooper, R.G. 1990. Stage-gate systems: a new tool for managing new products, *Business Horizons*, *33*(3), pp. 44-54.

Cooper, R.G. 1994. Perspective: Third-generation new product processes, *Journal of Product Innovation Management*, *11*(1), pp. 3-14.

Cooper, R.G. 2014. What’s next? After stage-gate. *Research Technology Management*, *57*(1), pp. 20-31.

Cui, A.S. and Wu, F., 2016. Utilizing customer knowledge in innovation: Antecedents and impact of customer involvement on new product performance. *Journal of the Academy of Marketing Science*, *44*(4), pp. 516-538.

Davenport, T.H., 2015. *Process management for knowledge work. Handbook on Business Process Management* 1 (pp. 17-35). Springer, Berlin, Heidelberg.

du Plessis, M. and Boon, J.A., 2004. Knowledge management in eBusiness and customer relationship management: South African case study findings. *International Journal of Information Management*, *24*(1), pp. 73-86.

Fidel, P., Schlesinger, W. and Cervera, A., 2015. Collaborating to innovate: Effects on customer knowledge management and performance. *Journal of Business Research*, *68*(7), pp. 1426-1428.

Goswami, M., 2017. An integrative product line redesign approach for modular engineering products within a competitive market space: A multi-objective perspective. *International Journal of Production Research*, pp. 1-22.

Gu, X.F., Hou, X.J., Ma, C.X., Wang, A.G., Zhang, H.B., Wu, X.H. and Wang, X.M., 2015. Comparison and improvement of association rule mining algorithm. In *12th International Computer Conference on Wavelet Active Media Technology and Information Processing* (ICCWAMTIP) (pp. 383-386), IEEE.

Guo, Y., Wang, M. and Li, X., 2017. Application of an improved Apriori algorithm in a mobile e-commerce recommendation system. *Industrial Management & Data Systems*, *117*(2), pp. 287-303.

Houlihan, J.B., 1985. International supply chain management. *International Journal of Physical Distribution & Materials Management*, *15*(1), pp. 22-38.

Hutcheson, G.D., and Sofroniou, N., 1999. *The multivariate social scientist: An introduction to generalized linear models*. London, Sage Publications.

Iriana, R. and Buttle, F., 2007. Strategic, operational, and analytical customer relationship management: attributes and measures. *Journal of Relationship Marketing*, *5*(4), pp. 23-42.

Jiao, J., Zhang, L., Zhang, Y. and Pokharel, S., 2007. Association rule mining for product and process variety mapping. *International Journal of Computer Integrated Manufacturing*, *21*(1), pp. 111-124.

Jin, G., Jeong, Y. and Yoon, B., 2015. Technology-driven roadmaps for identifying new product/market opportunities: Use of text mining and quality function deployment. *Advanced Engineering Informatics*, *29*(1), pp. 126-138.

Joshi, A.W. and Sharma, S., 2004. Customer knowledge development: Antecedents and impact on new product performance. *Journal of Marketing*, 68(4), pp. 47-59.

Kass, G. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, *29*(2), 119.

Khodakarami, F. and Chan, Y.E., 2014. Exploring the role of customer relationship management (CRM) systems in customer knowledge creation. *Information & Management*, *51*(1), 27-42.

Knight, K. E. (1967) A Descriptive Model of Intra-Firm Innovation Process. *Journal of Management*, *41*, pp. 478-496.

Kuo, Y.H. and Kusiak, A., 2018. From data to big data in production research: the past and future trends. *International Journal of Production Research*, pp.1-26.

Lee, C., Song, B. and Park, Y., 2012. Design of convergent product concepts based on functionality: An association rule mining and decision tree approach. *Expert Systems with Applications*, 39(10), pp. 9534-9542.

Lee, J.C., Choy, K.L. and Leung, K.H., 2018. An intelligent fuzzy decision support system for flexible adjustment of dye pricing to manage customer-supplier relationship. In *IEEE International Conference* *on* *Smart Manufacturing, Industrial & Logistics Engineering* (SMILE) (pp. 7-11), IEEE.

Liao, S.H., Chu, P.H., Chen, Y.J. and Chang, C.C., 2012. Mining customer knowledge for exploring online group buying behavior. *Expert Systems with Applications*, 39(3), pp. 3708-3716.

Liao, S.H., Hsieh, C.L. and Huang, S.P., 2008. Mining product maps for new product development. *Expert Systems with Applications*, *34*(1), pp. 50-62.

Liu, Z. (2009). Association rules mining method from XML based on ontology. *Journal of Computer Applications*, *28*(9), pp. 2318-2320.

Loh, W. and Shih, Y. (1997). Split selection methods for classification trees. *Statistica Sinica*, *7*, pp. 815-840.

Maidique, M.A. and Zirger, B.J., 1985. The new product learning cycle. *Research Policy*, *14*(6), pp. 299-313.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H., 2011. *Big data: The next frontier for innovation, competition, and productivity*. San Francisco: McKinsey Global Institute.

McKay, M.D., Beckman, R.J. and Conover, W.J., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), pp. 239-245.

Miller, W.L., 2001. Innovation for business growth. *Research Technology Management*, *44*(5), pp. 26-31.

Mingers, J. (1989). An empirical comparison of selection measures for decision-tree induction. *Machine Learning*, *3*(4), pp. 319-342.

Moore, W.L., Louviere, J.J. and Verma, R., 1999. Using conjoint analysis to help design product platforms. *Journal of Product Innovation Management*, *16*(1), pp. 27-39.

Mourtzis, D., Vlachou, E. and Zogopoulos, V., 2018. Mobile apps for providing product-service systems and retrieving feedback throughout their lifecycle: a robotics use case. *International Journal of Product Lifecycle Management*, 11(2), pp.116-130.

Niosi, J. (1999). Fourth-generation R&D: From linear models to flexible innovation. *Journal of Business Research*, *45*(2), pp. 111-117.

Nonaka, I. and Takeuchi, H., 1995. *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Boston, Oxford University Press.

Olson, D.L., 2018. View of IJPR contributions to knowledge management in supply chains. *International Journal of Production Research*, 56(1-2), pp.733-742.

Pfeffer, J. and Sutton, R.I., 1999. *The knowing-doing gap: How smart companies turn knowledge into action*. Boston, MA: Harvard Business Press.

Piri, S., Delen, D., Liu, T. and Zolbanin, H.M., 2017. A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble. *Decision Support Systems*, *101*, pp. 12-27.

Polanyi, M., 1966. The logic of tacit inference. *Philosophy*, *41*(155), pp. 1-18.

Quinlan, J.R., 1979. Discovering rules by induction from large collections of examples. In D. Michie (Ed.), *Expert systems in the micro electronic age.* Edinburgh: Edinburgh University Press.

Quinlan, J.R., 1992. *C4.5 - programs for machine learning.* San Mateo, CA: Kaufmann.

Roberts, D.L., and Candi, M., 2014. Leveraging social network sites in new product development: Opportunity or hype? *Journal of Product Innovation Management*, *31*(S1), pp. 105-117.

Rothwell, R. and Soete, L., 1983. Technology and economic change. *Physics in Technology*, *14*(6), p. 270.

Rothwell, R., 1994. Towards the fifth-generation innovation process. *International Marketing Review*, 11(1), pp. 7-31.

Schuetz, J. (2015). Why are Walmart and Target next-door neighbors? *Regional Science and Urban Economics*, *54*, pp. 38-48.

Seo, W., Yoon, J., Park, H., Coh, B.Y., Lee, J.M. and Kwon, O.J., 2016. Product opportunity identification based on internal capabilities using text mining and association rule mining. *Technological Forecasting and Social Change*, *105*, pp. 94-104.

Şeref, M.M., Carrillo, J.E. and Yenipazarli, A., 2016. Multi-generation pricing and timing decisions in new product development*. International Journal of Production Research*, *54*(7), pp. 1919-1937.

Shahbaz, M., Srinivas, M., Harding, J.A. and Turner, M., 2006. Product design and manufacturing process improvement using association rules. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *220*(2), pp. 243-254.

Shang, S.S., Li, E.Y., Wu, Y.L. and Hou, O.C., 2011. Understanding Web 2.0 service models: A knowledge-creating perspective. *Information & Management*, *48*(4-5), pp. 178-184.

Shimomura, Y., Nemoto, Y., Ishii, T. and Nakamura, T., 2018. A method for identifying customer orientations and requirements for product–service systems design. *International Journal of Production Research*, *56*(7), pp. 2585-2595.

Steiner, W. and Hruschka, H., 2003. Genetic algorithms for product design: How well do they really work? *International Journal of Market Research*, *45*(2), pp. 229-243.

Su, C.T., Chen, Y.H. and Sha, D.Y., 2006. Linking innovative product development with customer knowledge: A data-mining approach. *Technovation*, *26*(7), pp. 784-795.

Tan, K.H. and Zhan, Y., 2017. Improving new product development using big data: A case study of an electronics company. *R&D Management*, *47*(4), pp. 570-582.

Tsai, C.Y., Chang, P.C. and Wang, S.J., 2003. Applying association-rule techniques and artificial neural networks to product development. *Journal of the Chinese Institute of Industrial Engineers*, *20*(2), pp. 101-112.

Utterback, J.M. and Abernathy, W.J., 1975. A dynamic model of process and product innovation. *Omega*, *3*(6), pp. 639-656.

Williamson, P.J. and Yin, E., 2014. Accelerated innovation: The new challenge from China. *MITSloan Management Review*, *55*(4), pp. 27-34.

Wong, D., 2012*. Data is the next frontier, analytics the new tool*. London: Big Innovation Centre.

Wynn, D.C. and Eckert, C.M., 2017. Perspectives on iteration in design and development. *Research in Engineering Design*, 28(2), pp.153-184.

Xu, Z., Frankwick, G.L. and Ramirez, E., 2016. Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*, *69*(5), pp. 1562-1566.

Yee Liau, B. and Pei Tan, P., 2014. Gaining customer knowledge in low cost airlines through text mining. *Industrial Management & Data Systems*, *114*(9), pp. 1344-1359.

Zhan, Y. and Tan, K.H., 2018. An analytic infrastructure for harvesting big data to enhance supply chain performance. *European Journal of Operational Research*. [doi.org/10.1016/j.ejor.2018.09.018](https://doi.org/10.1016/j.ejor.2018.09.018)

Zhan, Y., Tan, K.H., Ji, G., Chung, L. and Tseng, M., 2017. A big data framework for facilitating product innovation processes. *Business Process Management Journal*, *23*(3), pp. 518-536.

Zhan, Y., Tan, K.H., Li, Y. and Tse, Y.K., 2016. Unlocking the power of big data in new product development. *Annals of Operations Research*, pp. 1-19.

Zhang, C., Zhou, G., Lu, Q. and Chang, F., 2017. Graph-based knowledge reuse for supporting knowledge-driven decision-making in new product development. *International Journal of Production Research*, *55*(23), pp. 7187-7203.

Zhu, Q. and Sarkis, J., 2004. Relationships between operational practices and performance among early adopters of green supply chain management practices in Chinese manufacturing enterprises. *Journal of Operations Management*, *22*, pp. 265-289.

Zirger, B.J. and Maidique, M.A., 1990. A model of new product development: An empirical test. *Management Science*, *36*, pp. 867-883.





Type A Type B Type C

Figure 2. Types of smartphone presented in the questionnaire

|  |
| --- |
| Table 1: Evolution of the new product development process |
| Period | **Approach to NPD** | **Characteristics** | **Features** | **Examples** |
| First generation: the post-Second World War period to the mid-1960s  |  | The process is perceived as a linear progression from scientific discovery to marketplace. | * No customer connection
* High degree of uncertainty and serendipity
* Focus on R&D
* Linear progression
 | Flash of genius or brainstorming (Knight, 1967)Multi-divisional and sequential process (Utterback and Abernathy, 1975) |
| Second generation: the mid-1960s to the late 1970s  |  | The process is generally organized as multidisciplinary projects, in a linear sequential process, starting with market needs. | * Increased market focus
* Linear progression
* Starts with market needs
* Well organized to reduce uncertainty
 |
| Third generation: the late 1970s to the early 1990s  | Market | The process is essentially linear, with sequential progress; NPD is driven by market feedback and requirements. | * Resources-based knowledge management
* Logically sequential but not necessarily a continuous process
* Functional departments work interactively
* Gathers feedback from each department to improve
 | Stage-gate (Cooper, 1990, 1994)Open innovation (Chesbrough, 2006) |
| Fourth generation: the early 1990s to the present Open sources such as the market, customers, third parties, competitors |  | NPD involves simultaneous processing through a wide range of networks and partners to increase speed. | * Collective customer knowledge management
* Simultaneous and cross-functional processing
* Gathers feedback from customers and market quickly
* Data-driven decision making
 |

|  |
| --- |
| Table 2: Data mining to support customer knowledge management in NPD |
| **Literature** | **Knowledge types** | **Data mining techniques** | **NPD application** |
| Moore et al., 1999 | Consumer knowledge to product features/attributes | Conjoint analysis | Product platform development |
| Steiner and Hruschka, 2003 | Knowledge extraction about potential optima design | Conjoint analysis and genetic algorithms | New product ideal generation |
| Tsai et al., 2003 | Customer knowledge from different product combinations | Association rules and neural networks | Customer identification  |
| Agard and Kusiak, 2004 | Knowledge about product requirements | Association rules | Design of new product families |
| Shahbaz et al., 2006 | Knowledge of normal and abnormal operational patterns | Association rules | Manufacturing process improvement |
| Su et al., 2006 | Customer knowledge from different market segments | Clustering | Product development process |
| Jiao et al., 2007 | Knowledge about product and process variety | Association rules | Product process mapping |
| Liao et al., 2008 | Knowledge extraction of cosmetic products | Association rules | Product maps mining |
| Chen, 2009 | Knowledge based on intuitionistic fuzzy | Fuzzy decision trees | New product identification  |
| Bae and Kim, 2011 | Knowledge integration of customers’ needs  | Decision trees | Product process mapping |
| Liao et al., 2012 | Customer knowledge of online group buying behavior | Clustering and association | Customer online behavior |
| Jin et al., 2015 | Knowledge identification of new product concepts | Text mining  | New product mapping |
| Seo et al., 2016 | Knowledge management of firms’ internal capabilities | Text mining | Product internal capability identification |
| Zhang et al., 2017 | Knowledge reuse for product development decision making | Personalized PageRank algorithm | Knowledge  |
| Shimomura et al., 2018 | Knowledge extraction about customer orientations and requirements | Topic analysis and scenario approaches | Product-service systems design |

|  |
| --- |
| Table 3. Key features of a smartphone as determined in the pre-test |
| **Feature number** | **Features** | **Label** |
| 1 | Price | PR |
| 2 | Weight | WT |
| 3 | Size | SZ |
| 4 | Camera resolution | CR |
| 5 | Screen resolution | SR |
| 6 | Operating system | OS |
| 7 | Battery | BT |
| 8 | Color | CL |
| 9 | Material | MT |

|  |
| --- |
| Table 4. Characteristics of the respondents  |
| **Measures** | **Items** | **No. of respondents** | **Frequency (%)** |
| Gender | (1) Male | 92 | 55.0% |
|  | (2) Female | 75 | 44.9% |
| Age | (1) 20-30 | 93 | 55.7% |
|  | (2) 30-40 | 55 | 32.9% |
|  | (3) Over 40 | 19 | 11.4% |
| Marital status | (1) Single | 118 | 70.7% |
|  | (2) Married | 49 | 29.3% |
| Employment status | (1) Undergraduate | 31 | 18.6% |
|  | (2) Postgraduate/MBA | 106 | 63.5% |
|  | (3) Employee | 30 | 18.0% |
| Internet access per day | (1) < 1hr | 49 | 29.3% |
|  | (2) 1-3 hrs | 85 | 50.9% |
|  | (3) > 3 hrs | 33 | 19.8% |
| Monthly income | (1) < $1,000  | 60 | 35.9% |
|  | (2) $1,000-5,000 | 30 | 18.0% |
|  | (3) > $5,000 | 77 | 46.1% |
| Education level | (1) No high school diploma | 4 | 2.4% |
|  | (2) High school graduate | 13 | 7.8% |
|  | (3) University or college | 150 | 89.8% |

|  |
| --- |
| Table 5: Association rule dataset |
| **ID** | **Feature** | **Label** | **ID** | **Feature** | **Label** |
| 1 | Size | SZ | 5 | Weight | WT |
| 1 | Price | PR | 5 | Size | SZ |
| 1 | Operating system | OS | 5 | Screen resolution | SR |
| 1 | Type C | STC | 5 | Type A | STA |
| 2 | Size | SZ | … | … | … |
| 2 | Weight | WT |
| 2 | Operating system | OS |
| 2 | Type C | STC |
| 3 | Screen resolution | SR | 166 | Price | PR |
| 3 | Material | MT | 166 | Battery | BT |
| 3 | Operating system | OS | 166 | Material | MT |
| 3 | Type C | STC | 166 | Type B | STB |
| 4 | Camera resolution | CR | 167 | Price | PR |
| 4 | Material | MT | 167 | Screen resolution | SR |
| 4 | Operating system | OS | 167 | Material | MT |
| 4 | Type C | STC | 167 | Type C | STC |

|  |
| --- |
| Table 6: Decision tree dataset |
| **Feature** | **Label (Range)** | **1** | **2** | **3** | **…** | **165** | **166** | **167** |
| Price | PR (1-5) | PR1 | PR1 | PR1 | … | PR2 | PR2 | PR4 |
| Weight | WT (1-4) | WT3 | WT2 | WT2 | WT3 | WT2 | WT3 |
| Size | SZ (1-4) | SZ3 | SZ2 | SZ3 | SZ3 | SZ2 | SZ1 |
| Camera resolution | CR (1-4) | CR4 | CR1 | CR1 | CR1 | CR1 | CR1 |
| Screen resolution | SR (1-4) | SR3 | SR2 | SR2 | SR1 | SR2 | SR1 |
| Operating system | OS (1-3) | OS2 | OS1 | OS2 | OS2 | OS1 | OS1 |
| Battery | BT (1-5) | BT5 | BT1 | BT2 | BT1 | BT3 | BT3 |
| Color | CL (1-5) | CL5 | CL4 | CL1 | CL4 | CL1 | CL5 |
| Material | MT (1-3) | MT1 | MT2 | MT2 | MT1 | MT2 | MT2 |
| Type | TY (A, B, C) | STB | STC | STC | STB | STC | STC |
| Notes: *PR1 = $300 or less; PR2 = $300-500; PR3 = $500-800; PR4 = over $800; PR5 = Does not matter; WT1 = 110 g or less; WT2 = 110-150 g; WT3 = over 150 g; WT4 = Does not matter; SZ1 = 5.0 in or less; SZ2 = 5.1-5.5 in; SZ3 = over 5.5 in; SZ4 = Does not matter; CR1 = 15 million pixels or less; CR2 = 15-20 million pixels; CR3 = over 20 million pixels; CR4 = Does not matter; SR1 = 1920p and above; SR2 = 1280-1920p; SR3 = 1280p or less; SR4 = Does not matter; OS1 = IOS; OS2 = Android; OS3 = Does not matter; BT1 = 1,400 mAh or less; BT2 = 1,400-1,800 mAh; BT3 = 1,800-2,000 mAh; BT4 = 2,000 mAh and above; BT5 = Does not matter; CL1 = White; CL2 = Gold; CL3 = Silver; CL4 = Black; CL5 = Does not matter; MT1 = Metal; MT2 = Plastics; MT3 = Does not matter; TYA = Type A smartphone; TYB = Type B smartphone; TYC = Type C smartphone.* |

|  |
| --- |
| Table 7: Results of the association rules analysis (min sup = 5%, min con = 30%) |
| **Smartphone type** | **Association rule** | **Sup. (%)** | **Conf. (%)** | **Lift** | **Antecedent** | **Consequent** |
| *TYA* | RA1 | 16.77  | 41.59 | 1.29 | SR | TYA |
| RA2 | 13.77  | 39.13 | 1.23 | MT | TYA |
| RA3 | 6.59  | 39.02 | 1.21 | WT | TYA |
| RA4 | 5.99  | 35.16 | 1.08 | BT | TYA |
| RA5 | 5.39 | 30.03 | 1.04 | CL | TYA |
| RA6 | 11.98  | 47.18 | 1.47 | SR & MT | TYA |
| RA7 | 8.38  | 39.77 | 1.24 | SZ & SR | TYA |
| RA8 | 7.19  | 38.05 | 1.13 | SR & CR | TYA |
| *TYB* | RB1 | 11.98  | 36.16 | 1.25 | PR | TYB |
| RB2 | 10.18  | 34.17 | 1.17 | WT | TYB |
| RB3 | 6.59  | 32.04 | 1.09 | BT | TYB |
| *TYC* | RC1 | 23.95  | 52.84 | 1.12 | SZ | TYC |
| RC2 | 23.35  | 71.27 | 1.38 | OS | TYC |
| RC3 | 22.16  | 56.71 | 1.10 | CR | TYC |
| RC4 | 15.57  | 57.56 | 1.20 | PR | TYC |
| RC5 | 17.37  | 87.86 | 1.71 | OS & CR | TYC |
| RC6 | 13.77  | 58.07 | 1.12 | SZ & MT | TYC |
| RC7 | 12.57  | 71.60 | 1.37 | SZ & OS | TYC |
| RC8 | 11.38  | 60.19 | 1.15 | MT & CR | TYC |
| RC9 | 10.78  | 81.35 | 1.59 | OS & MT | TYC |
| RC10 | 8.98  | 51.23 | 1.04 | WT & SR | TYC |
| Notes: *Sup. = Support; Conf. = Confidence; TYA = Type A smartphone; TYB = Type B smartphone; TYC = Type C smartphone; SR = Screen resolution; MT = Material; WT = Weight; BT = Battery; CL = Color; SZ = Size; CR = Camera resolution; PR = Price; OS = Operating system.* |

|  |
| --- |
| Table 8: Results of the decision tree modeling |
| **Smartphone type** | **Decision tree rule** | **IF** | **THEN** | **No. of observations** | **Percentage (%)** |
| *TYA* | RA1 | Screen Resolution = SR1; Material = MT2; Operating System = OS2 | TYA | 33 | 78.8% |
| RA2 | Material = MT3 | TYA | 21 | 57.1% |
| RA3 | Operating System = OS3 | TYA | 9 | 44.4% |
| *TYB* | RB1 | Weight = WT4; Material = MT2 | TYB | 11 | 72.7% |
| *TYC* | RC1 | Material = MT1; Operating System = OS1 | TYC | 42 | 90.5% |
| RC2 | Screen Resolution = SR1; Material =MT1 | TYC | 25 | 76.0% |
| RC3 | Material = MT3 | TYC | 14 | 64.3% |
| Notes: *TYA = Type A smartphone; TYB = Type B smartphone; TYC = Type C smartphone; PR1 = $300 or less; PR2 = $300-500; PR3 = $500-800; PR4 = over $800; PR5 = Does not matter; WT1 = 110 g or less; WT2 = 110-150 g; WT3 = over 150 g; WT4 = Does not matter; SZ1 = 5.0 in or less; SZ2 = 5.1-5.5 in; SZ3 = over 5.5 in; SZ4 = Does not matter; CR1 = 15 million pixels or less; CR2 = 15-20 million pixels; CR3 = over 20 million pixels; CR4 = Does not matter; SR1 = 1920p and above; SR2 = 1280-1920p; SR3 = 1280p or less; SR4 = Does not matter; OS1 = IOS; OS2 = Android; OS3 = Does not matter; BT1 = 1,400 mAh or less; BT2 = 1,400-1,800 mAh; BT3 = 1,800-2,000 mAh; BT4 = 2,000 mAh and above; BT5 = Does not matter; CL1 = White; CL2 = Gold; CL3 = Silver; CL4 = Black; CL5 = Does not matter; MT1 = Metal; MT2 = Plastics; MT3 = Does not matter.* |

|  |
| --- |
| Table 9: Proposed approach for innovative product development |
|  | **Proposed approach** |
|  | Data collection | Data pre-processing and model development | Model assessment and decision making |
| **Knowledge management** | Knowledge from customers  | Knowledge about customers | Knowledge for customer |
| **Product development**  | Generation of concepts and ideas | Product design and manufacturing | Market testing and product launch |
| **Data-driven procedures** | Web-based survey and questionnaire | Identification of product features and design details via data mining techniques | Deploy differentiated product features in product development |