Technological Forecasting & Social Change 170 (2021) 120871

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

A hybrid recommendation model for successful R&D collaboration: Mixing machine learning and discriminant analysis

Seung-Pyo Jun [a,*], Hyoung Sun Yoo [b], Jeena Hwang [c]

[a] Data Analysis Platform Center, Korea Institute of Science and Technology Information and Science & Technology Management & Policy, University of Science & Technology (UST), 66, Hoegi-ro, Dongdaemun-gu, Seoul, 130-741, Korea
[b] Korea Institute of Science and Technology Information and Science & Technology Management & Policy, University of Science & Technology (UST), 66, Hoegi-ro, Dongdaemun-gu, Seoul, 130-741, Korea
[c] Korea Institute of Science and Technology Information, 66, Hoegi-ro, Dongdaemun-gu, Seoul, 130-741, Korea

A R T I C L E I N F O

Keywords:
R&D collaboration
Recommendation model
Discriminant analysis
Decision tree analysis
Machine learning
Small and medium enterprises

A B S T R A C T

Seeking to stimulate and improve the rate of success of R&D collaboration by SMEs, this study developed a method of recommending types of external collaboration organizations that are optimal partners for SMEs. We began by examining the current data on R&D collaboration by partner type to effectively classify the types of R&D partners engaged with South Korean SMEs. Next, we applied machine learning and discriminant analysis to develop a hybrid model for recommending firms that will likely achieve high satisfaction from collaboration with four representative types of R&D partners (universities, public research institutes, large firms, and SMEs). Lastly, we used new data that had not been included in the model development stage, to perform additional evaluations of the model. In our research results, the hybrid recommendation model, designed to identify SMEs that will achieve high satisfaction by R&D partner type, demonstrated outstanding accuracy exceeding 91%. By applying the model proposed in this paper, firms will be able to select their R&D partner types more efficiently and improve the likelihood of achieving success in R&D collaboration. Meanwhile, those responsible for implementing public policies may use the proposed model to improve the efficiency of public investments that support R&D collaboration.

1. Introduction

R&D (research and development) or technological collaboration is based on inter-organizational collaboration, which refers to the sharing of resources, information and ideas between independent organizations with the expectation of mutual benefits (Hausman et al., 2002). Some scholars have explained that such a collaboration is undertaken by firms voluntarily because an increasing number of firms have become aware that such an inter-organizational collaboration is essential to achieve competitive superiority amid intensifying competition in the market (Lemmens, 2004).

Among the various instances of inter-organizational collaboration, technology collaboration is a type of collaboration specifically developed for technical, economic and people-related purposes (Brockhoff and Teichert, 1995). Small and medium-sized enterprises (SMEs), which are the main target of this study, have also attempted to engage in technical collaboration to maximize profits or supplement their insufficient capabilities (Calcagnini et al., 2019; Galende, 2006; Hottenrott and Lopes-Bento, H. 2016). Continuous research has also been conducted on these SMEs' goals and motives, which have a significant influence on the outcomes or determinants of such technological collaboration. Transaction Cost Economics (TCE), first developed by Williamson (1981), is a key theory that is frequently applied when researching alliances between firms or other economic agents. TCE explains that the motivation for R&D alliances lies in the desire of firms to choose an inter-firm alliance that minimizes transaction costs and achieves cost efficiency (Williamson, 1991). On the other hand, The Resource-based View (RBV) is founded on the idea that the purpose of inter-firm alliances is to harness the potential of generating value by combining the resources of firms (Das and Teng, 2000). Whereas TCE, informed by organizational economics, focused on transaction cost minimization driven by a competitive environment, RBV focuses on the utilization of various forms of resources belonging to a firm (Das and Teng, 2000). Of course, the motives and purposes driving such technical

* Corresponding author.
E-mail addresses: spjun@kisti.re.kr, spjunn@msn.com (S.-P. Jun), hsyoo@kisti.re.kr (H.S. Yoo), jeena@kisti.re.kr (J. Hwang).

https://doi.org/10.1016/j.techfore.2021.120871
Received 2 September 2020; Received in revised form 3 May 2021; Accepted 5 May 2021
Available online 21 May 2021

collaboration will also affect the selection of partners (Dong and Glaister, 2006; Nielsen, 2003).

Due to the recent advent of the 4th Industrial Revolution and the accelerating development of information communication technology, the scope of inter-organizational R&D collaboration has expanded beyond the conventional dyadic relationship to broader networks involving multiple interested parties (Schwab, 2017). Even in activities such as new product development, broader R&D collaboration—encompassing various parties including customers, buyer companies, competitor companies, and firms in complementary relation to one another—has been fundamentally transforming the approaches to generating new value (Fritsch and Lukas, 2001; Prahalad and Ramaswamy, 2000; Souder et al., 1997). This expansion of R&D collaboration led to the emergence of more diverse external sources (or types) that help drive R&D collaboration. Specifically, these types may now include supplier companies, buyer companies, competitor companies, and customers as well as public research institutes, private research institutes, universities, and professional consultants (Chatterji, 1996; Davenport and Miller, 2000). Such diversification of partner selections has generated research not only regarding inter-organizational technological collaboration with market, science and supplier partners, but also regarding functional diversity, that is, on portfolios of collaboration that involve a mix of partners – including competitors, private research centers and consultants, and institutional partners (public research centers and universities) (Sarpong and Teirlinck, 2018).

This study adopts the idea of open innovation and the framework of innovation system theory and is theoretically grounded on the view that R&D collaboration (joint research) is necessary to redress failures in the innovation system and strengthen the innovation capabilities (such as an absorptive capacity and an organization's ability to transform knowledge into new products) of SMEs (Chesbrough, 2006; Konsti-Laakso et al., 2012; Malerba, 2002). Collaboration in R&D innovation activities is thus regarded as an effective strategy for technology transfer that enables SMEs to overcome their limitations (Xu et al., 2019), and there have been various studies dedicated to improving the success of R&D collaboration or joint projects conducted by SMEs (Leischnig and Geigenmüller, 2020). However, SMEs' resource constraints and their inadequate ability to prevent knowledge leakage render it difficult for SMEs to select partners for collaboration (Sarpong and Teirlinck, 2018). SMEs still experience difficulties arising from issues related to partners in R&D collaboration, and there remains a need for research to redress this situation. For example, in a survey of South Korean SMEs regarding the difficulties they experience in R&D collaboration, 32.7% of the respondents cited difficulties related to partners (lack of suitable partners, lack of relevant information, communication problems with partners), which ranked greater than other challenges such as lack of funds (32.4%), prolonged research periods (23.9%), and information leakage and intellectual property disputes (11.0%) (MSS and Kbiz, 2018). R&D collaboration increases the complexity of R&D coordination, which may have a negative effect on innovation performance (Hagedoorn et al., 2018). This study aims to present a data analysis methodology using machine learning based decision tree analysis and discriminant analysis for effectively identifying optimal R&D collaboration types to recommend to SMEs. In this study, the phrase "optimal partner types" (or "optimal alternatives") refers to the types of partners that are anticipated to yield maximum satisfaction to SMEs that collaborate with them, determined by comparisons to all the other types of partners available for R&D collaboration (Leischnig and Geigenmüller, 2020; Sarpong and Teirlinck, 2018).

In this paper, we propose a hybrid model that uses a mixture of two methods to develop a recommendation model based on satisfaction prediction. We use a machine learning model (ensemble decision tree analysis) with relatively excellent predictive power to find the optimal determinant and apply discriminant analysis with excellent interpretation power to construct a final recommendation model. It should be noted that when using the ensemble model alone, the expected predictive power may be excellent, but there would also be a clear limit in that the effect of the independent variables would not be explainable (James et al., 2013); it is for the purpose of overcoming this limit that we used discriminant analysis in parallel. To demonstrate that this hybrid model also has an acceptable level of predictive power, we verified the predictive power of our model using the survey results for the last three years, a dataset that we had not used during model development.

Prior to developing the model for recommending optimal R&D collaboration types to SMEs, we reviewed the existing literature related to the development of recommendation systems (models) and partner recommendations in Section 2. Next, from among the various theoretical approaches that can be used to identify the target variable and determinants for the R&D partner type recommendation model, and we also present the results of our longitudinal analysis of the types of R&D collaboration partners adopted by small and medium South Korean firms in Section 2. Through this process, we identified the four types of partners that SMEs consider to be potential R&D collaboration partners: universities, public research institutes, large firms, and SMEs. Section 3 explains our research process and methodology and the data used in our analysis. In Section 4, we use machine learning and discriminant analysis to derive a model for generating the recommendations of the optimal types of R&D partners. Finally, in Section 5, we discuss the recommendation model developed for the four types: we begin with an evaluation of the predictive capabilities of the model, discuss the potentials for practical utilization of this model and then analyze the determinant factors identified by the recommendation model from a scholarly perspective.

2. Analysis of preceding studies and current findings

This study was directed by our interest in classifying the types of partners available for R&D collaboration with SMEs and developing a method of recommending such partners. In the following, we begin by examining the existing literature on recommendation systems (or models), since that is the main object of this paper. Next, to explain the basis for our selection of the determinants and target variables used in our R&D collaboration partner recommendation model for SMEs, we shall review the preceding studies focusing on the issues of 1) SMEs' motives for R&D collaboration and partner selection, 2) the measurement of the outcomes of SMEs' R&D collaboration, and 3) the determinants of success in R&D collaboration by SMEs. In this section, we also help develop a deeper understanding of the current conditions of R&D collaborations among South Korea's SMEs by longitudinally analyzing changes in annual utilization rates and differences in satisfaction levels depending on the types of R&D collaboration partners chosen by the SMEs. Through this process, we derived the operational definition of R&D collaboration partner types for this study.

2.1. Preceding studies related to data-based recommendation model research

With the increase in the quantity and accessibility of data in digital format, there has been expanded research into methods of automatically filtering for high quality data. Studies of customized information filtering (IF) devised for this purpose generally consider factors such as the user's interests, familiarity, novelty, importance, or urgency to predict which information will be of high priority for the user. According to Belkin and Croft (1992), research on information filtering systems differ from research on conventional information retrieval (IR) systems in the following ways. Whereas the user profiles in information filtering systems generally indicate long-term interests, user queries in IR systems tend to reflect short-term interests. Furthermore, whereas user queries in IR systems apply the method of finding information with the strongest similarity, IF systems proceed by eliminating unnecessary information. Such information filtering systems have evolved into recommendation systems, which implement filtering methods that can be broadly classified into two approaches (Park and Kim, 2017). The first is the content-based filtering approach, which recommends only items that have a high degree of relevancy based on user profiles. Leading examples include NewsWeeder, Infofinder, and NewsDude. The key to this approach is the level of expression of the data item and the technological level of individualized preferences in the user profile. The second approach is referred to as collaborative filtering: instead of calculating the similarity between a data item and user profiles, this approach utilizes similarity among user profiles. Individuals with similar user profiles are grouped together and information is shared among them. The main purpose of this approach is to generate recommendations among users within the same group. Representative examples include Ringo and Siteseer. Since the objective of this study is to share information in the form of recommendations for SMEs that are users within the same group, it can be classified as adopting a collaborative filtering approach.

Recommendation systems refer generally to computer-based systems that function as personalization tools by matching users with items for potential purchase or customized information, based on users' preferences. Such systems can provide further useful information, such as information on job openings or investment opportunities, which had been previously inaccessible due to asymmetry of information. In the R&D domain, there have been multiple attempts to research the utilization of recommendation systems. Xu et al. (2016), for example, researched a two-stage recommendation approach that includes the stage of identifying a set of suitable R&D project candidates, followed by the stage of actually recommending suitable R&D projects from that set for project applicants. Jeong et al. (2016) designed a recommendation system that evaluates committees for national R&D projects. Tang et al. (2012) researched a Cross-domain Topic Learning (CTL) model that aims to solve the difficulties that impeded researchers from engaging in interdisciplinary research collaboration. This model generated recommendations by considering the unique patterns formed among domains classified from large quantities of research publications. Notably, Wang et al. (2017) built a system that recommends researchers who have background knowledge and contextual awareness relevant for a given domain, to optimize R&D project collaboration between universities and industry. Jun et al. (2017) applied machine learning to identify and compare the profiles of firms that benefited from government support programs in their R&D planning stage, and firms that were in demand of such support. Their results demonstrated that there were differences between the profiles of beneficiary firms and in-demand firms, and they also argued that it would be possible to use the profiles of in-demand firms to select or to recommend the firms that are best suited to be a beneficiary of the governmental R&D support program.

The recommendation model that this study aims to develop is a model to be applied to a partner recommendation system that will improve the likelihood of successful R&D collaboration. One example of the available research on recommendation models and more specifically on partner recommendation is the study by Wu et al. (2013), which proposed an online service-based model for recommending patent partners to researchers working in firms. Their study adopted a 3-stage approach. It began by generating the candidate recommendation targets, which means identifying the set of potential collaborators, then identified the factors that adjust the candidate ranking. Based on these factors, the study proposed a method of sub-dividing the candidates. Lastly, Wu et al. (2013) proposed an interactive learning method that makes it possible to incorporate the user's feedback in the recommendation model. To develop this recommendation model, Wu used content similarity, collaborative filtering, and the SVM-Rank (support vector machine for ranking) method. Lu et al. (2013) also offered a web-based recommendation model that recommends business partners for SMEs. The study presents a hybrid fuzzy semantic recommendation (HFSR) approach which combines item-based fuzzy semantic similarity and item-based fuzzy collaborative filtering (CF) similarity techniques. Meanwhile, Yuan et al. (2015) developed a model for recommending alliance partners for firms, based on customer experience. This study utilized alliance partner recommendations to enable firms to deliver their aimed customer experiences; here, customer experiences were represented by images that can be analyzed and created based on customers' feedback and their interactions with companies. Even within this active field of recommendation systems research, however, there has been a relative lag in attempts to develop a model for recommending R&D collaboration partners, as our present study endeavors to do.

As in preceding studies, this study provides a partner recommendation model for firms or governments; what differentiates it is that it offers partner type recommendations for R&D collaboration. There is some continuity with preceding studies, however, in that we constitute a pool of partners (types) and derive the factors for determining their rankings. In this study, the determinant factors (independent variables) are established based on the preceding studies reviewed in the following section, but the target variables (dependent variables) are based on feedback on the global satisfaction level with R&D collaboration experiences, similar to the study by Yuan et al. (2015), and our recommendation model is developed based on collaborative filtering, thus building on and further enhancing the approach seen in preceding studies.

2.2. Preceding studies related to SMEs' R&D collaboration

2.2.1. SMEs' motivations for R&D collaboration and partner selection

Gulati and Singh (1998) defined strategic alliances as "voluntary arrangements between firms involving exchange, sharing or co-development of products, technologies, or services" and Hagedoorn and Schakenraad (1994) defined a technology alliance as the activity of inter-firm alliance for the purpose of achieving innovation for product development. Strategic alliance refers to a broader concept of collaborative activity compared to technology alliance; in other words, technological alliance could be considered to be subsumed under the concept of strategic alliance. Das and Teng (2000) distinguished the stages of alliances as follows: motive for alliance, alliance partner selection, alliance governance, and alliance performance. Kale and Singh (2009) divided inter-firm alliances into three stages. The first stage is the stage of formation and partner selection, the second stage is the selection of alliance governance and design, and the last stage is post-formation alliance management. Alliance performance refers to the evaluation of the outcome of the alliance, assessed in the post-alliance stage.

The purpose of this study is to propose a recommendation model for the partner selection step, which is the first step in inter-firm alliances, including R&D collaboration, as described above. Therefore, it is necessary to gain a clearer understanding of the stage of R&D collaboration (alliance), and as explained above, the motives and purposes of collaboration are important. In regards to the motives driving firms to form a technology alliance, from the TCE perspective, we can identify the desire to reduce R&D costs as a motive (Hagedoorn and Van Kranenburg, 2003) and from the RBV perspective, we understand that there is the motive to combine complementary technology with other firms in major fields of R&D to acquire external technological knowledge (Hagedoorn, 2006). Nielsen (2003) researched how the motive of firms in Denmark forming international strategic alliances affected their selection of partner and concluded that the motive for the strategic alliance did indeed affect the selection criteria for the partner. Dong and Glaister (2006) analyzed Chinese firms to find out how the motive behind a strategic alliance influenced the firm's partner selection criteria. Their research results indicated that the motive of the strategic alliance affected the partner selection criteria and especially had a greater effect on "task-related criteria." Based on these findings, we can conclude that the motive of technology alliance (from the perspective of TCE and RBV) affects the selection criteria of the technology alliance partner.

Gulati and Singh (1998) argued that the selection of the alliance governance structure is affected by coordination costs and appropriation concerns in relation to the partner. When a firm finds it difficult to trust the partner firm, and there are concerns of appropriation, the firm will likely select a governance structure with a high degree of control, such as EJV (equity joint venture). According to Das and Teng (2000), the preference for a specific governance structure will be determined by the manner in which the respective resources of the main firm and the partner firm are combined: Das and Teng (2000) argued that if the partner firm's combined assets consist of more valuable knowledge-based resources while the main firm's combined assets are asset-based resources, then the main firm will prefer EJV. These studies indicate that both TCE and RBV can help explain the governance structure of technology alliances.

Lai and Chang (2010) studied the Taiwanese machinery industry and analyzed how the motive of a technology alliance affected technology alliance performance. They found that the motive behind a technology alliance had a positive effect on that alliance's performance. Although there may be slight differences depending on the industry sector of each high-tech firm (Hagedoorn, 1993)—for example, the difference between capital intensive and knowledge intensive sectors—generally, a high-tech firm's motive for technology alliance can be expected to affect the performance of the alliance.

Kim and Kim (2013) also found that both TCE and RBV could affect the governance structure of a technological alliance. They found that the motivation of technology alliances in Korean high-tech firms had an effect on the criteria, governance, and performance of technology alliance partners in terms of TCE and RBV.

Based on the research findings discussed above, we can see, from the TCE perspective, that the transaction costs may vary for each type of partner that engages in R&D collaboration with SMEs. For example, the transaction cost will be very different when we compare the case of R&D collaboration with SMEs that are able to achieve commercialization, and therefore may become a competitor, against the case of R&D collaboration with universities or public research institutes, since it is almost impossible for such organizations to become a competitor to the firm. In this study, therefore, we consider a wide range of variables to generate recommendations of partners or types, including asset specificity, uncertainty, and information. The selection of a partner type is also very important from the RBV perspective, since each type of partner for R&D collaboration will respectively have significantly differing financial, technological and marketing capabilities. Also, task-related criteria will also inevitably affect the selection of the types of collaboration partners. From the theoretical perspective of TCE and RBV (the motive of R&D collaboration), this study classifies the types of partners or alternatives for R&D collaboration into ① government-funded research institutes (herein after referred to as public research institutes) ② universities, ③ large firms, ④ SMEs, and ⑤ other institutions (private research institutes, etc.) (Chatterji, 1996; Davenport and Miller, 2000; MSS and Kbiz, 2018; SMBA and Kbiz, 2015).

2.2.2. Measurement of the outcomes of R&D collaboration by SMEs

As described above, the results of collaboration may be measured differently depending on the goals of R&D collaboration: in cases driven by technological aims, know-how and patent acquisition through technology development, as well as the discovery of technological opportunities, can be counted as achieved outcomes (Brockhoff and Teichert, 1995; Caloghirou et al., 2021; Dyer et al., 2007; Eom and Lee, 2010; Jun et al., 2020; Zacharias et al., 2020). In cases with economic aims, outcomes can be measured based on various performances including increased sales from the development of new products as well as reduction in R&D costs, shortened time to market entry, improved R&D success rates, and entry into new markets (Arranz and de Arroyabe, 2008; Brockhoff and Teichert, 1995; Fernández-Olmos and Ramírez-Alesón, 2017; Greco et al., 2020; Hottenrott and Lopes-Bento, H. 2016; Jun et al., 2020; Sarpong and Teirlinck, 2018; Zacharias et al., 2020). In other cases, such as those with people-related aims, even indices such as expansion of networking, strengthening of R&D capabilities through learning, and building trust can be regarded as outcomes (Brockhoff and Teichert, 1995; Caloffi et al., 2018). Of course, attempts have also been made to adopt approaches that measure the outcomes for various different purposes simultaneously (Jun et al., 2020; Park et al., 2020).

However, not all innovative achievements result in patents (Fukugawa, 2016), and patents do not necessarily lead to new products or the commercial success of new products (Hottenrott and Lopes-Bento, H. 2016). To overcome these shortcomings of measurements based on patents, the results of SMEs' R&D collaboration are often more generally measured based on production or service innovation, with a focus on the ratio of sales from products that are new to the market over total sales (Fernández-Olmos and Ramírez-Alesón, 2017; Hagedoorn et al., 2018; Jun et al., 2020). However, concerns have continuously been raised regarding the limitation that it is difficult to assess outcomes based solely on quantitative performance indicators such as economic success, let alone based only on patents, considering the wide variety of goals driving R&D collaboration. As an alternative, the success of R&D collaboration has also been measured by qualitative success or global satisfaction (Dyer et al., 2007; Mora-Valentin et al., 2004).

The model for R&D collaboration partner recommendations for SMEs developed in this study did not presuppose or exclude specific motives for collaboration. Furthermore, the selection of SMEs' collaboration partners, including R&D partners, should be approached from the perspective of diversity and portfolio composition as well as based on the outcomes of collaboration (Hagedoorn et al., 2018; Sarpong and Teirlinck, 2018). Therefore, we chose the overall satisfaction level in R&D cooperation as the target (or independent) variable of our model, encompassing various indicators rather than specifying the outcome based on an economic perspective (or goal), such as production innovation. One particular reason for selecting satisfaction was that the recommendation model based on collaborative filtering operated well based on feedback (satisfaction) regarding the experience of utilization (Yuan et al., 2015). This point is one important aspect of the recommendation model presented in this study, which distinguishes it from previous studies.

2.2.3. Determinants of the success of R&D collaboration by SMEs

When considering the determinants of the success of R&D collaboration, we broadly categorized the factors into the characteristics of R&D, technology and firms, respectively. First, *R&D characteristics* are characteristics that appear in the process of performing R&D, and in general, investment, procurement, and collaboration experience are often considered as determinants. The size of R&D investment (or intensity) and the types of organizations with which collaboration was experienced are key determinants to be considered (Caloffi et al., 2018; Gkypali et al., 2017; Hottenrott and Lopes-Bento, H. 2016; Kafouros et al., 2020; Markovic et al., 2020; Sarpong and Teirlinck, 2018). The intensity of R&D collaboration is particularly important: since we need to take account of the complexity of R&D collaboration, the proportion of internal or external collaboration in R&D is also an important factor influencing the performance outcomes of collaboration (Caloghirou et al., 2021; Hottenrott and Lopes-Bento, H. 2016; Sarpong and Teirlinck, 2018). In addition, the necessity and effects of public funding for R&D collaboration by SMEs has also been a subject of active research (Caloffi et al., 2018; Jun et al., 2020; Yoo et al., 2018): in considering R&D characteristics, the source of R&D financing and the portfolio of financing are also key concerns (Arranz and de Arroyabe, 2008; Greco et al., 2020; Sarpong and Teirlinck, 2018). Therefore, for R&D characteristics, this study considered the scale of R&D costs, the R&D cost utilization, the technology development implementation ratio (R&D collaboration intensity), R&D cost procurement, and experiences categorized by the type of collaboration R&D partner (see Table 1).

Next, *technology characteristics* can be divided into the technological features of the core technologies themselves and the technological capabilities of firms. First, the technological capability of a firm is often analyzed based on the level of intellectual property rights held by the firm (Bellucci et al., 2019; Kafouros et al., 2020) and has also been

Table 1
Variable Names and Explanations.

Variable name			Variable explanation	Related literature
Dependent variable (Group variable)		High satisfaction group resulting from collaboration R&D	Whether they are highly satisfied with the joint or cooperative research partner types (more than 4 out of 5)	2015; Yuan et al., 2015
Independent/Control variable				
R&D characteristics		R&D costs scale	Total amount invested in technology development in 2014	Caloffi et al., 2018; Hottenrott and Lopes-Bento, H. 2016; Sarpong and Teirlinck, 2018
		Technology development implementation ratio	Ratio of specific development methods (internal research, joint research, commissioned research, technology adoption, etc.) among R&D investments	
		R&D cost utilization	Investment of costs in specific areas of R&D (internal research, joint research, commissioned research, etc.)	
		R&D cost procurement	Procurement from specific sources of R&D investment (government sponsored, government loans, private financing, private investments, etc.)	
		Experience of technological cooperation	Experience engaging in technological cooperation with external institutions	
Technology characteristics	Technological capabilities	Intellectual property rights in ownership	Intellectual property rights in ownership: domestic/foreign, patent/design, applications/ registrations	Bellucci et al., 2019; Gkypali et al., 2017; Greco et al., 2020; Kafouros et al., 2020; Zacharias et al., 2020; Yoo et al., 2018
		Number of R&D project performances	Number of R&D project performances (attempts, in progress, success, commercialization, etc.)	
		Technology capabilities (global comparison)	Competency compared globally to countries advanced in specific technology development fields (Product planning, new technology development, testing inspection, manufacturing and processing, design, commercialization, parts and process design, maintenance and repair, etc.)	
		Equipment availability	The level of testing and experimental equipment required for R&D currently available at the firm	
		Number of times test equipment was used	Number of times test equipment was used: professional testing institutions, universities, public research institutes, large firms, SMEs, etc.	
		Level of technology compared to previous year	Level of the firm's technology compared to 1 or 3 years ago	
		Technology development performance outcomes	Technology development performance outcomes: No. of intellectual property rights registered, no. of intellectual property rights applied, etc.	
	Technological features	Core technology life cycle	Technology life cycle of the core technology	Arranz and de Arroyabe, 2008; Hagedoorn et al., 2018; Yoo et al., 2018
		Product life cycle	Product life cycle of the main product	
		Technological innovativeness	Level of innovation demonstrated in the technology	
		Time required for imitation	The time required for an external party to imitate the main technology	
		Time required for each stage of R&D	Time required: development planning, development progression (or implementation), commercialization and total technology development	
		Time required: opening new sales markets	Time to develop a market after completion of development	
Firm characteristics		Sales	Total sales in 2014	Chapman et al., 2018; Fernández-Olmos and Ramírez-Alesón, 2017; Un and Asakawa, 2015
		Main product sales ratio	Ratio of main (flagship) products in sales	
		New technology sales ratio	Ratio of new technology outcomes in sales in 2014	
		Operating/ordinary profit	Operating/ordinary profit in 2014	
		Transaction client sales ratio	Ratio of specific client groups (large firms, SMEs, Public, Consumers, exports) in sales	
		Year of establishment	Year of the company founding	
		Age of CEO / representative	Birth year of CEO / representative	
		Production based areas	8 areas of production-based technologies	
		Total number of employees	Total number of full-time employees	
		Total number of researchers	Total number of research personnel	
		Innovation SMEs status	Status of registration as Innovation SMEs companies	
		Venture status	Venture company registration status	
		Growth stage	Company growth stage (market entry phase, growth phase, maturation phase, declining phase)	

analyzed in terms of the absorptive or adaptive capacity of the firm in relation to technology (Gkypali et al., 2017; Greco et al., 2020; Zacharias et al., 2020), as well as in terms of equipment availability or the level of technology relative to the global level (Yoo et al., 2018). Likewise, in this study, to take account of the overall capability for technology development, we considered the patents acquired, technology development performance outcomes, the number of R&D projects and their performances. To incorporate specific capabilities, including absorptive or adaptive capabilities, we considered the technology capabilities (global comparison) as presented in Table 1.

When analyzing technological features, the technology or product characteristics of the core technologies may be considered, which includes technological characteristics such as the level of technological innovativeness (Arranz and de Arroyabe, 2008; Yoo et al., 2018), the life cycle of a business project or product (Fernández-Olmos and Ramírez-Alesón, 2017; Yoo et al., 2018), the speed of technological change, and the time required for imitation (Hagedoorn et al., 2018; Yoo et al., 2018). This study likewise considers the product life cycle as well as the life cycle of the core technology as reflective of the rate of change of a technology, and also considers technological innovativeness and the time required for imitation. We also adopted as potential determinants the time required for each stage of R&D (planning, implementing, commercialization) and the time required for opening new sales markets, which Yoo et al. (2018) identified as important determinants affecting R&D collaboration attempts by SMEs (see Table 1).

Firm characteristics are determinants included in nearly all studies on SMEs' R&D collaboration, and leading examples include the year a firm was established (or the firm's age) as well as sales figures or information on employees, which are relevant to the scale of the firm (Chapman et al., 2018; Fernández-Olmos and Ramírez-Alesón, 2017; Hottenrott and Lopes-Bento, H. 2016). Information on product portfolios such as the proportion of new product sales or industry classification may also be included (Caloffi et al., 2018; Hottenrott and Lopes-Bento, H. 2016), and some studies have considered the share of sales by customer (Un and Asakawa, 2015; Yoo et al., 2018). In our study, exports were excluded from consideration since foreign countries were already reflected in the proportion of customer sales, and while we had already included the year of establishment and the number of employees, to further consider absorptive capacity (Kafouros et al., 2020), we added consideration of the corporate growth stage, the level of corporate innovation (registration as an innovative company or venture company), and the number of researchers (Dyer et al., 2007). Some studies have considered credit rating as an indicator of the profitability or stability of a firm (Hottenrott and Lopes-Bento, H. 2016), but in this study, we instead included operating /ordinary profits (Bellucci et al., 2019; Gkypali et al., 2017).

The Table 1 presents the type of variables that were used as determinant (input) variables in the partner recommendation model for SME R&D collaboration and the related literature.[1]

2.3. Current R&D collaboration by SMEs

SMEs in South Korea occupy a highly important position in the country's industrial structure, as indicated by the fact that 99.9% of private companies are SMEs. They are also important in terms of employment, since 87.9% of workers are employed by SMEs (SMBA, 2020).

Fig. 1 presents data on the R&D collaboration conducted by SMEs in South Korea during a recent decade. Building on the existing research reviewed above (Chatterji, 1996; Davenport and Miller, 2000; MSS and Kbiz, 2018; SMBA and Kbiz, 2015), this study classifies the types of partners or alternatives for R&D collaboration (joint research or commissioned research) into 5 types. Fig. 1 includes data on joint research and R&D collaboration undertaken by SMEs over a period of ten years, from 2007/2008 (2009 survey) to 2017 (2018 survey).[2]

First, the data shows that the ratio of SMEs that have experience in R&D collaboration through joint research has been on the decline, for all types of collaboration partners. Some may attribute this phenomenon to the increase in the number of SMEs but considering that during the same period (2007–2017), the growth rate in the number of SMEs was only 2.31% (SMBA, 2020), the more plausible explanation is that these firms failed to expand their R&D collaboration activities. Although the South Korean government's R&D collaboration budget and projects for SMEs increased from 2013 to 2015 (Kim and Yang, 2017), the R&D collaboration of SMEs failed to expand. This result can be explained by the expansion of companies which preferred internal R&D (Yoo et al., 2018) or the phenomenon that government support was concentrated on specific companies with excellent capabilities (Caloffi et al., 2018; Jun et al., 2017). Comparing the individual categories of partner types, we see that among the institutions available for R&D collaboration in South Korea, universities were most often selected by SMEs, and public research institutes and SMEs were the next highest in percentage. The data confirms the high percentage comprised by universities and public research institutes in instances of R&D collaboration in South Korea. Meanwhile, the data on the satisfaction level regarding joint research showed somewhat different trends in Fig. 1: although the ratio of firms with joint research experience declined overall, the satisfaction level has been incrementally improving. There are also notable differences when we compare the types of partners: public research institutes significantly outranked universities by generating the highest level of satisfaction regarding R&D collaboration from SMEs (excluding 2017), though universities followed closely behind, earning consistently high satisfaction for R&D collaboration. The number of collaborations with other SMEs were similar in percentage compared to collaboration with public research institutes, but the partnered SMEs ranked the lowest in terms of satisfaction, lagging even behind large firms. Based on these results, it must be concluded that from the point of view of the South Korean government, R&D collaboration with universities and public research institutes must be valued as critically important for SMEs. Not only do universities and public research institutes contribute most in terms of how often they are selected as partners in R&D collaboration, they also earn the highest rankings in satisfaction. Private research institutions, which could be potential substitutes in R&D collaboration for universities or public research institutes, have failed to show significantly improved outcomes in terms of usage rate.

Based on these existing research results, the partner that is most suitable for SMEs seeking R&D collaboration will vary depending on the transactions cost and the types of resources that are needed. Some SMEs may have chosen a specific partner due to consideration of cost or due to insufficient capabilities, even though they wished to engage a different partner if possible. Also, insufficient access to information may lead firms to choose a partner that is not an optimal fit for R&D collaboration, and this will lead to low satisfaction outcomes. Therefore, for policymakers, it is important to not only expand the scope of R&D collaboration but to assist beneficiary firms in making effective selections of partners. In other words, there is a need for a method of recommending the optimal types of R&D collaboration partners to benefit SMEs.

Out of the five types (or groups) of R&D collaboration partners for SMEs we listed in Fig. 1, we excluded private research institutes because these had a relatively low frequency of collaboration. We then proceeded to use data on satisfaction levels experienced by SMEs from the remaining four types of partners, to recommend the optimal type of R&D collaboration partner. This classification of the types of R&D partners of SMEs can be linked to numerous preceding studies related to technology transfers. Many researchers have already shown keen interest in examining not only inter-firm R&D collaboration but also the characteristics of R&D collaboration distinguished by type, such as universities or public research institutes (or as public research organizations) (Arranz and de Arroyabe, 2008; Caloffi et al., 2018; Caloghirou et al., 2021; Jun et al., 2020; Sarpong and Teirlinck, 2018; Un and Asakawa, 2015).

3. Method and scope of research

3.1. Process and methodologies of research

We propose a model for recommending the type of R&D collaboration partner optimal for SMEs. Based on previous research, 4 types were selected, global (overall) satisfaction regarding R&D collaboration was chosen as our target variable, and the 82 variables presented in Appendix Table A were considered as the determining factors. In this study, in order to consider a relatively large number of determinants while also considering nonlinear relationships (Caloghirou et al., 2021; Hagedoorn et al., 2018), we used a machine learning-based ensemble model, while

[1] The variable names presented in Table 1 are for the purpose of introducing and classifying the variables; more specific variable names are presented in Appendix Table A.

[2] The survey, which had previously been conducted every two years, began to be conducted annually beginning in 2013; this is the reason the data results presented in Figure 1 have variations in the time intervals.

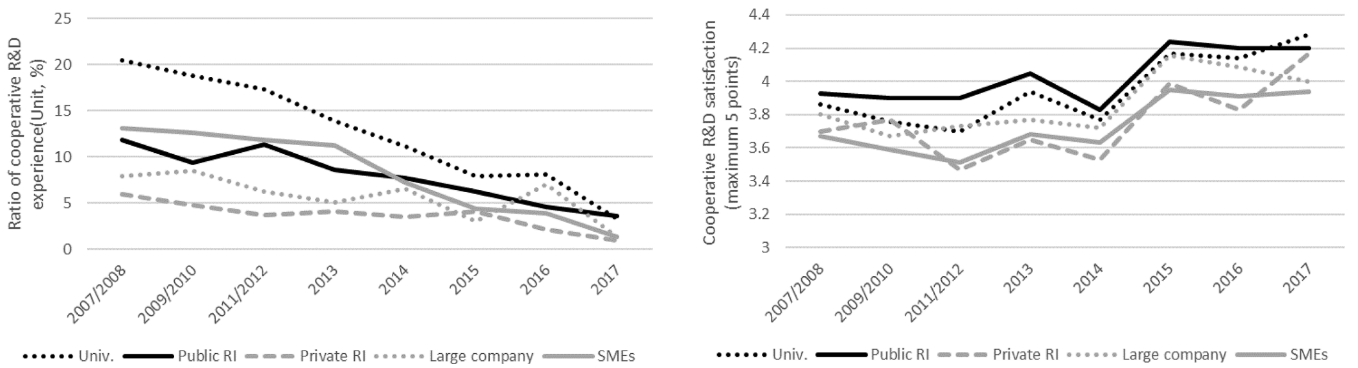

Fig. 1. Comparison of the ratio of SMEs with joint research experience (left) and their satisfaction levels (right) Data source: Compiled from data released by the Small and Medium Business Administration and Korea Federation of SMEs (MSS and Kbiz, 2017, 2018; SMBA and Kbiz, 2009, 2011, 2013, 2014, 2015, H. 2016).

applying decision tree analyses for the detailed models. In general, when a machine learning-based ensemble model is used alone, the expected predictive power is excellent, but there is also a clear limitation in that the effect of determinants cannot be explained (James et al., 2013). To overcome this limitation, we proposed a hybrid model that uses discriminant analysis in parallel and evaluated this hybrid model. This study utilized a large number of independent variables and applied machine learning to derive a set of significant candidate independent variables. Next, the narrowed field of independent variables was used to develop a model for recommending R&D collaboration partners or alternatives to SMEs by performing discriminant analysis. To recommend an optimal partner, this study used not only conventional discriminant analysis but also applied profiling techniques utilizing machine learning based decision tree analysis to identify the variables that are suitable for discriminant analysis.

Decision tree analysis, which is a type of data mining technique, searches for and identifies relations and patterns that exist in each data set and generates a model. This is a non-parametric method that does not require the assumption of linearity, normality or equality of variance (Choi et al., 2002). Both methods used in this research were multivariate analysis methods: multivariate analysis is a statistically based technique of simultaneously performing multiple univariate analyses, taking into consideration the correlations of the dependent variables, and the data used in this technique must be assumed to have multidimensional stationarity (Jun et al., 2017). In discriminant analysis, the predictor variables are assumed to have multivariate normal distribution. Although this assumption is disregarded in many actual cases, as in the case when a binary predictor variable is used, discriminant analysis performs well even when such issues exist (Shmueli et al., 2011). Discriminant analysis performs well even when the assumption of multivariate normal distribution is violated, as in the case of binary variable utilization, because in discriminant analysis the data is generally used only to find simple classification boundaries, such as a linear boundary (Friedman et al., 2001).

As discussed above, discriminant analysis is a type of parametric analysis and therefore is limited in the ability to perform direct analysis of categorical variables and it also requires caution when utilizing continuous variables. For this reason, to uphold the assumption of normality in the continuous variables (such as monetary values), conversions were performed, using a natural logarithm (refer to Appendix Table A). In the case of categorical variables, decision tree analysis was implemented to minimize the use of such variables. There are various methods available for performing decision tree analysis. In this study, the classification and regression tree method herein after referred to as CART (classification and regression tree) was used as well as CHAID, C5.0, and QUEST.[3] An optimal ensemble model that encompasses all the above these methods was adopted in this study (Breiman et al., 1984).

Fig. 2 explains the detailed process of developing a hybrid recommendation model using decision tree analysis and discriminant analysis. Steps 1 through 4 constitute the basic stage of preparing the data, and data conversion is a key component of this process. Among the analysis methods used in this research, discriminant analysis is a parametric analysis and a continuous variable (or dummy variable) is used as the independent variable. Among the independent variables (refer to Appendix Table A), there are multiple continuous variables, and in the descriptive statistics analysis results for continuous variables such as sales value, the skewness and kurtosis were found to be very high.[4] Therefore, it was difficult to uphold the assumption of normality in the distribution of the variables. For this reason, to perform discriminant analysis, some of the continuous variables that were used as the independent variables were converted to satisfy the assumption of normality. All variables measured by monetary value (e.g. sales value, R&D costs scale, etc.) were converted with a natural logarithm for use in the discriminant analysis.[5]

In Fig. 2, steps 5 through 8 are part of the process of identifying independent variables to be used in the recommendation model, through the supervised learning that takes place during machine learning. For each type of partner, we identified the high-satisfaction firms that were the subjects of our interest in this study, performed data balancing, then divided the data into training and test data to develop a decision tree for each type of partner.

Data mining builds a model through the learning of the analysis data, and in cases where the number of samples that belong to a specific class in the target variable is far greater than the number of samples that belong to another class, the learning for the class with the larger number of samples is performed more than the learning for the class without such a large number, resulting in a prediction model that is biased to be stronger in prediction for a specific class. Because of this issue, before establishing the prediction model, it is necessary to perform data balancing, which means balancing the ratio of the classes that exist in

[3] The data was split into analysis targets to perform a validation analysis of the decision tree analysis. 70% was set as the training sample and 30% was set as the test sample. The Gini index was used to measure impurity, with the minimum variation set at 0.0001 and the maximum depth set at 5. The cost of misclassification was not added in weight.

[4] Due to restrictions of page length, the descriptive statistics were omitted from Appendix Table A, but the majority had skewness and kurtosis of over ±2.

[5] If the minimum value was 0, we added 1 before conversion with a natural logarithm.

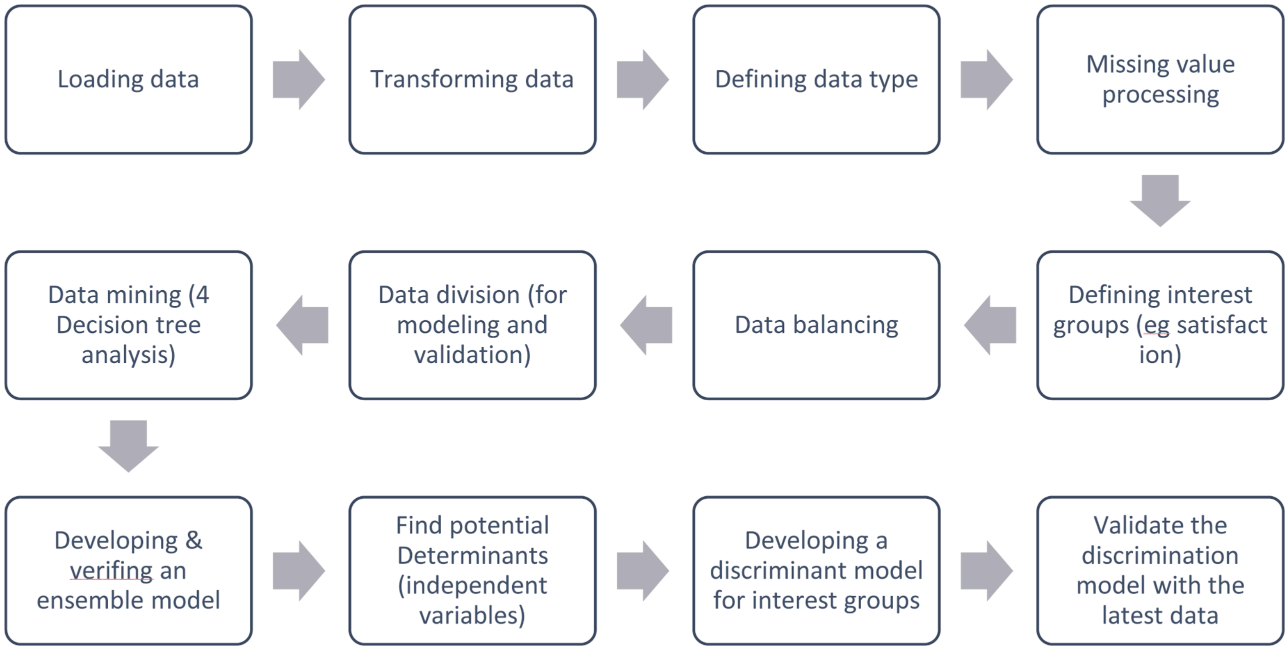

Fig. 2. Detailed process for developing a hybrid model to recommend types of R&D collaboration partners to SMEs.

the target variables.

Since the amount of data for a small class was insufficient for data mining, we chose to perform oversampling. To avoid the problem of overfitting, we endeavored to minimize data overlaps.[6] Increases in overlaps were restricted to a minimum of 30% and the oversampled data set was adopted as the final data set to which the data mining technique was applied.

In Steps 9 through 12 shown in Fig. 2, we identified the significant candidates for independent variables based on the results of the decision tree analysis we developed and completed the recommendation model through discriminant analysis. In this study, we did not specify a single decision tree analysis, and instead selected the variables that contribute strongly to prediction by developing an ensemble model that utilizes all four types of models. In Appendix Table B, we presented detailed results of ensemble modeling development, applied to each tree in the decision tree-based ensemble model development.[7] Although ensemble models greatly enhance the prediction capability of decision trees, a drawback is that ensemble models significantly lower interpretability (James et al., 2013). Therefore, we used an ensemble model only to select the variables that contribute to prediction to maximize predictive capability, then shifted to using discriminant analysis, which offers higher interpretability, to develop (or express) our recommendation model. We refer to the model presented in Fig. 2 as a "hybrid model" that is a mixture of two analytical models.

3.2. Research data and the scope of data

The data analyzed in this research was from the "8th Small and Medium Enterprises Technology Statistics Survey" released in 2015. This annual survey is conducted jointly by South Korea's Small and Medium Business Administration and the Korea Federation of SMEs in accordance with Article 8 of the *Act on the Promotion of Technology Innovation of Small and Medium Enterprise* ("Compilation of Statistics on Technologies of Small and Medium Enterprises").

The surveyed population consisted of 43,204 SMEs that were engaged in technology development, out of all firms in both manufacturing and non-manufacturing industries with 5 or more workers and less than 300 workers (34,745 manufacturing firms and 8459 non-manufacturing firms). From this population, a sample was extracted consisting of 3300 SMEs that were engaged in technology development as of December 31, 2014 (2636 manufacturing and 664 non-manufacturing) and the surveys were conducted by visiting. The surveys included information on the firms' technology innovation activities, current investment and technology level, and firm performance, which served as the raw data analyzed for this study.[8] To extract the sample from the population, we used the stratified sampling method. Sorting was performed according to stratified sampling classification variables from 29 industry divisions and four classifications based on the size of firms (number of workers) and then systematic sampling was performed. The survey covered a 1-year period from January 1, 2014 to December 31, 2014 and the survey was conducted from late June to late October in 2015 (SMBA and Kbiz, 2015). Table 2 shows our analysis of the descriptive statistics regarding the 3, 300 firms included in this study.

In this study, the results obtained from the survey questions were analyzed using the group variables (predicted variables or dependent variables). As seen in Table 2, the status of being a group that utilizes the respective joint research partner types or the status of demonstrating high satisfaction (4 points or more out of a maximum of 5 points) was

[6] Oversampling and undersampling are two techniques used to adjust the distribution of the ratio of classes in a data set. To perform oversampling, the class with a high ratio is taken as the basis and the classes with smaller ratios are repeatedly selected. One advantage is that the total data set may be increased, but since there may be many overlaps in the data, excessive oversampling may result in a problem of overfitting, generating a model that is suited only for the given data set (Jun et al., 2017).

[7] In Table B of the Appendix, we presented the details of the ensemble model development of CHAID and CART in the recommended model for PRI as an example, but we omitted the tree figures, which are results obtained from the analysis of individual decision trees, due to the length restriction of this paper.

[8] The survey results used in this study consisted of the analyses of 3,300 firms of survey raw data (primary data) received from the Korea Federation of SMEs for this study.

Table 2
Attributes of the analysis data.

Variables	Sub-group		Distribution ratio
Industry (Top 5)	Other machines and equipment		15.1%
	Metalworking products		10.9%
	Publishing (including software development and supply)		7.7%
	Electrical equipment		7.1%
	Electronic components, computers, video, sound and communication equipment		6.6%
Business age (Top 3)	More than 5 ~ 10 years		25.6%
	More than 10 ~ 15 years		23.9%
	More than 15 ~ 20 years		19.5%
Sales scale (Top 4)	More than 1 ~ 5 billion KRW		47.2%
	More than 10 ~ 50 billion KRW		21.6%
	More than 5 ~ 10 billion KRW		19.8%
	More than 500 million ~ 1 billion KRW		7.4%
Core Technology Sector (Top5)	Machinery & materials		35.6%
	Electrical & electronics		16.8%
	Chemistry		14.7%
	Information communication		12.1%
	Knowledge service		8.2%
Cooperative R&D experience group (Top 5)	University	Experience (367 cases)	11.1%
		High satisfaction (More than 4 out of 5) 250 cases	7.5%
	Public Research Institute (Pub. RI)	Experience (253)	7.7%
		High satisfaction (More than 4 out of 5) 176	5.4%
	Conglomerate company (large firm)	Experience (213)	6.5%
		High satisfaction (More than 4 out of 5) 155	4.7%
	Small & Medium Enterprise (SME)	Experience (236)	7.2%
		High satisfaction (More than 4 out of 5) 138	4.2%
	Private Research Institute	Experience (115)	3.5%
		High satisfaction (More than 4 out of 5) 62	1.9%

Source: (SMBA and Kbiz, 2015), here $ 1 is 1150 to 1200 KRW.

established as the dependent variable in the discriminant analysis. In this study, the interest group, which is judged to be the optimal partner, was the group that had engaged in joint research with the partner in question and expressed high satisfaction in the process or in response to the results; meanwhile, a group that did not have such experience or expressed low satisfaction was classified as the comparison group.[9] As seen in Fig. 1, out of the five types of R&D collaboration partners, private research institutes were found to have a notably low percentage in Table 2 as well. Therefore, in this study, our analysis focused on the top four types of R&D collaboration partners as indicated in Table 2. Although the survey explained in Table 2 was used for modeling and testing, we utilized the 9th, 10th and 11th "Descriptive Statistical Survey of Small and Medium Enterprises" released from 2016 to 2018 to perform additional verifications of the predictive capability of our developed model. We shall omit detailed explanations regarding the latter survey because it was only used for testing. We used SPSS statistics 24 to analyze and refine the data and used IBM Modeler 18 to develop ensemble models (decision tree analysis) and discriminant analysis models. The descriptive statistics of variables used in our machine learning-based decision tree analysis and discriminant analysis are presented in Appendix Table A.

4. Development of the recommendation model

The same survey results were also used for the discriminant analysis of SMEs' (technology) collaboration with other SMEs, and the data was derived from 3300 survey results. The discriminant analysis was performed with Fisher's linear discriminant function and the Wilks' Lambda value was adopted for the stepwise selection. The prior probability was calculated using the group sample size stated above. Among the 82 independent variables prepared from steps 1 to 4 of the research process presented in Fig. 2, there were independent variables that were identified through machine learning based decision tree analysis, as described in steps 5 to 10 of Fig. 2, and there were group variables (dependent variables), identifying the groups with high satisfaction for each of the institutions engaged in R&D collaboration, and on these discriminant analysis was performed using four types of partners as the criteria, and a discriminant analysis model was derived for each (Step 11 of Fig. 2) . As in the case of regression analysis, in discriminant analysis, forward or backward selection can be adopted for each of the variables, and in this study, the stepwise selection method was adopted for entering the variables (Wilks' Lambda value). Wilks' Lambda was used to evaluate the fitness of the discriminant of the clusters of high satisfaction for each R&D collaboration partner, using the stepwise selected variables. The results were confirmed to be statistically significant in all cases (with a difference between the two groups) ($p<0.000$).[10] For each of the discriminant models for the four respective types of partners, a minimum of 12 and a maximum of 19 variables were found to be significant variables in the model. To facilitate our understanding of the results of the discriminant analysis, we paired the partners of R&D collaboration with other organizations that had similar characteristics for comparison. Table 3 presents a model that predicts which firms will likely experience high satisfaction from R&D collaboration with universities or public research institutes. Table 4 distinguishes between cases of R&D collaboration with large firms and cases of collaboration with other SMEs. This table presents standardized and non-standardized coefficients together. Only standardized coefficients were used in our comparative analysis of the impact of coefficients; the non-standardized coefficients and the center point of the distribution shown at the bottom of the table are provided so readers may utilize this data to perform predictions using the model.

Table 3 presents the standardized (STD) canonical coefficients for the model that discriminates the firms that exhibited high satisfaction regarding the experience of R&D collaboration with universities and public research institutes. According to these results, the SMEs that engaged in technology collaboration with universities and expressed high satisfaction were those that invested more in joint research than in R&D costs, relied on a high ratio of commissioned research, and conducted active research despite the relatively long total period required for technology development. The firms that experienced high satisfaction from collaboration with universities and firms were the firms with a large number of research personnel but with a relatively small number of full-time employees, in other words, firms which were characterized by high research concentration (the ratio of researchers compared to other employees). The analytic results regarding public research institutes, shown on the right-hand side of Table 3, showed many similarities to the trends we found in the analysis results regarding universities, shown on the left. In both cases, the R&D cost for joint research had the largest standardized coefficient; this was followed by the ratio of commissioned research, which had a large influence. Also,

[9] In general, the success of joint research can be specified in several ways (e.g. technology development, sales, etc.). In this study, we considered satisfaction as the most appropriate indicator in terms of satisfying the meaning of various successes, and we used high satisfaction as operational definition of the success of joint research.

[10] Tables 3 and 4 show statistical values (Canonical correlation coefficient, Wilks' Lambda, Chi-square) that can be used to evaluate the suitability and performance of the discriminant.

Table 3
Comparison of the canonical discriminant function coefficients for the R&D collaboration recommendation model developed for SMEs for partnership with universities and public research institutes.

High satisfaction from R&D collaboration with universities			High satisfaction from R&D collaboration with public (government-funded) research institutes		
Variables	Coefficients		Variables	Coefficients	
	STD	Non-STD		STD	Non-STD
Transaction client sales ratio: large firms	−0.091	−0.002	Transaction client sales ratio: exports	−0.059	−0.003
			Technology development implementation ratio: joint research	0.086	0.004
Technology development implementation ratio: commissioned	0.405	0.034	Technology development implementation ratio: commissioned	0.491	0.034
Technology development implementation ratio: tech. adoption	−0.065	−0.008			
Commercialization capabilities (global comparison)	0.073	0.006			
			Design capabilities (global comparison)	0.143	0.011
			Parts and process design capability (global comparison)	−0.093	−0.007
			Intellectual property rights in ownership: domestic design	0.054	0.008
Intellectual property rights in ownership: domestic patent application	0.073	0.014	Intellectual property rights in ownership: domestic patent application	0.070	0.011
Time required: total technology development	0.218	0.020			
			Time required: opening new sales markets	0.133	0.023
No. of times test equipment was used: professional testing institutions	0.112	0.009			
No. of times test equipment was used: universities	0.157	0.022	No. of times test equipment was used: public research institutes	0.158	0.036
			No. of times test equipment was used: large firms	−0.065	−0.048
			Technology development performance outcomes: No. of IPR registered	0.051	0.050
Total number of researchers (LN)	0.091	0.098	Total number of researchers (LN)	0.125	0.133
Total number of employees (LN)	−0.122	−0.114			
R&D costs scale (LN)	−0.156	−0.113	R&D costs scale (LN)	−0.223	−0.157
R&D costs utilization - joint research (LN)	0.985	0.548	R&D costs utilization - joint research (LN)	0.908	0.490
			R&D costs procurement - government sponsored (LN)	0.164	0.072
(Constant)		−1.129	(Constant)		−1.121
Center point of distribution	Yes: 1.495 No: −0.753		Center point of distribution	Yes: 1.530 No: −0.844	
Canonical correlation coefficient	0.728		Canonical correlation coefficient	0.751	
Wilks' Lambda	0.470		Wilks' Lambda	0.436	
Chi-square (significant probability)	2319 (p<0.000)		Chi-square (significant probability)	2679 (p<0.000)	

Note: STD coefficients mean standardized canonical coefficients.

the total R&D cost incurred by firms had a negative effect on satisfaction in both cases.

In Table 3, there were some variables that showed divergent trends in public research institutes and universities: the time spent in the stage of finding sales outlets was longer, the firms' design capability was higher, and they had a larger number of design rights. However, since the firms that had a smaller scale of export and were lagging in design capabilities for parts and processes tended to exhibit high levels of satisfaction, it was possible to deduce that the public research institutions had supplemented the firms in these aspects.

Table 3 thus demonstrates that while the profiles of firms that were highly satisfied with R&D collaboration with universities and public research institutes have some common features, there are also differences in specific areas. First of all, there were differences in the forms of technological development: in the cases of high-satisfaction R&D collaboration with universities, only the ratio of commissioned research had a significant influence, but in the case of public research institutes, joint research as well as commissioned research had significant influence, confirming that joint research is likely to yield satisfaction more strongly in the case of public research institutes than universities. There was also a great difference in the firms' capabilities: commercialization capabilities were important for R&D collaboration with universities, but design capabilities and part process designing capabilities were found to be important for R&D collaboration with public research institutes. This confirmed that there are differences in the capabilities that can be provided by each of these types of R&D collaboration. In addition, there were differences in the time required for technological development and the pattern of utilizing external testing equipment. Notably, firms reporting high satisfaction from R&D collaboration with public research institutes obtained many registered patents as the outcome of technology development performance, indicating that public research institutes serve as useful partners for acquiring intellectual property rights. Another noteworthy difference in the profiles of the firms that respectively showed high satisfaction regarding universities and public research institutes is the difference in the method of procuring research expenses: firms experiencing high satisfaction with public research institutes had a strong contrasting tendency to procure the research cost through government sponsorship.

Table 4 compares the standardized canonical coefficients of the model that discriminates the firms that respectively exhibited high satisfaction from R&D collaboration with large firms and other SMEs. These results shown on the left side of Table 4 indicate that SMEs that experienced high satisfaction from technology collaboration with large firms tended to invest a significantly large portion of their research cost in joint research, but also invested a large amount in internal research. Also, the ratio of large firms among their transaction clients was high, the firms' product planning capabilities and maintenance and repair capabilities were strong, and the firms spent a long time in the stage of pioneering sales outlets. By contrast, firms that expressed high satisfaction regarding R&D collaboration with large firms tended to be weaker in their testing inspection capabilities and manufacturing capabilities and had a low ratio of internal research development. Compared to the data presented earlier in Table 3, it was notable that the cost spent on joint research and the ratio of commissioned research was not high and the age of the CEO and year of establishment were significant, while the ownership of intellectual property rights and the number of researchers were negligible.

Firms that showed strong satisfaction about their R&D collaboration with other SMEs, shown on the right side of Table 4, had very marked differences compared to the cases of R&D collaboration with large firms, shown on the left. The former firms tended to invest a large portion of their R&D costs in joint research or commissioned research and the ratio

Table 4
Comparison of the canonical discriminant function coefficients for the R&D collaboration recommendation model developed for SMEs for partnership with large firms or other SMEs.

High satisfaction from R&D collaboration with large firms **Variables**	**Coefficients** **STD**	**Non-STD**	**High satisfaction from R&D collaboration with SMEs** **Variables**	**Coefficients** **STD**	**Non-STD**
Age of CEO / representative	−0.127	−0.015			
Main product sales ratio	0.048	0.006			
Year of establishment	0.075	0.004			
Transaction client sales ratio: large firms	0.344	0.010			
Transaction client sales ratio: SMEs	0.064	0.002	Transaction client sales ratio: SMEs	0.189	0.005
Technology development implementation ratio: internal research	−0.284	−0.011			
Technology development implementation ratio: joint research	0.380	0.015			
Technology development implementation ratio: commissioned	0.056	0.006	Technology development implementation ratio: commissioned	0.350	0.024
			Technology development implementation ratio: tech. adoption	0.049	0.006
			New technology development capabilities (global comparison)	0.107	0.009
Product planning capabilities (global comparison)	0.077	0.007			
Testing inspection capabilities (global comparison)	−0.098	−0.007			
Manufacturing and processing capabilities (global comparison)	−0.092	−0.008			
Maintenance and repair capabilities (global comparison)	0.158	0.014			
			Intellectual property rights in ownership: design applications in foreign countries	0.120	0.048
			Intellectual property rights in ownership: domestic patent applications	−0.132	−0.024
			Time required: development planning	−0.192	−0.058
			Time required: development progression	0.103	0.020
Time required: opening new sales markets	0.186	0.033	Time required: opening new sales markets	0.135	0.025
No. of times test equipment was used: large firms	0.197	0.115			
Ordinary profit	−0.059	−0.000			
New technology sales ratio	−0.050	−0.003	New technology sales ratio	0.074	0.004
			Technology development performance outcomes: No. of IPR applied	0.131	0.081
R&D costs scale (LN)	−0.320	−0.224	R&D costs scale (LN)	−0.192	−0.139
R&D costs utilization – internal research (LN)	0.387	0.179			
R&D costs utilization - joint research (LN)	0.553	0.319	R&D costs utilization - joint research (LN)	0.958	0.484
			R&D costs utilization – commissioned research (LN)	0.384	0.251
			R&D costs procurement - private financing (LN)	−0.055	−0.035
			R&D costs procurement - private investment (LN)	0.121	0.303
(Constant)		17.547	(Constant)		−1.546
Center point of distribution	Yes: 1.915 No: −1.078		Center point of distribution	Yes: 1.360 No: −0.805	
Canonical correlation coefficient	0.821		Canonical correlation coefficient	0.723	
Wilks' Lambda	0.326		Wilks' Lambda	0.477	
Chi-square (significant probability)	3682 (p<0.000)		Chi-square (significant probability)	2526 (p<0.000)	

Note: STD coefficients mean standardized canonical coefficients.

of other SMEs among their sales transaction clients was also high. Furthermore, these firms responding with high satisfaction levels had strong capabilities in new technology development and tended to spend a long time in development and in opening new sales markets. When compared to cases of R&D collaboration with large companies, these firms even exhibited significance in a different direction for the same variable. Observing the new technology sales ratio, we found that there was a negative effect on the satisfaction regarding R&D collaboration with large companies, in contrast to the positive effect on R&D collaboration with SMEs. Compared to the cases of R&D collaboration with large firms shown above in Table 3 and Table 4, in terms of the technology development method, the technology implementation of these firms significantly tended toward the positive direction while their performance outcomes measured in terms of applications for domestic intellectual property rights leaned significantly toward the negative direction.

Let's compare the two cases compared in Table 4 in more detail, based on the significant variables and standardized coefficients. In the cases of high satisfaction from R&D collaboration with large firms and SMEs, there were of course important differences in the ratio of the transaction client: in cases where the ratio of the transaction client was high, the satisfaction levels were also high. There were also significant differences in the forms of technological development. In cases of high satisfaction from R&D collaboration with large firms, the ratio of joint research was high, but firms with high satisfaction with SMEs were found to have a high ratio of commissioned research. Between these two cases, there were also differences in the types of technological development capabilities and intellectual properties possessed by the firms and the time required for each step of technological development. The comparative results presented in Table 4 shows that firms gaining high satisfaction from R&D collaboration with large firms did not have strong in-house innovation capabilities. The new technology sales ratio was low, the levels of intellectual property and R&D personnel were not significant, and the manufacturing and processing capabilities were also low. By contrast, firms that experienced high satisfaction from R&D collaboration with SMEs were found to have relatively strong innovative capabilities of their own and strong capabilities for new technology development; they also had registered intellectual property rights in other countries and accomplished a high number of intellectual property rights applications as the outcome of R&D development. Therefore, we deduced that the collaboration aims of firms with high satisfaction from R&D collaboration with large firms were likely to be more focused on securing capabilities through transactions with the large firms, rather than on improving the firm's own technological capabilities.

5. Discussion

5.1. Evaluation of the prediction capability of the recommendation model

To evaluate how well this discriminant method performs predictions, we can adopt the cross-validation method of dividing the same data set into training test and test data, or the method of using completely different data sets. We used both of these methods to verify whether the prediction and recommendation method we developed in this study is suitable (Step 12 in Fig. 2).

To test how well predictions were performed by the classification function for high-satisfaction firms in relation to each type of partner, proposed in Table 3 and Table 4, we performed an analysis using a classification confusion matrix. Table 5 presents the verification results for the modeling of high satisfaction groups in relation to the types of partners. Although the accuracy of the high satisfaction firm model developed through discriminant analysis was verified to some degree using cross-validation values, to achieve a more objective verification of the model's accuracy, we performed additional verification using new data that had not been utilized during the model's development. All the data (3300 cases) from the "9th Descriptive Statistical Survey of Small and Medium Enterprises," released in 2016, were applied to the high satisfaction firm prediction model that identifies optimal R&D collaboration partners across the various types of partners. Table 5 presents the results of the analysis of the prediction model's classification accuracy.

Table 5 shows that in the case of R&D collaboration with universities, the verification data for the discriminant results also found the accuracy to be 87.11% in the case of the group with high satisfaction in relation to universities. The new data set from the 2016 survey (including 3300 cases) also yielded a high accuracy rate of 88.42%. Notably, in cases where the response was "NO (group with no prior experience or dissatisfaction)," the specificity was found to be high, respectively 87.40% and 88.77%. Furthermore, the sensitivity (or recall) in cases where the response was "YES (high satisfaction)" was also high, respectively 83.12% and 83.84%, which verified that this model has strong discriminant capability and reliability.

Table 5 shows the verification results for the hybrid model developed for the firms with high satisfaction in collaboration with public research institutes, indicating that the model's discriminant results predicted the high satisfaction group. The accuracy of the discriminant results for the high satisfaction group in relation to public research institutes was 87.08% in the case of the data used in the modeling and an even higher percentage of 88.76% in the case of the new data from the 2016 survey. The specificity to the cases responding "NO (no experience or dissatisfaction)" were high, respectively 87.04% and 88.68%, and furthermore, the sensitivity in the cases of "YES (high satisfaction)" were also high, respectively 87.23% and 90.06%. These results verified that the model has excellent discriminant capability and is highly reliable.

The results of model verification performed on the group of firms with high satisfaction in relation to large firms, presented in Table 5, also proved the model to perform well, with accuracy levels of 89.76% and 89.58% respectively and specificity levels of 90.70% and 89.69% respectively. The sensitivity levels in cases where the response was "YES (high satisfaction)" were also strong, respectively 73.21% and 85.19%. The verification results for the model designed for the group of firms with high satisfaction in relation to SMEs were also overall strong, but the accuracy levels were respectively 85.65% and 79.21%, which means that the accuracy was slightly lower in the results from the new data.

These results demonstrated the strong predictive capability of the model for the four types. The hit ratio was 80% or more not only for the training data set but even for the completely new data that had not been used in training, indicating that the model is able to perform predictions stably. Based on these analysis results, we concluded that the discriminant model proposed in Table 3 and Table 4 succeeded in generating results that were useful for predicting the high satisfaction firms for each type of collaboration partner.

We performed testing on two additional aspects. First, considering that when the hybrid model was compared with the ensemble model alone, it was found that the interpretation power improves but the predictive power can be halved (James et al., 2013), we verified the difference between the predictive power of the ensemble-only model and the hybrid model with the addition of discriminant analysis. The second issue was to address the question of the degree to which the predictive power has been sustained since 2016 until recently. Beginning with the 2017 survey, the criteria for classifying SMEs changed greatly from the size of employees to the size of sales, so there was inevitably a difference between the population and samples of the survey. We tested whether the predictive power persisted even in this severe validation set. Table 6 presents the results of these additional tests performed to address the challenge presented by the aforementioned two aspects.[11]

The results in Table 6, relevant to the first challenge of comparing the hybrid model's predictive power to that of the ensemble-only model (2015 Survey), demonstrate that the accuracy declined slightly, as expected. However, in terms of recall (or sensitivity) and the F1-Score, it can be seen that the hybrid model is superior in most cases. Especially in cases where the data set is imbalanced as in the case of our study, recall, precision, and the F1-Score may be more appropriate as a basis for determining the predictive power of a model than accuracy (Jeon et al., 2020; Jun et al., 2020). In that regard, we can conclude from the results of Table 6 that the hybrid model proposed by this study did not have a significant decrease in accuracy compared to the ensemble model and that it was an excellent model yielding a very significant improvement in the recall value.[12]

Table 6 shows the test results relevant to the second challenge of sustaining predictive power (2016–2018 survey), which indicate that accuracy increased in the 2017 and 2018 surveys, surveys in which the population and sample were different, as explained above. This can be attributed to a distortion phenomenon that appeared due to the decrease in the target group (ratio), as confirmed in Fig. 1. As the imbalance of the data set became excessively severe, the accuracy increased (increase in specificity levels). This effect also led to a decline in recall values. In terms of maintaining predictive power, considering the change of the test set, the results confirmed that the predictive power of the recommendation model for universities and government-supported public research institutions continued to be excellent up to the 2018 survey. While the persistence of predictive power was somewhat weak in the recommendation model for large firms and SMEs, this can be explained by the fact that the distribution of the target group (high satisfaction firms) was very low, rendering it difficult to maintain predictive power. In Table 6, it can be seen that the target ratio, which was 4% in the 2015 survey used in the model development, decreased significantly to 0.5% and 1.1% in the 2018 survey.

In summary, compared to the ensemble-only model, the hybrid model we proposed demonstrated superior predictive power in an imbalanced data set as used in our case and maintained some degree of predictive power even under severe conditions where the characteristics of the data set changed significantly. Even when the ranged of the validation data was expanded from 2016 to 2018, the case weighted average of accuracy was 91.3% and the recall (or sensitivity) was 73.7%. This confirmed that the approach of the hybrid model we proposed is suitable for recommending R&D collaboration partners to SMEs. These findings are significant in that they confirm the possibility that the

[11] In the 2018 Survey, certain questionnaire items were also removed, and therefore we performed predictions ignoring the variables corresponding to the removed questionnaire items.

[12] The fact that accuracy or recall was greatly improved in the hybrid model, in which the variables used for model development were reduced, could be explained by the effect of the improvement of the overfitting tendency found in the ensemble-only model.

Table 5
The classification confusion matrix for discriminant analysis.

Partner	Actual Class	Predicted Class Modeling (2015 survey) No	Yes	Hit ratio (%)	Validation (2016 survey) No	Yes	Hit ratio (%)
Universities	**No**	848	122	87.40%	2726	345	88.77%
	Yes	13	64	83.12%	37	192	83.84%
	Overall%	82.20%	17.80%	87.11%	83.73%	16.27%	88.42%
Public Research Institutes	**No**	869	129	87.04%	2766	353	88.68%
	Yes	6	41	87.23%	18	163	90.06%
	Overall%	83.73%	16.24%	87.08%	84.36%	15.64%	88.76%
Large firms	**No**	897	92	90.70%	2887	332	89.69%
	Yes	15	41	73.21%	12	69	85.19%
	Overall%	87.23%	12.73%	89.76%	87.85%	12.15%	89.58%
SMEs	**No**	858	140	85.97%	2524	664	79.17%
	Yes	10	37	78.72%	22	90	80.36%
	Overall%	83.06%	16.94%	85.65%	77.15%	22.85%	79.21%

Table 6
Comparative analysis of the models' predictive power.

Partner	Statistic	Unit	Modeling (test set) Ensemble –only Model (2015 survey)	Hybrid Model(2015 survey)	Validation Hybrid Model(2016 survey)	Hybrid Model(2017 survey)	Hybrid Model(2018 survey)
Universities	**Recall**	%	42.9	83.1	83.8	71.3	74.8
	Precision	%	47.1	34.4	35.8	41.1	36.9
	Accuracy	%	92.2	87.1	88.4	90.5	95.5
	F1-Score	%	44.9	48.7	50.1	52.1	49.4
	Test cases	Cases	1045	1045	3300	3300	3800
	Target ratio	%	11.1	11.1	6.9	7.3	2.9
Public Research Institutes	**Recall**	%	34.0	87.2	90.1	70.7	76.0
	Precision	%	27.6	24.1	31.6	25.0	42.8
	Accuracy	%	93.0	87.1	89.1	90.3	96.0
	F1-Score	%	30.5	37.8	46.8	36.9	54.8
	Test cases	Cases	1045	1045	3300	3300	3800
	Target ratio	%	7.7	7.7	5.5	4.0	3.2
Large firms	**Recall**	%	48.2	73.2	85.2	69.8	52.6
	Precision	%	64.3	30.8	17.2	22.6	7.1
	Accuracy	%	95.8	89.8	89.6	93.0	96.3
	F1-Score	%	55.1	43.4	28.6	34.2	12.5
	Test cases	Cases	1045	1045	3300	3300	3800
	Target ratio	%	4.7	4.7	2.5	2.6	0.5
SMEs	**Recall**	%	12.8	78.7	80.4	66.0	67.5
	Precision	%	18.2	20.9	11.9	17.6	13.0
	Accuracy	%	88.3	85.6	79.2	89.9	94.9
	F1-Score	%	15.0	33.0	20.8	27.8	21.8
	Test cases	Cases	1045	1045	3300	3300	3800
	Target ratio	%	4.2	4.2	3.4	2.9	1.1

hybrid model can be utilized in an information filtering-based recommendation model as well as for the purpose of R&D collaboration partner recommendation for SMEs.

5.2. Utilization of the recommendation model

The following is a detailed explanation of how to use the model for predicting firms that will experience high satisfaction, shown in Table 3 and 4. Whether a firm is included among firms with high satisfaction regarding R&D collaboration with each type is determined by constructing a model based on the non-standardized (Non-STD) coefficients among the models shown in Table 3 and 4, inputting the profile of the individual firm, and finding the group with center value (for example, in Table 3, in the case of universities, "Yes" is 1.495, and "No" is −0.753) to which the derived prediction value (discriminant score) is closest, in order to identify the discriminated group. Using this method, we can predict the cases in which a specific firm will be discriminated to be a high-satisfaction firm, among the four models shown in Tables 3 and 4. Based on these findings, we can recommend partners for R&D collaboration. To ensure that this study is useful for practical application by firms and policy leaders, in Section 4 we propose several specific means of utilizing the recommendation model we developed. The recommendation models for the four types of collaboration partners presented in Section 4 indicate a discriminant value for each respective type. We propose a method of expressing the suitability of each partner type in an easily comprehensible manner, using distance and distribution. Here, distance refers to the distance from each center point. Providing the distance from the respective center points of the four types facilitates comparisons of the relative values of the four types. On the other hand, distribution is used when one wants to know the absolute fitness of a specific type. Distribution here means the distribution of the 3300 discriminant scores, obtained when the information on the 3300 firms that had been used for training in the model development process was applied to recommendation model. When the information on new firms (or profiles) that were not used for data training is input, this can be compared to the distribution of the discriminant scores that had been used in previous training, allowing us to determine the fitness of a single type relatively objectively.

Thus far, we demonstrated that the recommendation model of R&D partner types for SMEs appeared to have predictive capability, and we also presented ideas for practically utilizing this model. By adopting the decision tree analysis and discriminant analysis used in this study, it will be possible to construct an intelligent system for recommending R&D collaboration partners to SMEs. Based on our research results, it would be possible to design a system or service in which specific firms can participate in a survey to identify their profiles and match them to the type of SMEs known to gain high satisfaction from R&D collaboration with each of the four alternative types of partners and compare the firm's discriminant scores for each of the four partner types to recommend a specific partner.

5.3. Evaluation of the determinants in the recommendation model

The hybrid model proposed herein maintained predictive power, and furthermore, one particularly notable value of this model is that it made it possible to perform a statistical analysis of the effects of determinants that could not be attempted when using only the ensemble model. Although the main purpose of this study was to develop a recommendation model, the interpretation of the determinant factors found in the recommendation model is also of scholarly significance.

Observing the determinant factors of firms that experience high satisfaction from R&D collaboration with the respective types, in the recommendation model presented in Tables 3 and 4, we see that the research results cited above confirm that R&D costs and the amount invested in joint research during R&D were the variables that were significant in all aspects. The scale of R&D costs had a negative correlation, whereas joint research costs had a positive correlation. It is to be expected that joint research costs would have a positive correlation since the existence of this expense indicates that the firm has joint research experience, and this finding verifies that the model is reliable. On the other hand, the fact that the scale of R&D costs, that is, total expenditure, appeared to have a negative influence on satisfaction may be perceived as a result contrary to previous studies analyzing R&D cooperation (Eom and Lee, 2010; Fernández-Olmos and Ramírez-Alesón, 2017; Gkypali et al., 2017). The research results of Gkypali et al. (2017) demonstrated that R&D investment has a positive effect on innovation performance indicators such as asset and profit margins and according to Eom and Lee (2010), R&D intensity has a positive effect on different types of performance depending on the partner. Considering that R&D investment or intensity can represent a company's R&D or even its absorptive capability (Kafouros et al., 2020), companies with high R&D capabilities are generally expected to have high innovative performance resulting from R&D collaboration (Eom and Lee, 2010; Fernández-Olmos and Ramírez-Alesón, 2017; Gkypali et al., 2017), but the results of our study indicated that firms with low R&D capability may experience higher satisfaction. This can be explained from an RBV perspective, namely that satisfaction may actually be higher when an SME encounters external R&D resources at a level it does not currently have. This is one important implication that our study was able to reveal because we adopted satisfaction as our target variable. From a conventional perspective that is oriented toward economic or technical performance, SME R&D collaboration should be encouraged in superior companies with large R&D investment, but by contrast, our research results show that even companies with inferior R&D capabilities, even if they are unable to derive innovation results right away through R&D collaboration, need R&D collaboration and should be encouraged to pursue it, based on assessments of intangible performance and satisfaction.

Comparing the differences among the alternative partners, calibrated with their respectively standardized sizes, we were able to rank the types in the order of their utilization of joint research development costs as follows: universities > SMEs > public research institutes > large firms. The rankings were the same when ordered by the size of R&D costs. The comparison of the types indicated that while SMEs that engage in technology collaboration with universities tend to have a relatively large amount of R&D funding and invest a lot of resources into joint research, SMEs that engaged in technology collaboration with large firms tended to be the firms that made the least amount of investment in R&D and in technology collaboration, compared to other SMEs that engage in technology collaboration with other alternative types of partners. Firms found to experience high satisfaction regarding technology collaboration with SMEs were found to be at an inferior status not only in terms of research capabilities but also in terms of their financial environment: Table 3 shows that these firms had inadequate access to private financing when procuring funding.

Table 3 and Table 4 presents the differences among the alternative types of partners and reveals the variables that conform well to the theoretical explanation provided by TCE and RBV. The percentage of transaction client sales, which exists in Table 4 but is not found in Table 3, is shown to be a positive significant variable, and the TCE concept of the frequency of transaction provides a strong explanation for this. By contrast, the number of times testing equipment was used, a variable that is missing only in the cases of technology collaboration with SMEs according to Table 3 and Table 4, is a typical example explainable through RBV. The desire to strategically utilize the respective resources of the various alternative partners motivates technology collaboration, and results in high satisfaction. There are differences that can also be explained by asset specificity in the framework of TCE. First, in the case of intellectual property rights, universities and public research institutes exhibited a positive correlation as seen in Table 4. In Table 4, in contrast, SMEs are shown to have a negative correlation, and in the case of large firms, the correlation was not significant. This indicates that depending on whether the partner is a potential competitor, the satisfaction experienced from the collaboration is affected by the asset specificity of technology. Also, Table 3 and Table 4 indicate that the types of R&D capabilities that are statistically significant (in a global comparison) will vary greatly depending on the type of partner chosen. These variables corroborate the theory of RBV, which explains that the motivation for R&D collaboration with the types of partners will be affected by the respectively differing capabilities. Among the variables pertaining to development time periods, the time expended on finding sales outlets was identified to be a significant variable in the case of public research institutes, large firms, and SMEs. This time period was especially long in the case of large firms, which confirmed once more that when a SME engages in technology collaboration with large firms, the effort to maintain and secure a transaction partner will likely be a weighty issue.

These results verified that both TCE and RBV affects the technology alliance governance structure and confirmed that utilizing these attributes will enable us to recommend to SMEs the types of partners for R&D collaboration that are optimal for the firms' needs. Another significance of this study is that it underscored the need to consider determinants that had not received much attention in previous studies on SMEs' R&D collaboration, such as the time required for opening new sales markets, the number of times test equipment was used, and the transaction client sales ratio.

6. Conclusions and issues for further research

This study proposed a method of recommending R&D collaboration partners for SMEs, using machine learning based decision tree analysis and discriminant analysis. We also demonstrated that this hybrid method has strong prediction capability by completing cross validation as well as verification using a new data set. We verified that in the case of an imbalanced data set, the proposed hybrid model can have superior predictive power compared to the ensemble model alone and that the hybrid model maintains a relatively robust predictive power even under severe conditions, as seen when applied to a survey in which the characteristics of the data set changed significantly. Through these findings, this study significantly demonstrated that the proposed hybrid model is suitable as a partner recommendation model for SMEs and can an effective new method for predicting an imbalanced data set.

This study also discussed methods of utilizing the recommendation model to practically assist firms and policy makers. After the profile of an individual firm is input, it is compared to the profiles of firms that had high satisfaction with each of the types of R&D collaboration partners, based on machine learning training. In our proposed method, a discriminant score is assigned through discriminant analysis, and comparisons are made to the distribution (center point) of cases prepared in advance to discriminate whether a firm can be predicted to have high satisfaction. By thus using the recommendation model, firms will be able to select optimal R&D collaboration types on their own, while policymakers will be better able to select firms that are best suited to the R&D collaboration type available in public support programs.

We also sought to make a scholarly contribution by interpreting the determinant factors in the recommendation model. This study identified the features of firms that reported high satisfaction with each type of R&D collaboration partner. A common feature was that the ratio of joint research was high in these firms' use of R&D costs, and the ratio of commissioned research was also high, but the R&D expenditure was low. However, there were differences among the types of partners for some variables: the technological development agreement method, the types of R&D capabilities (in global comparison) and the time required for each stage of technological development were variables which exhibited very large differences. Among the four types of R&D collaboration partners, the profiles for universities were similar to those for public research institutes: firms with high satisfaction had relatively high R&D costs and invested a lot of resources into joint R&D. By contrast, in the case of public research institutes, the ratio of joint research was high and especially in cases where the research cost was procured from the government, the firms' satisfaction with R&D collaboration was high. Firms that gained high satisfaction from R&D collaboration with large firms showed especially notable differences compared to other partners. Compared to the cases with other partners, these firms had the smallest investments in R&D or R&D collaboration, and this indicated that these firms' motivation for R&D collaboration was to gain from transactions with the large firms rather than to strengthen their own technological capabilities. In cases where firms engaged in R&D collaboration with other SMEs, a notable feature was that if the firms procured R&D funding from private sources, the source had significant effects on satisfaction: private loans had negative effects, whereas private investments had positive effects. The characteristics of the firms that achieve high satisfaction level in relation to the respective R&D collaboration types may also be of scholarly interest for related studies.

We anticipate that this study will also contribute to enhanced policymaking. We analyzed current information on R&D collaboration by SMEs in South Korea by type and our research also proposed a method of efficiently identifying the optimal partner for R&D collaboration among the various types available, by correlating them to the individual attributes of the SMEs. This approach aims to improve the success rate of R&D collaboration undertaken by SMEs and will help enhance efficiency in administration and budget implementation.

One caveat to consider in the interpretation of the results of this research is that the profiles of the various alternative partners presented in Table 3 and Table 4 may not reflect the outcomes of collaboration. Although it is to be expected that sales values or investment values would rarely be incorporated in these profiles, some of the capabilities may indeed have been reflected in the profiles, and this poses a limit to interpretation. One issue that this study has not adequately addressed in the fact that, while it is important to identify firms with high potential to successfully engage in R&D collaboration with public research institutes, it is also important to address the system approach, from the perspective of the policy and administration of public research institutes and innovation system theory. In addition, beyond the recommendation of a partner, there will be bound to be demands for more specific recommendations regarding each actor, and this should be considered one of the current limitations of this study. We expect that further research into the detailed attributes of partner candidates will generate more studies that contribute to a wider range of fields. To generate recommendations that are not merely at the level of choosing the type of R&D collaboration partner but more specific in identifying potential collaboration partners, there needs to be access and sharing of detailed information (such as contract information) not only regarding the firms but about individual collaboration partners as well. If the method presented in this study is expanded by utilizing such detailed information, it will be possible not only to identify recommended types of partners but also to select individual partners to recommend. Another limitation of this study is that we did not distinguish between low-satisfaction firms and non-satisfaction groups because we concentrated on predicting high-satisfaction groups. Therefore, it should be kept in mind that the potential collaborative research high satisfaction group proposed by this study represents a mixed profile of the group that currently experiences collaborative research.

Funding source

This research was supported by Korea Institute of Science and Technology Information (KISTI).

Data reference

The survey raw data (primary data) in this study received from the Korea Federation of SMEs for this study.

Author statement

CRediT author statement

Term	Seung-Pyo Jun	Hyoung Sun Yoo	Jeena Hwang
Conceptualization	V		
Methodology	V		
Software	V		
Validation	V		
Formal analysis	V		
Investigation	V	V	V
Resources		V	
Data Curation	V		
Writing - Original Draft	V	V	V
Writing - Review & Editing	V	V	V
Visualization	V	V	V
Supervision	V		
Project administration	V		
Funding acquisition	V		

02. Sep. 2020. Seung-Pyo, Jun

Declarations of Competing Interest

None.

Acknowledgements

Seung-Pyo Jun, the first and corresponding author, acknowledges support from the Research Program at Korea Institute of Science and Technology Information. The authors would like to express special gratitude to Korea Federation of SMEs for sharing valuable survey data for our research and to our reviewers and editors who have generously provided helpful feedback to improve this paper.

Appendix Table A. Descriptive statistics for determinants

Continuous Variable Name		Units	N	Minimum	Maximum	Average	S.D.
R&D costs scale (LN)		Mil. KRW	3300	0.7	10.4	5.2	1.4
Technology development implementation ratio	Internal research	%	3300	0.0	100.0	87.5	26.1
	Joint research	%	3300	0.0	100.0	9.4	22.9
	Commissioned research	%	3300	0.0	100.0	1.6	8.9
	Technology adoption	%	3300	0.0	100.0	1.3	9.1
	Etc.	%	3300	0.0	95.0	0.2	2.7
R&D costs utilization (LN)	Internal research	Mil. KRW	3300	0.0	9.5	4.9	1.7
	Joint research	Mil. KRW	3300	0.0	8.3	1.0	2.0
	Commissioned research	Mil. KRW	3300	0.0	10.0	0.2	1.0
	Technology adoption	Mil. KRW	3300	0.0	7.0	0.2	0.8
	Etc.	Mil. KRW	3300	0.0	6.4	0.0	0.3
R&D costs procurement (LN)	Internal raising	Mil. KRW	3300	0.0	9.7	4.9	1.5
	Government sponsored\	Mil. KRW	3300	0.0	8.5	0.4	1.4
	Government loans	Mil. KRW	3300	0.0	9.7	0.9	1.9
	Private financing	Mil. KRW	3300	0.0	8.2	0.4	1.5
	Private investments	Mil. KRW	3300	0.0	5.1	0.0	0.2
	Etc.	Mil. KRW	3300	0.0	6.1	0.0	0.2
Intellectual property rights in ownership	Domestic patent registrations	Cases	3300	0.0	600.0	4.2	14.7
	Domestic design registrations	Cases	3300	0.0	200.0	1.2	6.7
	Domestic IP registrations total	Cases	3300	0.0	1600.0	7.7	33.0
	Foreign patent registrations	Cases	3300	0.0	100.0	0.3	3.1
	Foreign design registrations	Cases	3300	0.0	80.0	0.1	1.7
	Foreign IP registrations total	Cases	3300	0.0	250.0	0.6	6.6
	Domestic patent applications	Cases	3300	0.0	73.0	0.7	3.1
	Foreign patent applications	Cases	3300	0.0	185.0	0.2	3.7
Number of R&D Project Performance	R&D project attempts	Cases	3300	0.0	400.0	4.1	10.5
	R&D project in progress	Cases	3300	0.0	80.0	1.6	2.9
	R&D project failure	Cases	3300	0.0	100.0	0.5	3.0
	R&D project success	Cases	3300	0.0	300.0	2.0	7.4
	Commercialization in progress	Cases	3300	0.0	50.0	0.6	1.8
	Commercialization failure	Cases	3300	0.0	100.0	0.1	1.9
	Commercialization success	Cases	3300	0.0	200.0	1.3	5.4
Technology Capabilities (Global comparison)	Product planning	%	3290	30.0	100.0	77.0	11.6
	Design	%	3281	30.0	100.0	76.0	12.7
	New technology development	%	3285	30.0	100.0	76.4	12.2
	Product design	%	3290	30.0	100.0	77.1	11.7
	Parts and process design	%	3267	30.0	100.0	76.6	12.1
	Testing inspection	%	3277	20.0	100.0	74.7	14.0
	Manufacturing and processing	%	3281	20.0	100.0	79.0	12.1
	Product management	%	3276	20.0	100.0	78.8	11.7
	Maintenance and repair	%	3278	20.0	100.0	78.3	11.7
	Commercialization	%	3292	25.0	100.0	76.6	11.8
Number of times test equipment was used	Professional testing institutions	Cases	3300	0.0	250.0	4.5	12.0
	Universities	Cases	3300	0.0	100.0	0.7	4.8
	Public research institutes	Cases	3300	0.0	100.0	0.6	3.7
	Local SME administrations	Cases	3300	0.0	70.0	0.3	2.1
	Large firms	Cases	3300	0.0	50.0	0.3	1.4
	SMEs	Cases	3300	0.0	100.0	0.2	2.6
Level of technology compared to previous year	Compared to 1 year ago	%	3300	40.0	200.0	110.5	15.8
	Compared to 3 years ago	%	3053	20.0	300.0	125.6	26.8
Technology development performance outcomes	IPR appiled	Cases	3300	0.0	15.0	0.2	0.9
	IPR registered	Cases	3300	0.0	10.0	0.1	0.7
Time required for each stage of R&D	Development planning	Months	3300	0.0	36.0	4.6	3.4
	Development progression (or implementation)	Months	3300	0.0	48.0	6.4	4.5
	Commercialization	Months	3300	0.0	48.0	5.8	4.0
	Total technology development	Months	3300	3.0	96.0	16.8	10.0
Time required: opening new sales markets		Months	3300	1.0	48.0	7.2	5.5
Sales (LN)		Mil. KRW	3300	5.5	12.2	8.7	1.3
Main product sales ratio		%	3300	5.0	100.0	84.3	20.1
New technology sales ratio		%	3300	0.0	100.0	20.1	21.0
Operating profit		Mil. KRW	3300	(26,067.0)	27,826.0	682.7	2090.4
Ordinary profit		Mil. KRW	3300	(53,110.0)	50,139.0	587.7	2649.9
Transaction client sales ratio	Large firms	%	3300	0.0	100.0	30.7	36.8
	SMEs	%	3300	0.0	100.0	41.6	38.4
	Consumers	%	3300	0.0	100.0	10.7	24.8
	Public	%	3300	0.0	100.0	9.5	23.8
	Exports	%	3299	0.0	100.0	7.5	18.8
	Overall area	%	3300	30.0	100.0	78.0	10.0
Year of establishment		Year	3300	1942.0	2013.0	2000.9	9.4
Age of CEO / representative	Year of CEO birth	Year	3300	1926.0	1987.0	1961.9	8.6
Total number of employees (LN)		Persons	3300	0.0	4.8	1.3	0.9
Total number of researchers (LN)		Persons	3300	1.6	5.7	3.3	1.1

Discrete Variable Name	Scale	N	Mode (1st)	Ratio of Mode (1st),%	Mode (2nd)	Ratio of Mode (2nd),%

(continued on next page)

(*continued*)

Discrete Variable Name	Scale	N	Mode (1st)	Ratio of Mode (1st),%	Mode (2nd)	Ratio of Mode (2nd),%
Experience of technological cooperation	Nominal	3300	Universities	11.1	PRI	7.7
Equipment availability	Ordinal	3300	50~75%	21.9	Less than 25%	21.8
Core Technology life cycle	Ordinal	3300	Maturity	51.1	Growth	39.9
Product life cycle	Ordinal	3300	5~7 years	18.9	2~3 years	16.2
Technological innovativeness	Ordinal	3300	Middle level	60.2	Low Level	25.9
Time required for imitation	Ordinal	3300	0.5~1 year	34.7	1~1.5 years	20.8
Production based areas	Nominal	3300	Machinery & Materials	35.6	Electrical & Electronics	16.8
Innovation SMEs status	Nominal	3300	No	65.5	Yes	34.5
Venture status	Nominal	3300	No	80.2	Yes	19.8
Growth stage	Nominal	3300	Growth	47.9	Maturity	45.3

Appendix Table B. Examples of development details of the Ensemble Model (CHAID & CART Ensemble model for PRI)

Model	CHIAD Accuracy	No. of Variables	Nodes(Model Size)	CART Accuracy	No. of Variables	Nodes(Model Size)
1	92.1%	17	44	88.4%	32	15
2	85.7%	21	67	73.3%	34	19
3	84.2%	26	80	86.0%	26	13
4	76.4%	21	84	77.9%	23	15
5	87.7%	26	76	79.1%	51	35
6	85.8%	25	77	64.6%	34	23
7	68.9%	24	87	68.0%	58	33
8	88.9%	26	80	70.6%	53	37
9	78.3%	22	69	67.8%	49	31
10	74.6%	31	87	82.5%	47	27

References

Arranz, N., de Arroyabe, J.C.F., 2008. The choice of partners in R&D cooperation: an empirical analysis of Spanish firms. Technovation 28, 88–100.

Belkin, N.J., Croft, W.B., 1992. Information filtering and information retrieval: two sides of the same coin. Commun. ACM. Citeseer.

Bellucci, A., Pennacchio, L., Zazzaro, A., 2019. Public R&D subsidies: collaborative versus individual place-based programs for SMEs. Small Bus. Econ. 52, 213–240.

Breiman, L., Friedman, J., Olshen, R., Stone, C.J., 1984. Classification and regression trees.

Brockhoff, K., Teichert, T., 1995. Cooperative R&D and partners' measures of success. Int. J. Technol. Manag. 10, 111–123.

Calcagnini, G., Giombini, G., Liberati, P., Travaglini, G., 2019. Technology transfer with search intensity and project advertising. J. Technol. Transf. 44, 1529–1546.

Caloffi, A., Mariani, M., Rossi, F., Russo, M., 2018. A comparative evaluation of regional subsidies for collaborative and individual R&D in small and medium-sized enterprises. Res. Policy 47, 1437–1447.

Caloghirou, Y., Giotopoulos, I., Kontolaimou, A., Korra, E., Tsakanikas, A., 2021. Industry-university knowledge flows and product innovation: how do knowledge stocks and crisis matter? Res. Policy 50, 104195.

Chapman, G., Lucena, A., Afcha, S., 2018. R&D subsidies & external collaborative breadth: differential gains and the role of collaboration experience. Res. Policy 47, 623–636.

Chatterji, D., 1996. Accessing external sources of technology. Res.-Technol. Manag. 39, 48–56.

Chesbrough, H.W., 2006. Open innovation: The new Imperative For Creating and Profiting from Technology. Harvard Business Press.

Choi, J., Kang, H., Kim, E., Lee, S., Han, S., Kim, M., 2002. Prediction and Excess of Data Mining Using Decision Tree Analysis. SPSS Academy, Seoul.

Das, T.K., Teng, B.-.S., 2000. A resource-based theory of strategic alliances. J. Manage. 26, 31–61.

Davenport, S., Miller, A., 2000. The formation and evolution of international research alliances in emergent technologies: research issues. J. High Technol. Manag. Res. 11, 199–213.

Dong, L., Glaister, K.W., 2006. Motives and partner selection criteria in international strategic alliances: perspectives of Chinese firms. Int. Bus. Rev. 15, 577–600.

Dyer, J.H., Powell, B.C., Sakakibara, M., Wang, A.J., 2007. THE DETERMINANTS OF SUCCESS IN R&D ALLIANCES, Academy of Management Proceedings. Academy of Management Briarcliff Manor, NY 10510, pp. 1–6.

Eom, B.-.Y., Lee, K., 2010. Determinants of industry–academy linkages and, their impact on firm performance: the case of Korea as a latecomer in knowledge industrialization. Res. Policy 39, 625–639.

Fernández-Olmos, M., Ramírez-Alesón, M., 2017. How internal and external factors influence the dynamics of SME technology collaboration networks over time. Technovation 64, 16–27.

Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning. Springer series in statistics New York, NY, USA.

Fritsch, M., Lukas, R., 2001. Who cooperates on R&D? Res. Policy 30, 297–312.

Fukugawa, N., 2016. Knowledge spillover from university research before the national innovation system reform in Japan: localisation, mechanisms, and intermediaries. Asian J. Technol. Innovat. 24, 100–122.

Galende, J., 2006. Analysis of technological innovation from business economics and management. Technovation 26, 300–311.

Gkypali, A., Filiou, D., Tsekouras, K., 2017. R&D collaborations: is diversity enhancing innovation performance? Technol. Forecast. Soc. Change 118, 143–152.

Greco, M., Grimaldi, M., Cricelli, L., 2020. Interorganizational collaboration strategies and innovation abandonment: the more the merrier? Ind. Mark. Manag. 90, 679–692.

Gulati, R., Singh, H., 1998. The architecture of cooperation: managing coordination costs and appropriation concerns in strategic alliances. Adm. Sci. Q. 781–814.

Hagedoorn, J., 1993. Understanding the rationale of strategic technology partnering: nterorganizational modes of cooperation and sectoral differences. Strategic Manag. J. 14, 371–385.

Hagedoorn, J., 2006. Understanding the cross-level embeddedness of interfirm partnership formation. Acad. Manag. Rev. 31, 670–680.

Hagedoorn, J., Lokshin, B., Zobel, A.K., 2018. Partner type diversity in alliance portfolios: multiple dimensions, boundary conditions and firm innovation performance. J. Manag. Stud. 55, 809–836.

Hagedoorn, J., Schakenraad, J., 1994. The effect of strategic technology alliances on company performance. Strategic Manag. J. 15, 291–309.

Hagedoorn, J., Van Kranenburg, H., 2003. Growth patterns in R&D partnerships: an exploratory statistical study. Int. J. Ind. Org. 21, 517–531.

Hausman, W.H., Montgomery, D.B., Roth, A.V., 2002. Why should marketing and manufacturing work together?: some exploratory empirical results. J. Oper. Manag. 20, 241–257.

Hottenrott, H., Lopes-Bento, C., 2016. R&D partnerships and innovation performance: can there be too much of a good thing? J. Prod. Innovat. Manag. 33, 773–794.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning. Springer.

Jeon, H., Seo, W., Park, E., Choi, S., 2020. Hybrid machine learning approach for popularity prediction of newly released contents of online video streaming services. Technol. Forecast. Soc. Change 161, 120303.

Jeong, H., Kim, Y.-.K., Kim, J., 2016. An evaluation-committee recommendation system for national R&D projects using social network analysis. Cluster. Comput. 19, 921–930.

Jun, S.-.P., Kim, S.-.G., Park, H.-.W., 2017. The mismatch between demand and beneficiaries of R&D support programs for SMEs: evidence from Korean R&D planning programs. Technol. Forecast. Soc. Change 116, 286–298.

Jun, S.-.P., Lee, J.-.S., Lee, J., 2020. Method of improving the performance of public-private innovation networks by linking heterogeneous DBs: prediction using ensemble and PPDM models. Technol. Forecast. Soc. Change 161, 120258.

Kafouros, M., Love, J.H., Ganotakis, P., Konara, P., 2020. Experience in R&D collaborations, innovative performance and the moderating effect of different dimensions of absorptive capacity. Technol. Forecast. Soc. Change 150, 119757.

Kale, P., Singh, H., 2009. Managing strategic alliances: what do we know now, and where do we go from here? Acad. Manag. Perspect. 45–62.

Kim, H.-J., Kim, B.-K., 2013. A Study on the impact of technological alliance motivation on the partner selection criteria, governance structure and performance of Korean High-Tech Firms' technological alliances: a focus on the integration of TCE and RBV. J. Technol. Innovat. 21.

Kim, S.-w., Yang, H.-c., 2017. Small Business R&D Policy Special Series: sMEs' cooperative R&D performance status and network effect. STEPI Insight 1–35.

Konsti-Laakso, S., Pihkala, T., Kraus, S., 2012. Facilitating SME innovation capability through business networking. Creat. Innovat. Manag. 21, 93–105.

Lai, W.-.H., Chang, P.-.L., 2010. Corporate motivation and performance in R&D alliances. J. Bus. Res. 63, 490–496.

Leischnig, A., Geigenmüller, A., 2020. Examining alliance management capabilities in university-industry collaboration. J. Technol. Transf. 45, 9–30.

Lemmens, C.E., 2004. Innovation in Technology Alliance Networks. Edward Elgar Publishing.

Lu, J., Shambour, Q., Xu, Y., Lin, Q., Zhang, G., 2013. a web-based personalized business partner recommendation system using fuzzy semantic techniques. Comput. Intell. 29, 37–69.

Malerba, F., 2002. Sectoral systems of innovation and production. Res. Policy 31, 247–264.

Markovic, S., Jovanovic, M., Bagherzadeh, M., Sancha, C., Sarafinovska, M., Qiu, Y., 2020. Priorities when selecting business partners for service innovation: the contingency role of product innovation. Ind. Mark. Manag. 88, 378–388.

Mora-Valentin, E.M., Montoro-Sanchez, A., Guerras-Martin, L.A., 2004. Determining factors in the success of R&D cooperative agreements between firms and research organizations. Res. Policy 33, 17–40.

MSS, Kbiz, 2017. 10th Survey On Technology of Small & Medium Enterprises. Ministry of SMEs and Startups & Korea Federation of SMEs (Kbiz), Seoul.

MSS, Kbiz, 2018. 11th Small and Medium Enterprises Technology Statistics Survey. Ministry of SMEs and Startups & Korea Federation of SMEs (Kbiz), Seoul.

Nielsen, B.B., 2003. An empirical investigation of the drivers of international strategic alliance formation. Eur. Manag. J. 21, 301–322.

Park, J., Kim, J., Woo, H., Yang, J.-.S., 2020. Opposite effects of R&D cooperation on financial and technological performance in SMEs. J. Small Bus. Manag. 1–34.

Park, S., Kim, D.-.Y., 2017. Assessing language discrepancies between travelers and online travel recommendation systems: application of the Jaccard distance score to web data mining. Technol. Forecast. Soc. Change 123, 381–388.

Prahalad, C.K., Ramaswamy, V., 2000. Co-opting customer competence. Harv. Bus. Rev. 78, 79–90.

Sarpong, O., Teirlinck, P., 2018. The influence of functional and geographical diversity in collaboration on product innovation performance in SMEs. J. Technol. Transf. 43, 1667–1695.

Schwab, K., 2017. The fourth industrial revolution. Currency.

Shmueli, G., Patel, N.R., Bruce, P.C., 2011. Data Mining For Business intelligence: Concepts, techniques, and Applications in Microsoft Office Excel with XLMiner. John Wiley and Sons.

SMBA, 2020. Survey on the status of SMEs. STATISTICS KOREA, Korean Statstical Information Service (KOSIS).

SMBA, Kbiz, 2009. 4th Survey on Technology of Small & Medium Enterprises. Small and Medium Business Administration & Korea Federation of SMEs, Seoul.

SMBA, Kbiz, 2011. 5th Survey on Technology of Small & Medium Enterprises. Small and Medium Business Administration & Korea Federation of SMEs, Seoul.

SMBA, Kbiz, 2013. 6th Survey on Technology of Small & Medium Enterprises. Small and Medium Business Administration & Korea Federation of SMEs, Seoul.

SMBA, Kbiz, 2014. 7th Survey on Technology of Small & Medium Enterprises. Small and Medium Business Administration & Korea Federation of SMEs, Seoul.

SMBA, Kbiz, 2015. 8th Survey on Technology of Small & Medium Enterprises. Small and Medium Business Administration & Korea Federation of SMEs, Seoul.

SMBA, Kbiz, 2016. 9th Survey on Technology of Small & Medium Enterprises. Small and Medium Business Administration & Korea Federation of SMEs, Seoul.

Souder, W.E., Buisson, D., Garrett, T., 1997. Success through customer-driven new product development: a comparison of US and New Zealand small entrepreneurial high technology firms. J. Prod. Innovat. Manag. 14, 459–472.

Tang, J., Wu, S., Sun, J., Su, H., 2012. Cross-domain collaboration recommendation. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1285–1293. ACM.

Un, C.A., Asakawa, K., 2015. Types of R&D collaborations and process innovation: the benefit of collaborating upstream in the knowledge chain. J. Prod. Innovat. Manag. 32, 138–153.

Wang, Q., Ma, J., Liao, X., Du, W., 2017. A context-aware researcher recommendation system for university-industry collaboration on R&D projects. Decis. Support Syst. 103, 46–57.

Williamson, O.E., 1981. The economics of organization: the transaction cost approach. Am. J. Sociol. 87, 548–577.

Williamson, O.E., 1991. Comparative economic organization: the analysis of discrete structural alternatives. Adm. Sci. Q. 269–296.

Wu, S., Sun, J., Tang, J., 2013. Patent partner recommendation in enterprise social networks. In: Proceedings of the sixth ACM international conference on Web search and data mining, pp. 43–52.

Xu, L., Li, J., Zhou, X., 2019. Exploring new knowledge through research collaboration: the moderation of the global and local cohesion of knowledge networks. J. Technol. Transf. 44, 822–849.

Xu, W., Sun, J., Ma, J., Du, W., 2016. A personalized information recommendation system for R&D project opportunity finding in big data contexts. J. Netw. Comput. Appl. 59, 362–369.

Yoo, H.S., Lee, C., Jun, S.-.P., 2018. The characteristics of SMEs preferring cooperative research and development support from the government: the case of Korea. Sustainability 10, 3048.

Yuan, S.T., Hsieh, P.H., Yeh, Y.-.C., 2015. A Computing Metaphor Approach to Customer Experience Based Alliance Partner Recommendation. Kybernetes.

Zacharias, N.A., Daldere, D., Winter, C.G., 2020. Variety is the spice of life: how much partner alignment is preferable in open innovation activities to enhance firms' adaptiveness and innovation success? J. Bus. Res. 117, 290–301.

Seung-Pyo Jun is a principal researcher of Data Analysis Platform Center at the Korea Institute of Science and Technology Information (KISTI). Also, he is a professor of Science and Technology Policy in University of Science & Technology (UST). He received his Ph.D. from the Science and Technology Studies Program at Korea University. His-area of interest includes the demand forecasting, emerging technology detecting and new technology adoption analysis. Recently he is focusing on the science technology and SMES supporting policy using data intelligence.

Hyoung Sun Yoo is a principal researcher at the Korea Institute of Science and Technology Information (KISTI). Also, he is an assistant professor of Science and Technology Policy in University of Science & Technology (UST). His-area of interest includes demand forecasting, science and technology policy, complex network, and agent-based modeling.

Jeena Hwang is a researcher of Technology Commercialization Center at the Korea Institute of Science and Technology Information (KISTI). Her area of interest includes demand forecasting, innovation policy for SMEs, and R&D collaboration management using big-data.