Identifying New Product Ideas:

Waiting for the Wisdom of the Crowd or Screening Ideas in Real Time

Steven Hoornaert[a], Michel Ballings[b], Edward C. Malthouse[c], Dirk Van den Poel[a]

[a]Ghent University, Department of Marketing, Tweekerkenstraat 2, 9000 Ghent, Belgium, steven.hoornaert@ugent.be, dirk.vandenpoel@ugent.be
[b]The University of Tennessee, Department of Business Analytics and Statistics, 249 Stokely Management Center, 916 Volunteer Blvd, Knoxville, TN, 37996, michel.ballings@utk.edu, Corresponding author
[c]Northwestern University, 1845 Sheridan Road, Evanston, IL, 60208, ecm@northwestern.edu

Abstract

Crowdsourcing ideas from consumers can enrich idea input in new product development. After a decade of initiatives (e.g., Starbucks' MyStarbucksIdea, Dell's IdeaStorm), the implications of crowdsourcing for idea generation are well understood, but challenges remain in dealing with the large volume of rapidly-generated ideas produced in crowdsourcing communities. This study proposes a model that can assist managers in efficiently processing crowdsourced ideas by identifying the aspects of ideas that are most predictive of future implementation and identifies three sources of information available for an idea: its *content*, the *contributor* proposing it, and the *crowd*'s feedback on the idea (the "3Cs"). These information sources differ in their time of availability (content/contributor information is available immediately; crowd feedback accumulates over time) and in the extent to which they comprise structured or unstructured data. This study draws from prior research to operationalize variables corresponding to the 3Cs and develops a new measure to quantify an idea's distinctiveness. Applying automated information retrieval methods (latent semantic indexing) and testing several linear methods (linear discriminant analysis, regularized logistic regression) and nonlinear machine-learning algorithms (stochastic adaptive boosting, random forests), this article identifies the variables that are most useful towards predicting idea implementation in a crowdsourcing community for an IT product (Mendeley). Our results indicate that consideration of content and contributor information improves ranking performance between 22.6% and 26.0% over random idea selection, and that adding crowd-related information further improves performance by up to 48.1%. Crowd feedback is the best predictor of idea implementation, followed by idea content and distinctiveness, and the contributor's past idea-generation experience. Firms are advised to implement two idea selection support systems: one to rank new ideas in real time based on content and contributor experience, and another that integrates the crowd's idea evaluation after it has had sufficient time to provide feedback.

Keywords: idea selection, crowdsourcing, idea selection support system, innovation, real-time analysis, machine learning

Practitioner Points

- When using automated idea screening that incorporates crowd feedback as an initial step for subsequent human evaluation, practitioners should utilize nonlinear machine learning algorithms because these outperform classical linear methods.
- Ranking ideas in real time is a viable option, but waiting for the wisdom of the crowd is desirable. Therefore, firms should implement two idea selection support systems: one real-time system that can immediately rank new ideas based on content and contributor experience; and an additional one that integrates the crowd's idea evaluation after sufficient time for feedback. The two systems are complementary and

can be used simultaneously.

- When ranking ideas in real time, managers should use classical statistical methods, because their performance is similar to more computational intensive machine learning methods.

Introduction

The new product development (NPD) process, in which firms develop new products and improve existing ones, is dependent on firms' capacity to produce and identify good ideas. Firms have traditionally relied on in-house development teams and marketing research for this purpose, but they can now use consumer feedback and discussion collected from digital, social and mobile environments. Such consumer data flow from a variety of sources and are generated in large volumes and at a fast rate (velocity). To organize such data in one place, many firms create designated crowdsourcing environments where consumers can discuss products, propose new ideas, and evaluate ideas proposed by other consumers. Such so-called "crowdsourcing communities" are popular. For example, Google received 150,000 proposals in its 2008 Project 10^100 contest—a general call for ideas to "help as many people as possible"—and devoted 3000 employees to processing them (Blohm, Leimeister, and Krcmar, 2013).

It is highly challenging for firms to process the large volumes of information that flow through crowdsourcing communities and to identify the (relatively few) ideas that are worth implementing. In particular, attempts to automate or formalize idea processing are complicated by the fact that new ideas and discussions of those ideas are usually provided as written text, which is a form of unstructured data that cannot be analyzed with classical multivariate analysis methods (Calantone, Di Benedetto, and Schmidt, 1999; Cormican and O'Sullivan, 2004). Yet some aspects of an idea may be relatively easy to analyze; for example, crowdsourcing community platforms are likely to have information on whether a given user has submitted ideas in the past and whether those ideas were successful, and these characteristics may be indicative of the likelihood that subsequent ideas will succeed as well. Another challenge that firms face is determining the timing at which ideas should be processed. Presumably, a firm benefits from identifying good ideas in real time, as soon as possible after they are proposed. Yet, after an idea is proposed, it takes time for the community to respond to it. This leads to the question—does the firm gain substantial benefit from waiting for the community's response? Or could it just as easily evaluate the worthiness of ideas on the basis of other, immediately-available information?

Herein, this study seeks to shed light on the manner in which a firm presented with a flow of incoming ideas in a crowdsourcing community might assess those ideas most efficiently. To this end, a model is proposed in which the information available on a given crowdsourced idea is divided into three categories—*content*, *contributor*, and *crowd*—which are called the *3Cs*. The *content* category refers to the idea itself, which is usually expressed in unstructured, written text. The *contributor* category refers to information about the person who contributed the idea, such as whether he or she submitted ideas or discussed other ideas in the past, and whether those previous ideas were adopted. The *crowd* category refers to the reactions of other consumers. Some of these reactions may be structured—e.g., votes or ratings—and some may be unstructured, i.e., textual comments. Whereas information in the *content* and *contributor* categories are available at the moment an idea is submitted, information in the *crowd* category requires time to accumulate and cannot be used in real-time evaluations.

An attempt is then made to evaluate the relative roles of the different information categories in determining an idea's potential for success. To this end, building on existing text-mining approaches and implementing a variety of machine learning algorithms, this study integrates information from each of the 3Cs to predict an idea's likelihood of being implemented. In capturing the content dimension of ideas, a method of quantifying the innovativeness of a given idea, based on *k*-means clustering of text-mined idea content, is proposed.

Applying our approach to a data set from a crowdsourcing community for a software product called Mendeley, the results indicate that consideration of (real-time) information about the content and the contributor significantly improves predictive capacity as compared with random idea selection, even using logistic regression models. However, the accuracy of classification can be improved substantially by incorporating crowd information. Crowd evaluation is most predictive of idea implementation, followed by the content of the idea and then the contributor. Overall, more recent machine learning models substantially outperform linear models. To our knowledge, our study is the first to evaluate the predictive capacity that can be gained from waiting for the feedback of the crowd as opposed to processing ideas in real time, and to benchmark different methods for screening ideas in real time. Table 1 summarizes our contribution relative to other research in this area.

Table 1: Literature on Idea Selection in Crowdsourcing Communities					
Study	**3Cs Studied**	**Comparison of different analysis methods**	**Comparison of structured vs. unstructured data**	**Insights regarding the use of crowdsourcing in the NPD process**	**Value of real-time processing vs. waiting for crowd feedback**
Our study	All three	Classical and machine learning methods benchmarked. Classical methods have similar performance as ML methods for contributor and crowd, but ML methods are better for all 3Cs	Crowd (structured) > content (structured) > contributor (unstructured)	Idea scoring: models as support for decision makers as an initial ranking step for subsequent human evaluation	Real-time ranking acceptable, but waiting for crowd improves ranking considerably
Di Gangi and Wasko (2009)	Content and Crowd	No benchmarking	No benchmarking	Understand the technical requirements of the ideas and contributor's needs in order to extract value from the idea.	Not studied
Bayus (2013)	Contributor	No benchmarking	No benchmarking	Identifying and studying high quality contributors (called ideators) as a means to maintain an ongoing supply of quality ideas over time	Not studied
Walter and Back (2013)	Content	No benchmarking	No benchmarking	Idea scoring: models as support for decision makers as an initial ranking step for subsequent human evaluation	Not studied
Westerski, Dalamagas, and Iglesias (2013)	Content	No benchmarking	No benchmarking	Idea annotation for distinctive features in order to support subsequent human evaluation	Not studied
Klein and Garcia (2015)	Crowd	No benchmarking	No benchmarking	Idea scoring: models as support for decision makers as an initial ranking step for subsequent human evaluation	Not studied

Nagar, De Boer, and Garcia (2016)	All three	No benchmarking	Crowd > Content > Contributor	Idea scoring: models as support for decision makers as an initial ranking step for subsequent human evaluation	Not studied
Toubia and Netzer (2017)	Content	No benchmarking	No benchmarking	Idea scoring: models as support for decision makers as an initial ranking step for subsequent human evaluation. Recommend words to users to improve ideas	Not studied
Rhyn and Blohm (2017)	Content	No benchmarking	No benchmarking	Idea scoring: models as support for decision makers as an initial ranking step for subsequent human evaluation	Not studied

Literature

The role of consumer input in the NPD process

Product innovation is generally conceptualized as a five-stage NPD process consisting of idea generation and selection, concept development, product design, product testing, and product introduction (Urban and Hauser, 1993). As an idea for a new product passes through the NPD process, perceived risk, investments, and development time increase. Therefore, firms have been focused on integrating the customer’s perspective as a means of expediting this process and reducing the overall NPD cycle time (Alam, 2006), cost, and need for future product modifications. Historically, traditional marketing research techniques such as focus groups, surveys, prototyping, product testing, and test marketing have been used to gather consumer input (Urban and Hauser, 1993). In recent decades, the Internet has enabled larger numbers of consumers to participate in product innovation, while facilitating a more cost-effective, richer, and recurrent dialogue between the firm and its customers (Sawhney, Verona, and Prandelli, 2005). In this way, the firm benefits from consumer knowledge and expertise as additional input to its value-creation process, while consumers experience more value (Bharadwaj and Dong, 2014), are more loyal through increased perceptions of quality (Bharadwaj, Nevin, and Wallman, 2012), feel involved and acknowledged (Fuchs and Schreier, 2011), and get a sense of belonging and empowerment (Saarijärvi, Kannan, and Kuusela, 2013). Meanwhile, both benefit from better products.

In the last decade, firms have been experimenting with alternative ways to access consumer knowledge, of which crowdsourcing is a prime example. Over the years, different researchers have conceptualized crowdsourcing in various ways (see Tarrell et al., 2013;

Lang, Bharadwaj, and Di Benedetto, 2016). This study refers to the original and most dominant definition of this concept in literature: crowdsourcing *is* a firm's outsourcing of a task that was once performed by an employee, to a large, undefined group of people outside the firm in the form of an open call (Howe 2006, 2008). This construct covers a multitude of efforts, such as Amazon Mechanical Turk, IdeaStorm, Innocentive, and Threadless.

In the context of innovation, crowdsourcing can take the form of a one-time event to collect ideas (e.g., Cisco's I-Prize) or a continuous process in which idea generation and evaluation happen concurrently (Bayus, 2013). As noted above, platforms in which the latter process takes place are often called crowdsourcing or (user) innovation communities (Bayus, 2013; Di Gangi and Wasko, 2009; Westerski, Dalamagas, and Iglesias, 2013). Members of these communities can share suggestions for improvements on a company's products or services, or publicly express their opinions by voting, commenting, rating, ranking, or buying idea stocks (Klein and Garcia, 2015). When the firm decides to dismiss an idea or, conversely, to advance it to development, it typically notifies the community of the idea's new status, e.g., by posting a comment or a blog post. Examples of popular crowdsourcing communities are Dell's IdeaStorm, Starbucks' MyStarbucksIdea, and IBM's Innovation Jam.

Crowdsourcing communities are popular and can generate valuable ideas. Yet the number of ideas is large, and only a few are likely to be useful to the firm. For example, 25,186 ideas have been submitted to IdeaStorm, but only 549 (~2.2%) have been implemented since the platform's launch in 2007 (IdeaStorm, 2016). MyStarbucksIdea has collected 162,156 ideas, of which only 320 (~0.2%) have been implemented since the community's launch in 2008 (Hossain and Islam, 2015; MyStarbucksIdea, 2015). Identifying the ‘needles in the haystack’ can be a challenge. This challenge is further complicated by the common problem of duplicate ideas (Di Gangi, Wasko, and Hooker, 2010), which consume time and resources without enriching idea input.

This study proposes a method to automate idea implementation decisions that can deal with these challenges. Our method ranks ideas based on their probability for implementation using a range of multifaceted metrics identified by prior research. In particular, this study focuses on evaluating the relative contributions of the three information sources elaborated above—the 3Cs—towards predicting an idea’s likelihood of being implemented. Our findings may enable firms to derive value from crowdsourcing more efficiently (i.e., reduce evaluation time) and cost-effectively (i.e., reduce evaluation cost)[1].

The following paragraphs draw from literature on creativity research, applied psychology and product innovation research to provide a discussion of what can be measured about each of the 3Cs and survey literature relevant to the use of each information source to select high-quality ideas. Table 1 summarizes the studies discussed and highlights our incremental contribution, showing that this study is the first to integrate the 3Cs in one idea selection model, thereby providing insight into the value of real-time idea processing and of analysis of structured vs. unstructured data, which are questions identified in the call for manuscripts for this special issue (Bharadwaj and Noble, 2015).

The 3Cs

Content-based idea selection

The content dimension of an idea refers to the text written by the contributor to describe the idea. It can include additional media such as an image or video. Two broad types of evaluations have been discussed in the literature to detect the quality of the content of an idea: *human* evaluations and *text mining*. The first involves staff, experts or consumers formally scoring an idea against predefined decision criteria (Carbonell-Foulquié, Munuera-Alemán, and Rodríguez-Escudero, 2004), informally evaluating an idea relying on a 'gut' feeling by drawing from prior experience (Magnusson, Netz, and Wästlund, 2014), or a combination of both (Eling, Langerak, and Griffin, 2015).

Humans can interpret ideas beyond word use or grammatical structure, and, given that submitted ideas are often vague (Sternberg and Lubart, 1999), undeveloped (Jouret, 2009) and immature (Di Gangi, Wasko, and Hooker, 2010; Magnusson, 2009), this capacity offers an advantage over automated approaches. Human evaluations, however, require complex cognitive effort and are time-intensive and costly. Furthermore, human evaluators can suffer from fatigue or loss of focus. In general, performance of the idea-content evaluation depends on the instructions given to evaluators, the possibility of collaboration among them, and their level of expertise (Jouret, 2009; Magnusson, Wästlund, and Netz, 2014; Onarheim and Christensen, 2012; Rietzschel, Nijstad, and Stroebe, 2010). Delegating idea-content evaluation to lower-level employees or external parties (e.g., Amazon Mechanical Turk) may be more cost efficient than relying on in-house expert evaluators, but increases the odds of missing out on opportunities (Eling, Langerak, and Griffin, 2015). The use of multiple raters can mitigate this risk by bringing in different views of the same idea, but reconciling those views can be challenging (Moreau, Lehmann, and Markman, 2001).

Text-mining-based methods of idea content evaluation are promising because they can be used to process large amounts of unstructured content in real time. Recent research has found that text-mining-based analysis of *raw* product ideas captured directly from contributors can provide reliable predictions of the proposed products' commercial success and consumers' purchase intent, without intervention of human evaluators (Kornish and Ulrich, 2014). These observations motivate the use of more text mining in innovation.

Initial studies on text mining and innovation used unsupervised techniques (e.g., cluster analysis) for detecting new ideas in patent texts (e.g., Thorleuchter, Van den Poel, and Prinzie, 2010; see Christensen et al., 2016 for a review). Later on, other studies used supervised techniques (e.g., regression) to investigate the capacity of text mining to provide information on idea-content quality in crowdsourcing communities (Walter and Back, 2013; Westerski, Dalamagas, and Iglesias, 2013). More specifically, Westerski, Dalamagas, and Iglesias (2013) looked at word dissimilarity between ideas, and Walter and Back (2013) looked at an idea's set of unique words, to investigate how text mining features can predict human idea evaluation decisions. This study extends the research method of Walter and Back (2013) by performing latent semantic indexing (LSI) before applying *k*-means clustering to quantify an idea's level of novelty, which is subsequently used to predict the idea's likelihood of implementation by the firm.

Contributor-based idea selection

When a crowdsourcing community member contributes an idea, the firm is likely to have access to data about him or her that can be informative regarding the value of the proposed idea. For example, the idea contributor, also called the ideator, may have a history of contributing ideas. A contributor's record of previous suggestions may be predictive because of the notion that high-quality ideas come from new and original arrangements of a person's existing knowledge (Dahl and Moreau, 2002). People with diverse expertise pool their knowledge to come up with ideas that are more innovative than users who have narrower, domain-relevant skills (Amabile, 1983). The more competence and experience contributors possess, the higher the expected quality of their solutions (Magee, 2005; Poetz and Schreier, 2012). Correspondingly, a history of successful ideas may be indicative of a contributor who possesses expertise, suggesting a likelihood of proposing additional successful ideas in the future. However, applied psychological research has noted how prior knowledge can impede idea generation. People with knowledge in a given domain may *fixate* on prior examples and consequently neglect to explore the entire solution space, which results in less original and valuable ideas (Dahl and Moreau, 2002). In other words, people with a history of contributing ideas may produce subsequent ideas that are simply adjustments of previous ones, adapted according to prior experiences of success or failure (Marsh, Landau, and Hicks, 1996). This pattern of idea production tends to occur in settings where users need to suggest ideas on topics they are familiar with or on features that are commonly known (Jansson and Smith, 1991; Perttula and Sipilä, 2007; Purcell and Gero, 1996). Even though fixation is likely to occur in crowdsourcing communities, Bayus (2013) found that previous experience promotes rather than limits creativity. Yet, the same study found that, after a contributor's idea has been implemented, his or her subsequent ideas become less diverse. Thus, there is theoretical support for a contributor's number of previous contributions and/or previous implementations having either a positive or negative effect on the likelihood of future implementation.

Another potentially informative aspect of the contributor is the extent to which he or she has discussed other ideas in the community. As proposed by Osborn (1953) more than half a century ago, such discussion causes participants to revise their knowledge and to refine their own ideas (Kohn, Paulus, and Choi, 2011). Furthermore, solving consumer problems encourages members to stay active in the community (Lu, Singh, and Srinivasan, 2011). In addition, members who participate in discussions perceive more benefits, tend to feel a greater sense of community membership, and take their contributions more seriously (Preece, Nonnecke, and Andrews, 2004). These observations suggest that a tendency to engage in discussion might have a positive effect on future idea implementation.

Crowd-based idea selection

In crowdsourcing communities, members express their opinions of an idea by voting, commenting, rating, ranking, or buying idea stocks in prediction markets (Klein and Garcia, 2015). Different communities implement different systems: For example, IdeaStorm and MyStarbucksIdea use a system of up- and downvoting. Other communities, like Mendeley, give each user a budget of 10 votes to spend on ideas. Once an idea is implemented or declined, votes are returned. Commenting is unlimited and allows members to exchange thoughts on idea suggestions, and to further refine and develop ideas with the community. Only one study (Di Gangi and Wasko, 2009) investigated voting and commenting behavior in crowdsourcing communities and did not find a significant effect of either one on the likelihood of idea implementation. Zhu and He (2002), however, found that social context significantly influences implementation.

Several factors suggest that crowdsourcing community members can be a valuable resource in the idea selection process, potentially even outperforming experts or employees, in selecting ideas. A crowdsourcing community is likely to be made up of individuals who use the firm's products, and users often know their own needs and wants best and are in a premier position to understand how a product or service might create value for them (Magnusson, Wästlund, and Netz, 2014b). In particular, leading-edge users can detect product requirements earlier than the more occasional user because of their more advanced knowledge (Franke and Shah, 2003; von Hippel, 2005). Consequently, they can be powerful allies in idea selection (Pitta and Fowler, 2005). Surowiecki (2004) also refers to the value of the 'wisdom of the crowd' in idea selection, proposing that large numbers of individuals can make decisions that are superior to those of experts. However, his thesis refers to *independent* individuals, whereas in crowdsourcing communities users view and are influenced by the activity of others. Therefore, it is not clear whether Surowiecki's concept of the wisdom of the crowd is applicable to this context.

Research has shown that people perform better individually than collectively at generating ideas, yet no such pattern has been observed in the idea selection process (Faure, 2004; Rietzschel, Nijstad, and Stroebe, 2006). Studies attempting to explore this issue exposed a general problem with crowd-based idea selection, which is the misalignment of decision criteria. The crowd assesses ideas based on its own criteria (e.g., feasibility, desirability) or out of self-interest rather than the ideas' potential value to the firm (Rietzschel, Nijstad, and Stroebe, 2010). Providing participants with predefined decision criteria can mitigate this problem but increases time and cognitive complexity for the crowd (Riedl et al., 2010, 2013); likewise, it may disclose firm-sensitive information on innovation strategy.

In sum, it is unclear whether and to what extent data regarding the crowd's response to an idea is likely to be informative regarding an idea's probability of implementation.

Methodology

Sample

Data were obtained from the Mendeley crowdsourcing community, which has been in operation since 2008; these data are publicly available. This study uses the Mendeley community because its functionalities (i.e., idea submission, idea evaluation and user collaboration) are similar to those of previously-researched communities (e.g., IdeaStorm, MyStarbucksIdea) (Hrastinski et al., 2010). In addition, it contains sufficient data for our analysis methods (n =7,046 ideas). Mendeley, an Elsevier-owned product, offers users—predominantly scholars and college students—a library manager to collect and annotate reference material (e.g., articles, book chapters), and integrate citations in word processing systems. In addition, users can communicate, share resources, or collaborate on projects with other members. The company implements a freemium pricing strategy in which it offers a free version with 2GB of library space. Additional storage space can be purchased for a monthly fee.

The firm engages with its community to improve its software using a feedback forum built on the UserVoice platform. This forum (feedback.mendeley.com) enables users to suggest improvements for future releases. Upon registration, users can contribute their own ideas, and comment or vote on ideas from other members. No compensation or rewards are offered. Mendeley employees can provide feedback to the community by commenting on ideas and by assigning a status to each idea reflecting its stage in the consideration process.

The status options are *implemented, declined, under review, started* or *planned*. The latter three categories (*under review, started* or *planned*) are intermediate statuses, whereas the first two (*implemented, declined*) are more definitive. *Started* ideas refer to those that the firm began work on but were not yet completed and *implemented* ideas are ones that were completed and integrated into their software product. In our study, the focus is on ideas that were either implemented or declined and do not investigate the intermediate idea statuses further, given their limited use (2.59% of the total number of ideas in our sample)[2]. It should be noted that our focus on ideas that were actually implemented is in line with previous studies (Bayus, 2013; Di Gangi and Wasko, 2009; Hossain and Islam, 2015); this approach ensures that the ideas analyzed have a certain degree of value and usefulness (Franke, von Hippel, and Schreier, 2006; Levitt, 1963).

Data on 7,546 ideas were collected, comprising all ideas contributed between the launch of Mendeley's feedback forum on November 27, 2008 and December 5, 2014. After deleting ideas without a valid publication date (time stamp) and those whose status was not *implemented* or *declined*, 7,127 ideas remained. Some of those ideas (n = 81, 1.1%) had been contributed by the firm itself. Ultimately, the firm decides which ideas it implements; so for its own suggested ideas, it is both judge and jury. Therefore, ideas authored by the firm were dropped from the analysis. This resulted in a final sample of 7,046 ideas. These ideas were posted by 5,555 unique contributors during the observed period. The majority of those contributors (83.84%) posted one idea, 10.92% posted two ideas, and the remaining 5.24% posted three ideas or more.

Model/Variables

Idea implementation (DV)

During the observed period, 630 ideas (~9% of the final sample) were implemented, and 6,416 were declined. The dependent variable in our research, *idea implementation* (y_i), equals 1 if idea i was implemented and 0 if it was declined. Table 2 offers summary statistics and an overview of the variables included in our model. Ideas that are not considered relevant by the firm tend to linger in the forum and continue to receive votes and comments from the crowd. Eventually, Mendeley can decide to mark an idea as dead by assigning a *declined* status. This occurs, on average, after 3 years and 47 days (M = 1,142 days, SD = 507 days). Both the company and the community scan for duplicate ideas. This practice is reflected, for example, in the comment of a Mendeley employee on an idea suggested on November 8, 2012: *"... various duplicate tickets have been merged to increase its priority (thanks to those who pointed them out)."* As a result, duplication is limited among the ideas circulated in the Mendeley community, with only 15 detected duplicate ideas since the community's launch.

Idea content

Content is analyzed following a standard text mining procedure (Feldman and Sanger, 2007; Walter and Back, 2013). First, spaces, punctuation, stop words and numbers are removed from an idea's title and description. Next, sentences are broken down into terms (individual words) and each term is stemmed (e.g., the terms "usage", "using", and "used" are transformed into the root form "use"). Third, term occurrence for each idea is calculated, resulting in a term-frequency (tf) matrix. Next, sparse terms that occur in less than .5% of the idea corpus are deleted, and terms are inversely weighted by their relative occurrence in the idea corpus, producing a term frequency-inverse document frequency (tf-idf) matrix. As a last step, LSI is performed to scale down the terms to their underlying semantic meaning (Deerwester et al., 1990). LSI works well for term reduction and is good at grouping terms with similar (synonymy) or several (polysemy) meanings (Deerwester et al., 1990). Following Coussement and Van den Poel (2008), the number of LSI concepts (11 in our study) was determined using cross-validation.

In addition to using LSI concepts, a new way of characterizing the novelty of an idea is developed. A highly innovative idea contains new thoughts and concepts that are fundamentally different from previous ideas. In our study, an idea's relative novelty (also referred to here as 'distinctiveness') is operationalized by measuring how (dis)similar that idea is to past idea submissions. This operationalization ties into operationalizations used in previous text mining studies in innovation (e.g., Walter and Back, 2013). Using *k*-means clustering, three clusters of ideas are identified in the training set using the 11 LSI idea concepts. Next, ideas from the validation and test set are assigned to their closest clusters. Finally, an idea's score on distinctiveness is calculated as the number of prior ideas in the cluster to which the idea was assigned, divided by the number of prior ideas across all idea clusters. High values (e.g., >70% similarity to previous ideas) indicate that the idea is very similar to past ideas resulting in a low score on relative distinctiveness. Conversely, low values signal a highly dissimilar idea, resulting in a high score on relative distinctiveness.

Contributor experience

Contributor experience is operationalized similarly to Bayus (2013)[3]. More specifically, contributor experience is measured as the *number of past implemented ideas*, *number of previous contributor comments*, *number of past ideas*, and *tenure*. These variables are log-transformed to alleviate skewness and to allow for non-linear effects.

Number of past implemented ideas measures the success of a contributor in generating ideas that were implemented by the firm before submitting idea *i*. Since the start of the forum, 5,555 users contributed ideas. Approximately 10.3% (n = 553) had at least one idea implemented, of which 7.41% (n = 41) got two ideas implemented, 2.35% (n = 13) three ideas, and 0.54% (n = 3) four or five. Of the contributors who suggested more than one idea, 13.95% had more than half of their submitted ideas implemented. Compared with first-time contributors, contributors whose previous ideas were implemented more than double their odds (×2.09) of getting new ideas implemented ($\chi^2(1, n = 7{,}046) = 31.27$, $p<.001$) and receive 95.74% more crowd comments on their ideas (comments: $t(357.16) = 0.91$, $p = .026$). However, they have to wait 478 days longer for a decision ($W = 1{,}416{,}900$, $p < .001$). In sum, the firm seems to favor ideas from contributors who have created ideas in the past, yet there is no indication that their ideas are prioritized strategically over ideas from new contributors.

Number of previous contributor comments measures how many comments an idea contributor posted before submitting idea *i*. Compared with contributors who do not participate in conversations on other members' ideas, those who do participate prior to submitting their own ideas are 2.61 times more likely to have their ideas implemented ($\chi^2(1, n = 7046) = 59.24$, $p < .001$), receive 3.67 times more crowd comments ($t(301.88) = 4.64$, $p < .001$), but have to wait 165 days longer for their ideas to be implemented ($W = 1{,}053{,}300$, $p = .004$).

In addition to these variables, also included are the *number of past ideas*, the number of ideas the contributor posted before submitting idea *i*; *tenure*, the number of days a member has been in the crowdsourcing community; and *month* to account for seasonal effects.

Crowd feedback

In Mendeley's community, voting is allowed until the firm makes a decision to implement or decline an idea. Commenting, on the other hand, is still possible after an idea has been implemented or declined, yet only takes place for 5% of the ideas. Our analysis only includes crowd votes and crowd comments up until an idea received its implementation status. Figure 1 demonstrates that ideas accumulate crowd comments slowly. An idea receives 70% of its final number of comments after approximately 500 days (~1 year 5 months).

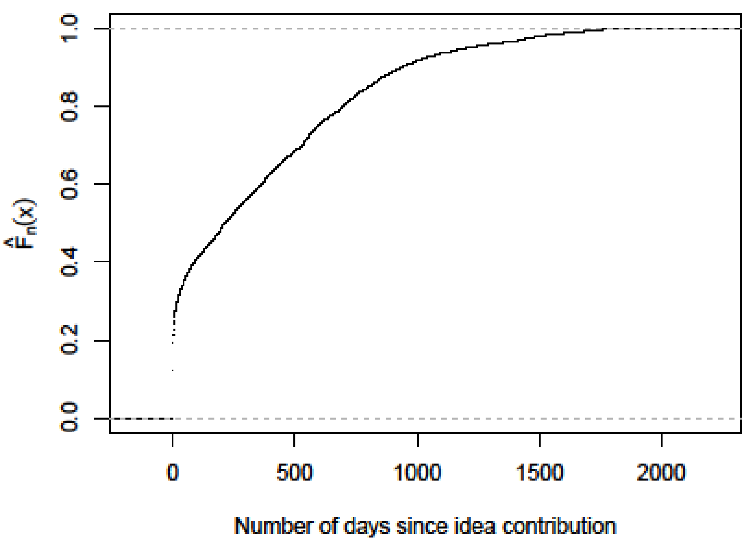


Figure 1: Cumulative Distribution of the Number of Crowd Comments

The time between idea contribution and final status (implemented/declined) varies significantly (M = 1,071 days, SD = 547 days, Min = 0, Max = 2,184), so some ideas have more time to gather crowd feedback than others. This problem is addressed by normalizing the number of votes and comments by the time it took the firm to implement or decline an idea (in days). In this way, the focus is on the rate rather than the quantity of votes and comments. Experiments with a logarithmic discount for time were conducted (based on Figure 1) and found the results and relationships to be consistent with the proposed linear normalization. Summarizing, two measures related to crowd feedback are include into our model: *number of crowd votes* (per day) and *number of crowd comments* (per day). Both measures were log-transformed prior to the analysis.

Table 2: Variables Included in the Model						
Variable category	**Variable name**	**Description**	***M***	**SD**	**Min**	**Max**
Idea implementation (dependent variable)	y_i	=1 if idea *i*'s status is marked as implemented; 0 if marked as declined.	0.09	0.29	0	1
Content	*LSI components*	11 idea concepts derived from previously submitted ideas.	-0.001	0.003	-0.89	0.84
	Relative distinctiveness	Degree of similarity idea *i* has compared to previously submitted ideas.	0.79	0.26	0.01	1
Contributor experience	*Number of past ideas*	Number of ideas contributor *c* posted before submitting idea *i*.	0.15	0.40	0	2.57
	Number of previous contributor comments	Number of comments contributor *c* posted on ideas before submitting idea *i*.	0.05	0.28	0	3.37
	Number of past implemented ideas	Number of implemented ideas contributor *c* posted before submitting idea *i*.	0.04	0.19	0	1.79
	Tenure	Number of days contributor *c* spent in the community before submitting idea *i*.	0.66	1.66	0	7.40
Crowd feedback	*Number of crowd votes*	Crowd voting activity idea *i* acquired during observation time,	0.02	0.12	0	2.49

		normalized by observation time.				
	Number of crowd comments	Crowd commenting activity idea *i* acquired during observation time, normalized by observation time.	0.01	0.005	0	1.39
Control variables	*Month*	Month idea *i* was submitted.	6.53	3.50	1	12

Data analysis methods

A benchmark of four methods was conducted in predicting, on the basis of the 3C data outlined above, whether an idea will be implemented or declined: linear discriminant analysis (*LDA*; Fisher, 1936); regularized logistic regression (*LR*; Friedman, Hastie, and Tibshirani, 2010); stochastic adaptive boosting (*AB*; Friedman, 2001); and random forests (*RF*; Breiman, 2001). Each of these methods can estimate a probability of implementation for a given new idea.

For LDA, maximum likelihood estimation was used and equal covariance matrices between implemented and declined ideas was assumed. LR uses lasso regularization, which reduces the probability of overfitting by shrinking the coefficients with an L1 penalty term, multiplied by λ, added to the log likelihood objective function[4].

The following two paragraphs provide background on the tree models used by RF and AB. Consider the hypothetical tree model in Figure 2. Starting from the top node, 20% of the ideas in the dataset are implemented. Breaking down the ideas according to the number of votes, 36% of those that received at least one vote are implemented, whereas only 4% of those that received no votes are implemented. Going further down the tree, ideas with at least one vote that have a relative distinctiveness score of 0.9 or higher are implemented in 89% of cases. An idea with no votes and a distinctiveness score of less than 0.9, on the other hand, is implemented only in 3% of cases. In sum, in this hypothetical example, one only needs a new idea's number of crowd votes and its relative distinctiveness to predict its probability of implementation. An idea with two votes and a relative distinctiveness of 0.95, for example, has an 89% probability of implementation.

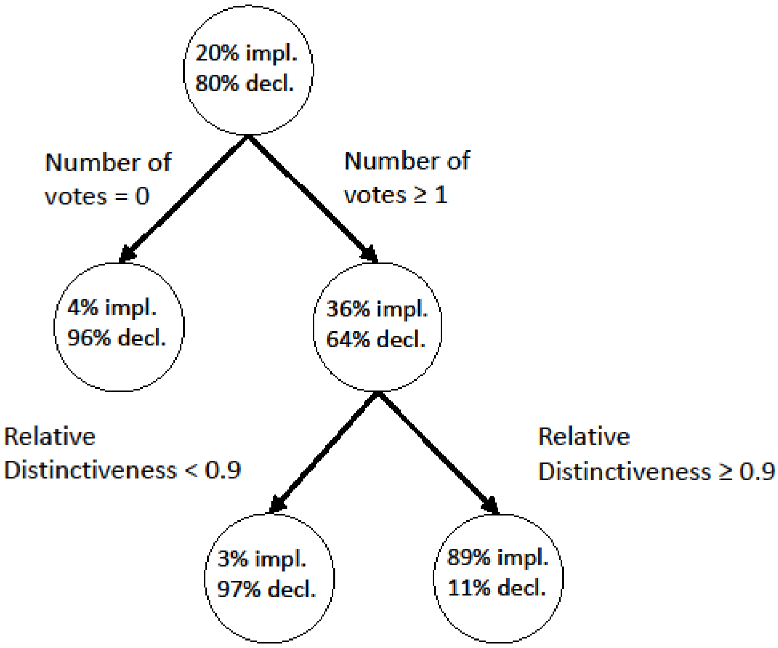


Figure 2: Hypothetical Tree Model

Both RF and AB grow a large number of single decision trees on a sample equal in size to the original data with replacement. Each decision tree makes a prediction when a new idea i is presented. The final probability for idea i is then the number of trees that predict this idea as implemented divided by the number of grown trees. This averaging step improves model performance by reducing the likelihood that a wrong model is chosen (Breiman, 1996). The main difference between the two algorithms is that RF only considers a random subset of candidate variables at each node of a decision tree (Breiman, 2001), whereas AB grows subsequent decision trees on samples of previously misclassified ideas (Friedman, 2001). Both algorithms only require two parameters to consider; for RF, the number of decision trees to grow (default: 500) and the number of variables to select at each node (default: *sqrt(k)*, where *k* is the number of variables in the dataset); and for AB, the number of iterations (default: 150) and the number of terminal nodes (default: 8). In a recent benchmark study, RF and AB were the top two performing classification algorithms currently available (Fernández-Delgado et al., 2014).

Model evaluation

Model performance is examined using the *Area Under the Receiver Operating Characteristic Curve* (*AUROC* or *AUC*), which gives the probability that a model will rank a randomly-selected implemented idea higher than a randomly-selected declined idea (DeLong, DeLong, and Clarke-Pearson, 1988)[5]. Its values range from .5 (no differentiation between ideas) to 1 (perfect idea differentiation). The null hypothesis states that the AUCs are equal and is tested with the DeLong test, which is a variant of the Mann-Whitney *U*-test (DeLong, DeLong, and Clarke-Pearson, 1988).

Results

The added value of crowd feedback: Real-time data vs. time-delayed data

As discussed above, one of our key aims was to determine whether a firm benefits from waiting to obtain *crowd* data, which take time to accumulate, or whether it can just as easily make decisions in real time, based on *contributor* and *content* information. Two separate models are built to assess the added value of crowd feedback in idea selection: the first contains content and contributor experience (scenario 1) and the second additionally includes crowd feedback (scenario 2). Figure 3 summarizes the results for the various prediction algorithms implemented.

In all cases, the performance of the model that included all of the 3Cs was superior to that of the model that included only content and contributor experience. The improvement in terms of AUC was by .113 (+17.9%), .186 (+29.6%), .295 (+48.1%), and .274 (+43.8%) in LDA, LR, AB, and RF, respectively. The DeLong test confirms that these increases are significant across all classifiers ($p < .001$). In scenario 1, the models can predict idea implementation accurately with an AUC ranging between .613 (AB) and .630 (LDA). This means that the probability that a model will rank an implemented idea higher than a declined idea is between 61.3% and 63.0%. After including crowd feedback (scenario 2), this probability increases to between 74.3% (LDA) and 90.8% (AB).

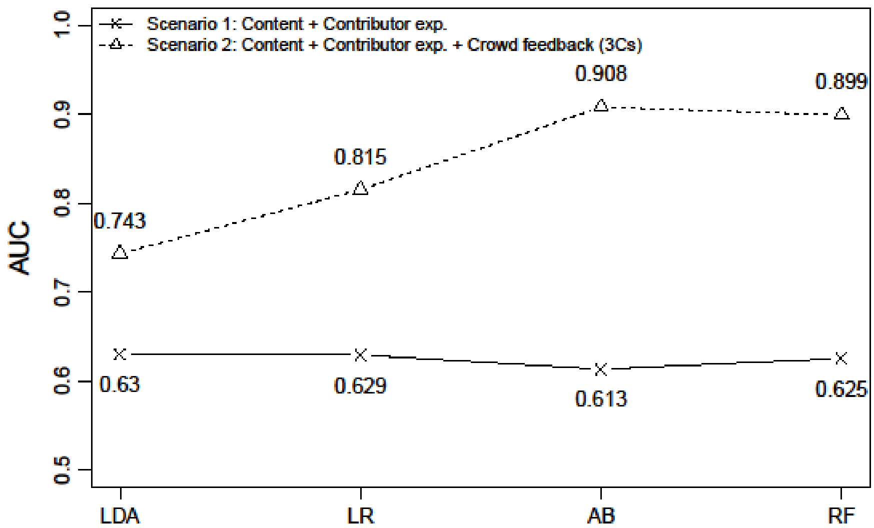


Figure 3: AUC for Scenarios 1 and 2

The performance of our automated models is further compared with the performance of convenient heuristics (Table 3). These heuristics are intuitive and easy to implement, and are used by firms for idea ranking (Jouret, 2009). Three idea ranking heuristics are investigated: the first was based on the number of crowd votes (high to low), the second was based on the number of crowd comments (high to low), and the third was based on random selection (equivalent to an unbiased coin toss). For scenario 1, the use of our model improved idea selection by between 8.7% and 11.7% over ranking by votes (AUC = .564), between -3.8% and -1.1% over ranking by comments (AUC = .637), and between 22.6% and 26.0% over random idea selection (AUC = .5). Thus, idea selection on the basis of content coupled with contributor experience is superior to random idea selection or idea selection based on the number of votes, but is marginally inferior to idea selection based on the number of comments. For scenario 2, the use of our model improved idea selection by between 31.7% and 61.0% over ranking by votes (AUC = .564), between 16.6% and 42.5% over ranking by comments (AUC = .637), and between 48.6% and 81.6% over random idea selection (AUC=.5). In sum, using all 3Cs performs systematically better than using idea ranking heuristics over several algorithms. Across algorithms, AB and LDA were, respectively, the best- and worst-performing classifiers for both scenarios.

Table 3: Benchmarking Model Performance over Heuristics				
	Linear discriminant analysis (LDA)	*Regularized logistic regression (LR)*	*Stochastic adaptive boosting (AB)*	*Random forests (RF)*
Scenario 1: Content + Contributor Experience				
AUC	.630	.629	.613	.625
Percentage improvement over *crowd vote ranking*	11.7%	11.5%	8.7%	10.8%

(AUC=.564)				
Percentage improvement over *crowd comment ranking* (AUC=.637)	-1.1%	-1.3%	-3.8%	-1.9%
Percentage improvement over *random idea selection* (AUC=.500)	26.0%	25.8%	22.6%	25.0%
Scenario 2: Content + Contributor Experience + Crowd Feedback				
AUC	.743	.815	.908	.899
Percentage improvement over *crowd vote ranking* (AUC=.564)	31.7%	44.5%	61.0%	59.4%
Percentage improvement over *crowd comment ranking* (AUC=.637)	16.6%	27.9%	42.5%	41.1%
Percentage improvement over *random idea selection* (AUC=.500)	48.6%	63.0%	81.6%	79.8%

The relative importance of the 3Cs

Next, the random forests (RF) algorithm is used to investigate the relative importance of each of the 3Cs in predicting idea implementation, and the nature of each variable's relationship with idea implementation. RF was selected for this purpose because it performs well on this problem and has convenient model interpretation tools. Instead of coefficients and *t*-statistics, RF offers variable importance measures and partial dependence plots. Variable importance in RF is most often measured using the mean decrease in the Gini measure of node impurity, defined as $p(1-p)$, where p is the estimated probability that an idea is implemented. This measure represents the total decrease in node impurity from all splits in the trees divided by the total number of trees. The more a variable decreases impurity (i.e., the higher the Gini) the more predictive it becomes.

Figure 4 displays variable importance values for all variables in the most comprehensive model (scenario 2). There is a clear break in the plot based on the mean decrease in Gini: higher than 80 for all crowd feedback variables, between 25 and 10 for all content variables, and lower than 5 for all contributor experience variables. The number of crowd votes and crowd comments are, respectively, ranked first and second, confirming our model-level results that crowd feedback is indeed important. The number of votes (Gini = 152.7) is 1.75 times more important than the number of comments (Gini = 86.9). The LSI concepts are ranked between 3 and 14, with relative distinctiveness at rank 5. The four lowest-ranked items all relate to contributor experience and have similar Gini scores.

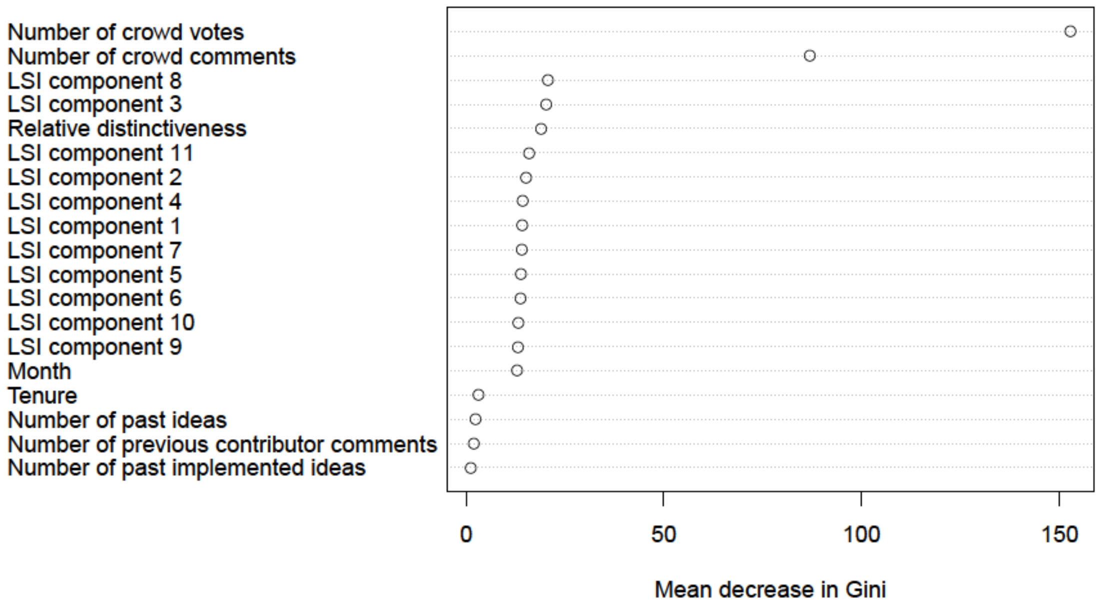
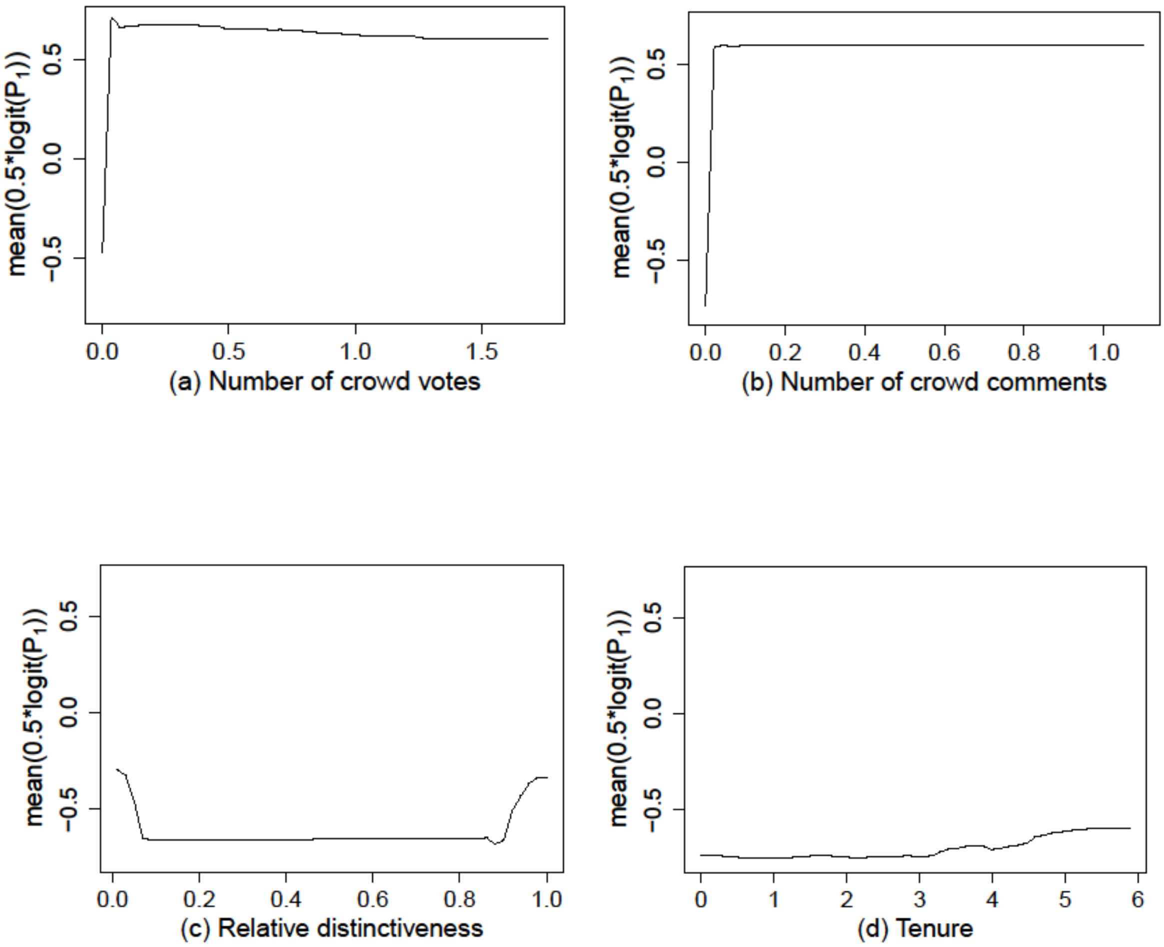


Figure 4: Variable Importance Plot for Scenario 2

Having assessed the relative importance of the 3Cs, our attention is turned to the relationships of the variables with idea implementation. Whereas the direction of the relationship in statistical models such as logistic regression would be determined by looking at the signs of the coefficients, the nature of the relationship can be explored in RF with partial dependence plots. A partial dependence plot shows the average probability of implementation for each value of a variable *x*, holding all other variables constant (Breiman, 2001).

Figure 5 displays the partial dependence plots for each variable of the 3Cs, except for the LSI components, which are highly specific to Mendeley. The probability of implementation stabilizes after 0.05 for the number of crowd votes and crowd comments (Figure 5a, 5b). This means that if an idea receives at least one (=exp(0.05)) vote (comment) per day, it is, on average, more likely to be implemented. Interestingly, receiving two or more votes (comments) per day does not improve the likelihood of implementation over receiving one. Regarding idea content, Figure 5c shows that if an idea is either very similar (i.e., less distinctive) or very dissimilar (i.e., more distinctive) to previous ideas, its odds of implementation increase. Ideas stuck in the middle however, have a lower probability of implementation.

Figures 5d-g depict the variables of contributor experience: *tenure*, *number of past implemented ideas*, *number of previous contributor comments*, and *number of past ideas*. All show similar, positive relationships with idea implementation: more interaction with the community or longer membership in the community is associated with a higher probability of implementation. Specifically, contributors have a higher probability of having their ideas implemented when they have been with the community for at least 25 (=exp(3.2)) days (Figure 5d), have had one idea implemented previously (Figure 5e), made at least one comment on another idea (Figure 5f) or posted at least 7 (=exp(2)) ideas (Figure 5g). Finally, Figure 5h shows the relationship between *month* and idea implementation. Given that it is a control variable and that its effects are likely to be firm-specific (and therefore not generalizable), this is not discussed in greater detail.



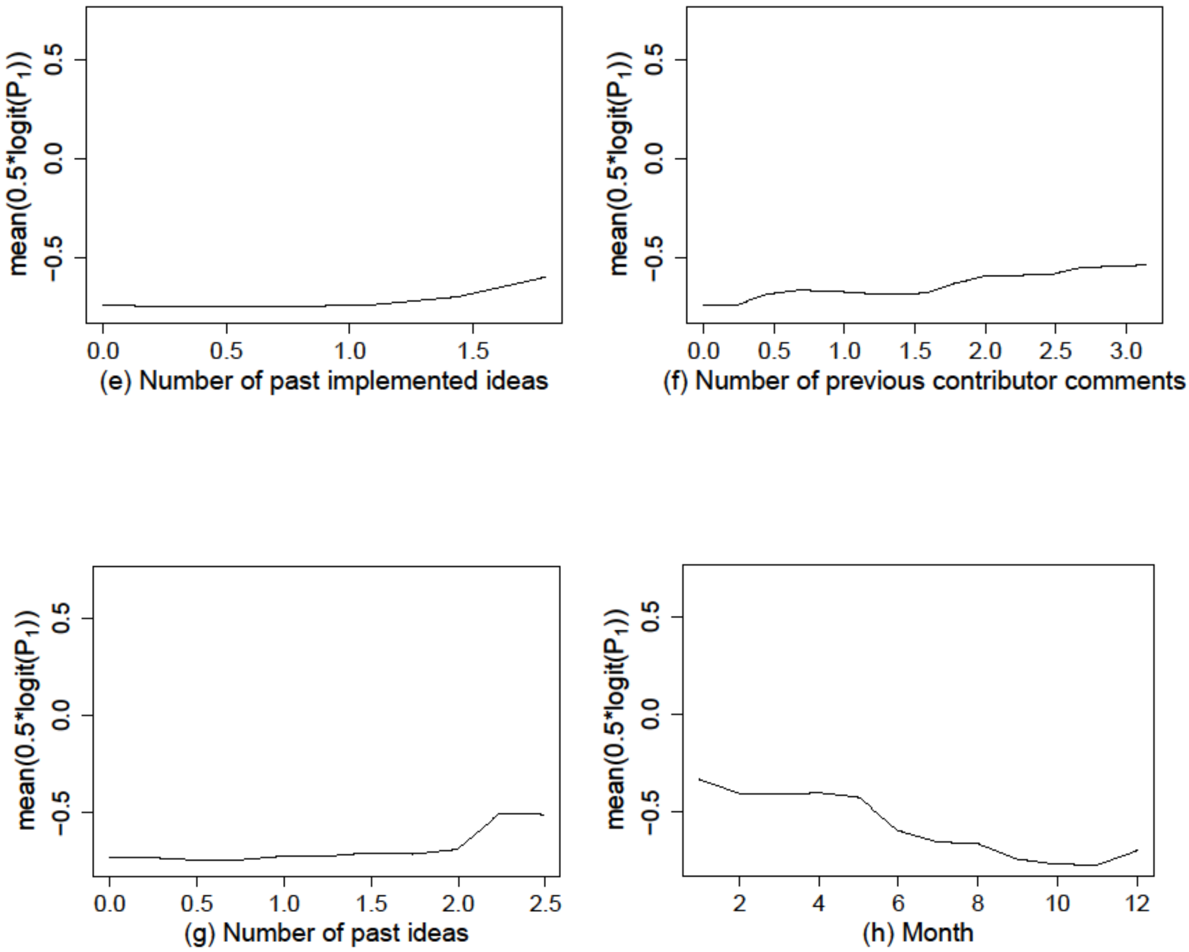


Figure 5: Partial Dependence Plots of the 3Cs

Discussion

Our results suggest that waiting for crowd data—and specifically, structured data, i.e., the number of votes and comments that an idea receives per day—may be worthwhile: including this information improves idea selection between 17.9% and 48.1% over using content and contributor experience. The nonlinear models (AB, RF) substantially outperformed the linear models (LDA, LR) when crowd data is incorporated, suggesting that the former can capture nonlinearities and interactions not captured by the latter. This finding necessitates more research on the use of nonlinear methods in idea selection. Our results further indicate that ideas need to surpass an initial threshold of obtaining one crowd vote (comment); achieving this improves the odds of implementation substantially. These findings suggest that, after controlling for content and contributor experience, the decision criteria of both the crowd and the firm are likely to be well aligned.

When the manager does not have time to wait for the crowd, then he or she can gain information from the idea's content, which is comprised of unstructured textual data. To distinguish between novel and similar ideas, this study developed a measure for ideas' relative distinctiveness as compared with previously-submitted ideas. This operationalization is similar to Walter and Back (2013) except that latent semantic indexing (LSI) is performed before applying *k*-means clustering. Both highly dissimilar (i.e., distinctive, new) and highly similar (i.e., more of the same) ideas have a higher implementation probability. Moderately distinctive ideas, however, are not likely to be implemented. The use of nonlinear data analysis methods exposed this previously undiscovered relationship between idea distinctiveness and idea implementation. Our observations here make sense, since firms look for both incremental ('do better, yet more of the same') and radical ('new to the firm or industry') ideas (Pisano, 2015; Tidd and Bessant, 2009). Highly innovative ideas typically demand more support from senior management (Ettlie, Bridges, and O'keefe, 1984), and require substantial time and investments to develop. Therefore, it makes sense to first identify ideas that the firm considers quick wins (e.g., bug fixes; Dahl, Lawrence, and Pierce, 2011). Since community members who vote and comment on ideas, it is reasonable to find that idea content is important, but not the top predictor of idea implementation compared to crowd feedback, since text mining approaches, unlike humans, can fail to accurately capture information in ideas (Westerski, Dalamagas, and Iglesias, 2013).

Contributor experience was also predictive of idea implementation; however, of the 3Cs, its role was smallest. Generally, similarly to Bayus (2013), more experience in generating ideas facilitates idea implementation: community members who have a history of generating ideas or have had ideas implemented previously have a higher probability of having new ideas implemented compared with contributors with no such history. This observation is supported by research of Simonton (2003, 2004), which argues that a contributor's productivity in generating implemented ideas is strongly associated with the number of submitted ideas. One reason for our observations could be that, as users generate more ideas and monitor the firm's response on these ideas, they get a better sense of what the firm considers valuable and likewise adapt their suggestions to the firm's feedback (Marsh, Landau, and Hicks, 1996). Although this may result in ideas that are less original and less valuable to the firm (Dahl and Moreau, 2002), this is not necessarily a problem in cases of small updates. Similarly, a positive effect of community membership (*tenure*) and community interaction (*number of previous contributor comments*) on idea implementation can be observed. This finding supports prior literature that stated that by communicating, contributors revise their own ideas (Perry-Smith, 2006; Perry-Smith and Shalley, 2003), get multiple views on their ideas, and are more exposed to problems faced by other consumers (Lu, Singh, and Srinivasan, 2011). As a result, they are more likely to generate ideas relevant to the firm (Osborn, 1953).

It is important to note that the positive effects of prior idea generation took time to develop. More specifically, contributors become better at generating valuable ideas in a nonlinear fashion. Therefore, in line with previous research (Bayus, 2013; Lu, Singh, and Srinivasan, 2011), firms are advised to aim to retain contributors for longer periods of time to be able to sufficiently capitalize on this effect. In our study, members who were active for about a month (25 days) prior to their idea suggestion had a higher probability of getting their idea implemented.

Theoretical Implications

Our study is the first to simultaneously include the 3Cs regarding new ideas—content, contributor experience, and crowd feedback—and benchmark classical linear methods and nonlinear machine learning methods in predicting idea implementation. Though extant literature has pointed to the potential contributions of each of the 3Cs to the idea selection process, our study is among the first empirical studies to integrate all three categories and assess their relative predictive importance. A key consideration in our approach is that the information categories differ in terms of the extent to which they contain structured data—which are easier to process compared with unstructured data—and the timing at which they become available. In particular, whereas idea content and contributor information are available immediately, crowd feedback takes time to accumulate.

To the best of our knowledge our approach of applying nonlinear machine learning algorithms (with superior accuracy) to the combined 3Cs is both unique and more robust than models in extant literature. Our results indicate that crowd feedback is the best predictor of idea implementation. This conclusion contradicts previous research that found that the crowd is unable to select valuable ideas (Di Gangi and Wasko, 2009; Faure, 2004; Rietzschel, Nijstad, and Stroebe, 2006). In the context of crowdsourcing communities, reliance on crowd feedback may be beneficial because popularity may reflect future demand for the innovation, thereby reducing the firm's uncertainty regarding idea selection (Di Gangi and Wasko, 2009; Di Gangi, Wasko, and Hooker, 2010); or the firm may feel pressure to be more user-oriented to ensure enduring community participation (Di Gangi, Wasko, and Hooker, 2010). In conclusion, empowering the crowd is important because it can help the firm in making better idea implementation decisions. This means that it is important for firms to manage the crowd in the long run by keeping them engaged and by avoiding resentment. Di Gangi, Wasko, and Hooker (2010) argue that the best way to do this is for the firm to comment on ideas, ask questions, pay attention to crowd feedback, and, especially, respond swiftly. A timely response is important because it signals to the community that its efforts are appreciated.

Managerial Implications

Innovation in today's data-rich environment presents unique challenges to firms. Crowdsourcing communities afford us to elicit large volumes of new product suggestions, ideas, and potential solutions, but selecting the best ideas can be expensive and time-consuming (Klein and Garcia, 2015). Calls have been made to examine which methods and data are best to make real-time NPD decisions—in our case selecting ideas—in data-rich environments (Bharadwaj and Noble 2015). From Table 1, it is apparent that the limited number of studies that focus on idea selection in crowdsourcing communities have directed little or no attention towards these suggestions. This study responds to these calls and next the managerial implication that stem from our results are described. This manuscript, to the best of our knowledge, marks the first study to: 1) benchmark classical and machine learning methods in idea scoring, and 2) compare the performance of idea scoring models incorporating data that is available in real time with models that utilize data that becomes available at a later stage. Our results assist managers in determining how to process large volumes of rapidly-generated ideas in crowdsourcing communities as efficiently as possible.

Our analysis indicates that the nonlinear machine learning methods utilized in this study substantially outperform classical statistical methods. This is only true when crowd data are included in the model. When only content and contributor variables are included both classes of methods show similar performance. Therefore, it is recommended to refrain from using more computational intensive machine learning methods in a real-time setting. For models that include the 3Cs it is recommended to use machine learning algorithms to capture the complex relationships that govern the data.

The information on idea content and the contributor's past idea generation experience (scenario 1) improves predictive performance up to 26.0% over random idea selection. If idea ranking can wait for the wisdom of the crowd, however, performance can be improved further by up to 48.1% (scenario 2). Hence, ranking ideas in real time is a viable option, but waiting for the wisdom of the crowd is desirable. Therefore, it is recommended that firms implement two idea selection support systems: one real-time system that can immediately rank new ideas based on content and contributor experience; and an additional one that integrates the crowd's idea evaluation after it has had sufficient time to provide feedback. The two systems are complementary and can be used simultaneously.

Limitations and Future Research

Despite our contributions to literature and practice, several limitations apply to our study.

First, our analysis regards ideas from a crowdsourcing community in the IT industry. This industry was chosen to enable us to compare our findings to those of prior research on crowdsourcing communities (e.g., IdeaStorm, MyStarbucksIdea) and because of the public availability of these ideas. Yet, it is possible that our findings are inherently specific to the IT industry and that the relative predictive roles of the 3Cs vary across industries or contexts. Future research could investigate whether our results hold in other industries.

In addition, Mendeley's business model may impact the generalizability of our findings. In contrast to Dell or Starbucks, Mendeley uses a freemium pricing strategy. As a result, price is unlikely to impact a consumer's decision whether or not to use Mendeley. In Dell's case, however, purchasing power does play a role. Furthermore, Mendeley focuses its efforts on a niche market (i.e., academic reference software), whereas Dell's products apply to a broader, more heterogeneous consumer market (i.e., personal computers and technology). Both elements, however, can influence user participation in crowdsourcing communities, which in turn impacts the quality of idea generation efforts (Lu, Singh, and Srinivasan, 2011). More research on this aspect is therefore required.

Finally, our study analyzed ideas that had either been implemented or declined. Advancing an idea to development is seldom a straightforward go/no-go decision, so the firm may make mistakes on idea implementation given this uncertainty. Instead of analyzing a binary decision (i.e., implementation or rejection), future research could use dependent measures in the implementation decision that are more proximal to the firm's bottom line such as sales, R&D investments, cost structure, or market uncertainty. In addition, future research could explore the intermediate statuses of idea implementation more thoroughly. In the case of Mendeley, these included *under review*, *started*, and *planned*. Future research to investigate these intermediate states in more detail is encouraged.

References

Alam, I. 2006. Removing the fuzziness from the fuzzy front-end of service innovations through customer interactions. *Industrial Marketing Management* 35(4): 468–80.

Amabile, T. M. 1983. The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology* 45(2): 357–76.

Bayus, B. L. 2013. Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management Science* 59(1): 226–44.

Bharadwaj, N., and Y. Dong. 2014. Toward further understanding the market-sensing capability-value creation relationship. *Journal of Product Innovation Management* 31(4): 799–813.

Bharadwaj, N., J. R. Nevin, and J. P. Wallman. 2012. Explicating hearing the voice of the customer as a manifestation of customer focus and assessing its consequences. *Journal of Product Innovation Management* 29(6): 1012–30.

Bharadwaj, N., and C.H. Noble 2015. Call for papers: Special Issue of Journal of Product Innovation Management: Innovation in Data-Rich Environments.

Blohm, I., J. M. Leimeister, and H. Krcmar. 2013. Crowdsourcing: How to benefit from (too) many great ideas. *MIS Quarterly Executive* 12(4): 199-211.

Breiman, L. 1996. Bagging predictors. *Machine Learning* 24(2): 123–40.

Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5–32.

Calantone, R. J., C. A. Di Benedetto, and J. B. Schmidt. 1999. Using the analytic hierarchy process in new product screening. *Journal of Product Innovation Management* 16(1): 65–

76.

Carbonell-Foulquié, P., J. L. Munuera-Alemán, and A. I. Rodríguez-Escudero. 2004. Criteria employed for go/no-go decisions when developing successful highly innovative products. *Industrial Marketing Management* 33(4): 307–16.

Christensen, K., S. Nørskov, L. Frederiksen, and J. Scholderer. 2016. In search of new product ideas: Identifying ideas in online communities by machine learning and text mining. *Creativity and Innovation Management* (2016).

Cormican, K., and D. O'Sullivan. 2004. Auditing best practice for effective product innovation management. *Technovation* 24(10): 819–29.

Coussement, K., and D. Van den Poel. 2008. Integrating the voice of customers through call center emails into a decision support system for churn prediction. *Information & Management* 45(3): 164–74.

Dahl A., J. Lawrence, and J. Pierce. 2011. Building an innovation community. *Reseach-Technology Management* 54(5): 19–27.

Dahl, D. W., and P. Moreau. 2002. The influence and value of analogical thinking during new product ideation. *Journal of Marketing* 39(1): 47–60.

Deerwester, S., S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41(6): 391–407.

DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3): 837–45.

Di Gangi, P. M., and M. M. Wasko. 2009. Steal my idea! Organizational adoption of user innovations from a user innovation community: A case study of Dell IdeaStorm. *Decision Support Systems* 48(1): 303–12.

Di Gangi, P. M., M. M. Wasko, and R. E. Hooker. 2010. Getting customers' ideas to work for you: learning from Dell how to succeed with online user innovation communities. *MIS Quarterly Executive* 9(4): 213–28.

Eling, K., F. Langerak, and A. Griffin. 2015. The performance effects of combining rational and intuitive approaches in making new product idea evaluation decisions. *Creativity and Innovation Management* 24(3): 1–18.

Ettlie, J. E., W. P. Bridges, and R. D. O'keefe. 1984. Organization strategy and structural differences for radical versus incremental innovation. *Management science* 30(6): 682-95.

Faure, C. 2004. Beyond brainstorming: Effects of different group procedures on selection of ideas and satisfaction with the process. *The Journal of Creative Behavior* 38(1): 13–34.

Feldman, R., and J. Sanger. 2007. *The text mining handbook: advanced approaches in analyzing unstructured data.* New York: Cambridge University Press.

Fernández-Delgado, M., E. Cernadas, S. Barro, and D. Amorim. 2014. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *Journal of Machine Learning Research* 15(1): 3133−81.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of eugenics* 7(2): 179-88.

Franke, N., and S. Shah. 2003. How communities support innovative activities: An exploration of assistance and sharing among end-users. *Research Policy* 32(1): 157–78.

Franke, N., E. von Hippel, and M. Schreier. 2006. Finding commercially attractive user innovations: A test of lead-user theory. *Journal of Product Innovation Management* 23(4): 301–15.

Friedman, J. H. 2001. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* 29(5): 1189–1232.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33(1): 1.

Fuchs, C., and M. Schreier. 2011. Customer empowerment in new product development. *Journal of Product Innovation Management* 28(1): 17–32.

Hossain, M., and K. M. Z. Islam. 2015. Generating ideas on online platforms: A case study of "My Starbucks Idea." *Arab Economic and Business Journal* 10(2): 102–11.

Howe, J. 2006. The rise of crowdsourcing. *Wired Magazine* 14(6): 1–4.

Howe, J. 2008. *Crowdsourcing: why the power of the crowd is driving the future of business.* New York: Crown Publishing Group.

Hrastinski, S., N. Z. Kviselius, H. Ozan, and M. Edenius. 2010. A review of technologies for open innovation: characteristics and future trends. *Proceedings from 2010 43rd Hawaii International Conference on System Sciences* 2010: 1–10.

IdeaStorm. Available at http://www.ideastorm.com. Accessed on September 27, 2016.

Jansson, D. G., and S. M. Smith. 1991. Design fixation. *Design Studies* 12(1): 3–11.

Jouret, G. 2009. Inside Cisco's search for the next big idea. *Harvard Business Review* 87(9): 43-45.

Klein, M., and A. C. B. Garcia. 2015. High-speed idea filtering with the bag of lemons. *Decision Support Systems* 78(2015): 39–50.

Kohn, N. W., P. B. Paulus, and Y. Choi. 2011. Building on the ideas of others: An examination of the idea combination process. *Journal of Experimental Social Psychology* 47(3): 554–61.

Kornish, L. J., and K. T. Ulrich. 2014. The importance of the raw idea in innovation: Testing the sow's ear hypothesis. *Journal of Marketing Research* 51(1): 14-26.

Lang, M., N. Bharadwaj, and C. A. Di Benedetto. 2016. How crowdsourcing improves prediction of market-oriented outcomes. *Journal of Business Research* 69(10): 4168-76.

Levitt, T. 1963. Creativity is not enough. *Harvard Business Review* 41(3): 72–83.

Lu, Y., P. V. Singh, and K. Srinivasan. 2011. How to retain smart customers in crowdsourcing efforts? A dynamic structural analysis of crowdsourcing customer support and ideation. Paper presented at Conference of Information Systems and Technology, Charlotte.

Magee, G. B. 2005. Rethinking invention: Cognition and the economics of technological creativity. *Journal of Economic Behavior and Organization* 57(1): 29-48.

Magnusson, P. R. 2009. Exploring the contributions of involving ordinary users in ideation of technology-based services. *Journal of Product Innovation Management* 26(5): 578–93.

Magnusson, P. R., J. Netz, and E. Wästlund. 2014a. Exploring holistic intuitive idea screening in the light of formal criteria. *Technovation* 34(5): 315–26.

Magnusson, P. R., E. Wästlund, and J. Netz. 2014b. Exploring users' appropriateness as a proxy for experts when screening new product/service ideas. *Journal of Product*

Innovation Management 33(1): 4-18.

Marsh, R. L., J. D. Landau, and J. L. Hicks. 1996. How examples may (and may not) constrain creativity. *Memory & Cognition* 24(5): 669–80.

Moreau, P., D. Lehmann, and A. Markman. 2001. Entrenched knowledge structures and consumer response to new products. *Journal of Marketing Research* 38(2): 14–29.

Mosteller, F., and J. W. Tukey. 1977. *Data analysis and regression: a second course in statistics.* Addison-Wesley Series in Behavioral Science: Quantitative Methods.

MyStarbucksIdea. Available at http://mystarbucksidea.force.com. Accessed on December 15, 2015.

Nagar, Y., P. De Boer, and A. C. B. Garcia. 2016. Accelerating the review of complex intellectual artifacts in crowdsourced innovation challenges. Paper presented at Thirty Seventh International Conference on Information Systems, Dublin.

Onarheim, B., and B. Christensen. 2012. Distributed idea screening in stage-gate development processes. *Journal of Engineering Design* 23(9): 660–73.

Osborn, A. F. 1953. *Applied imagination, principles and procedures of creative thinking.* New York: Charles Scribner's Sons.

Perry-Smith, J. E. 2006. Social yet creative: the role of social relationships in facilitating individual creativity. *The Academy of Management Journal* 49(1): 85–101.

Perry-Smith, J. E., and C. E. Shalley. 2003. The social side of creativity: a static and dynamic social perspective. *Academy of Management Review* 28(1): 89–106.

Perttula, M., and P. Sipilä. 2007. The idea exposure paradigm in design idea generation.

Journal of Engineering Design 18(1): 93–102.

Pisano, G. P. 2015. You need an innovation strategy. *Harvard Business Review* 93(6): 44–54.

Pitta, D. A., and D. Fowler. 2005. Online consumer communities and their value to new product developers. *Journal of Product & Brand Management* 14(5): 283–91.

Poetz, M. K., and M. Schreier. 2012. The value of crowdsourcing: Can users really compete with professionals in generating new product ideas? *Journal of Product Innovation Management* 29(2): 245–56.

Preece, J., B. Nonnecke, and D. Andrews. 2004. The top five reasons for lurking: improving community experiences for everyone. *Computers in Human Behavior* 20(2): 201–23.

Purcell, A. T., and J. S. Gero. 1996. Design and other types of fixation. *Design Studies* 17(4): 363–83.

Riedl, C., I. Blohm, J. M. Leimeister, and H. Krcmar. 2010. Rating scales for collective intelligence in innovation communities: why quick and easy decision making does not get it right. *Proceedings from the 31st International Conference on Information Systems* 2010: 98–104.

Riedl, C., I. Blohm, J. M. Leimeister, and H. Krcmar. 2013. The effect of rating scales on decision quality and user attitudes in online innovation communities. *International Journal of Electronic Commerce* 17(3): 7-36.

Rietzschel, E. F., B. A. Nijstad, and W. Stroebe. 2006. Productivity is not enough: A comparison of interactive and nominal brainstorming groups on idea generation and selection. *Journal of Experimental Social Psychology* 42(2): 244–51.

Rietzschel, E. F., B. A. Nijstad, and W. Stroebe. 2010. The selection of creative ideas after individual idea generation: Choosing between creativity and impact. *British Journal of Psychology* 101(1): 47–68.

Rhyn, M. and I. Blohm. 2017. A machine learning approach for classifying textual data in crowdsourcing. *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik* 2017: 1171-85.

Saarijärvi, H., P. K. Kannan, and H. Kuusela. 2013. Value co-creation: theoretical approaches and practical implications. *European Business Review* 25(1): 6–19.

Sawhney, M., G. Verona, and E. Prandelli. 2005. Collaborating to create: The internet as a platform for customer engagement in product innovation. *Journal of Interactive Marketing* 19(4): 4–34.

Simonton, D. 2003. Scientific creativity as constrained stochastic behavior: The integration of product, person and process perspectives. *Psychological Bulletin* 129(4): 475–94.

Simonton, D. 2004. *Creativity in Science: Chance, Logic, Genius, and Zeitgeist.* New York: Cambridge University Press.

Sternberg, R. J., and T. I. Lubart. 1999. The concept of creativity: Prospects and paradigms. *Handbook of creativity* (1999): 3–15.

Surowiecki, J. 2004. *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations little.* London: Little Brown.

Thorleuchter, D., D. Van den Poel, and A. Prinzie. 2010. Mining ideas from textual information. *Expert Systems with Applications* 37(10): 7182–8.

Tarrell, A., N. Tahmasbi, D. Kocsis, J. Pedersen, A. Tripathi, J. Xiong, O. Oh, and G.-J.

deVreede. 2013. Crowdsourcing: A snapshot of published research. *Proceedings of the Nineteenth Americas Conference on Information Systems* (2006): 1–14.

Tidd, J., and J. Bessant. 2009. *Managing innovation: Integrating technological, market and organizational change.* Chichester: Wiley.

Toubia, O., and O. Netzer. 2017. Idea generation, creativity, prototypicality. *Marketing Science* 36(1): 1-20.

Urban, G. L., and R. J. Hauser. 1993. *Design and marketing of new products.* New Jersey: Prentice Hall.

von Hippel, E. 2005. *Democratizing innovation.* Cambridge: MIT Press.

Walter, T. P., and A. Back. 2013. A text mining approach to evaluate submissions to crowdsourcing contests. *Proceedings of the Annual Hawaii International Conference on System Sciences*: 3109–18.

Westerski, A., T. Dalamagas, and C. A. Iglesias. 2013. Classifying and comparing community innovation in Idea Management Systems. *Decision Support Systems* 54(3): 1316–26.

Zhu, J. J. H., and Z. He. 2002. Perceived characteristics, perceived needs, and perceived popularity adoption and use of the Internet in China. *Communication research* 29(4): 466–95.

Footnotes

[1]Note that it also takes time and resources to set up these automated methods. Whereas for human evaluation the direct cost per idea is roughly the same, automated methods have a high setup cost after which the operational cost (e.g., machine processing time) of rerunning the model is marginal. The author team thanks an anonymous reviewer for pointing this out.

[2]The intermediate categories represent only 2.59% of the total number of ideas in the sample: *under review* (*n*=114, 1.58%), *started* (*n*=22, 0.30%) and *planned* (*n*=51, 0.71%).

[3]Given that contributors in Mendeley rarely assign their ideas to idea categories (~2% of contributors), the variables from Bayus (2013) related to idea categories are not operationalized in our study.

[4]The amount of shrinkage λ was chosen by maximizing the AUC over a range of 100 λ-values on a holdout sample. Note that regularized logistic regression is the same as (standard) logistic regression when λ=0.

[5]To mimic their operational deployment, our models are evaluated with out-of-period validation, which is the most stringent type of evaluation (Mosteller and Tukey 1977, p. 38). Ideas submitted in 2008-2009 were used to fit the models (training set; *n*=1040) and ideas submitted in 2010-2014 were used as a holdout sample (*n*=6006). A random selection of 50% of the holdout sample was used to determine the optimal shrinkage value λ for LR (validation set) and the remaining 50% was used to estimate final model performance (test set).

Appendix

ID	Date Posted	Idea Title and Description	Remark	Status
366509	Oct. 27, 2009	*Android* *An android app to access Mendeley from android phones would be welcome (when all bugs in the desktop app are fixed ;-). Best, Claude.*	Received most crowd votes (7837 votes)	Implemented
142951	Mar. 18, 2009	*Check for duplicates* *Please add a possibility for checking for duplicates and not adding those/modifying the existing entries when importing! Especially please check for duplicates when automatic PDF extraction!*	Clear problem description	Implemented
1396351	Jan. 20, 2011	*Share This on Twitter thru Desktop* *It's great that I can share on twitter through the web interface, but I usually close my browser when I'm working, to minimize interruptions. But I WOULD like to share what I'm reading on my twitter account, so it would be great if I could share directly from the Desktop application.*	Unclear problem description	Declined
2338399	Oct. 23, 2011	*Webos*	Incomplete idea	Declined