IPRPD
International Journal of Business & Management Studies
ISSN 2694-1430 (Print), 2694-1449 (Online)
Volume 04; Issue no 07: July, 2023
DOI: 10.56734/ijbms.v4n7a2

SEMI-AUTOMATED REQUIREMEMT ELICITATION IN DATA DRIVEN MARKETING FOR PRODUCT INNOVATION BY USING MACHINE LEARNING

Prof. Dr. Clotilde Rohleder[1], Dr. Indra Kusumah[2]
Prof. Dr. -habil Camille Salinesi[3], Michael Meier[4]

[1] *Professor, University of Applied Sciences Constance, Germany*

[2]*lecturer, University of Applied Sciences Constance, Germany*

[3]*Professor, University of Paris 1, France*

[4]*CEO of Schindler & Parent GmbH, International Marketing Agency, Meersburg, Germany*

Abstract

The creation of an innovation needs new input, new ideas which are not easy to be acquired. Meanwhile in the era of internet people express their mind, their feeling and their judgement about anything in the digital world. Among others the crowd (those people) gives also comments about a physical object (hardware) or digital object (software). Some of the comments may give a direction for a new innovation. The usage and extraction of these information (comments) for the elicitation of product requirement for a new innovation is an object of research works and there are already methods developed by scholars. After studying the state of the art we come up with an idea to enhance the performance of the existing methods. We develop a process chain based on machine learning for an automated requirement elicitation based on crowd input optimized by a specific domain knowledge, applied for e-commerce. We focus on the aspects related to digital disruption as well as agile management in innovation. The crowd input gained from a digital big data available in twitter from the past 5 years until now. We analyze the digital consumer behavior by extracting this data. By using an optimized specific domain knowledge, we extract product requirement for a new innovation in the domain of e-commerce. We compare our method with the available state of the art methods and show better performance of our approach.

Keywords

Product Innovation, Marketing, Machine Learning, Requirement Engineering, E-Commerce

1. Introduction

The digitalization of the economy is taking place very fast and macroeconomic changes are strongly influenced by the digital aspect (United Nations, 2019). Digital business will grow even more with the Covid-19 pandemic that is spreading throughout the world. In addition, the development of fiber optic networks that continues to expand is an enabler for digital business growth (OECD, 2020). Digital business services are very dependent on their acceptance, the wishes and needs of users [13]. The ability of digital service platforms to provide solutions and provide excellent service to users are the key to the acceptance and use of e-commerce [15], [14]. This is also a determinant of whether users switch to other providers or not depending on satisfaction as a result of the ability of the digital service platform to be able to meet user expectations and desires [9].

A wealth of potentially valuable information about product innovation exists on social media [5], [16]. E-commerce users also share their aspirations and perceptions online [3] [4], those papers focus on these users.

Because most efforts to take advantage of this data are manual, they cannot handle a large amount of data. One promising way to analyze this data is to employ machine learning [2]. The earlier data of user aspirations and desires are obtained, the earlier the engineering process and the innovation of the e-commerce platform system can be engineered.

Speaking about innovation, the innovations is the creation and implementation of new processes, products, services and methods of delivery, which result in significant improvements in outcomes, efficiency, effectiveness or quality [17]. Innovation, which is a key driver of productivity growth, is subject to several well-documented market failures that lead to under-investment in R&D activities [18].

The following sub chapter presents a way to identify potential product or service innovations in the e-commerce domain.

2. Background and literature review

Similar to HYVE project team in Netnography [11], [12], we focus on review and comment possibilities in Internet about products to be able to use the customer feedback in social media and eCommerce for innovation and product politics purposes. In our research, similar to [11], [12], we use the artificial intelligence algorithms to deal with big data, but we go further in direction requirements engineering (RE) for innovation than Netnography's researchers do.

Some researches published results regarding potentially valuable information about product innovation that exists on social media [5], [16]. E-commerce users also share their aspirations and perceptions online [3] [4]. Those papers focus on these users. We research on usable data content for innovation.

In the field of requirements engineering one could find many research results like Lim [7]'s researches. They published articles related to an automated and data driven requirement elicitation. Lim [7] analyzed 1848 articles related to an automated and data driven requirement elicitation. He extracted three aspects of the solution step, namely:

- Types of dynamic data sources used for automated requirements elicitation
- Techniques used for automated requirements elicitation
- The outcomes of automated requirements elicitation

Lim found out that there is a clear dominance of human-sourced data, compared to the process-mediated and machine-generated data sources. As a result of that the techniques used for data processing are based on natural language processing, while the use of machine learning for classification and clustering is prevalent. The dominant intention of the proposed methods was to automate the elicitation process fully, rather than to combine it with traditional stakeholder-involved approaches. The final results regarding the completeness and the readiness of the elicited data for use in system development or evolution are currently limited—most of the studies obtain some of the information relevant for requirement's content, some studies target the identification of the core functionality or quality in terms of features, and only a few of the studies achieve a high-level requirement content.

Especially in context of requirement elicitation method Lim identified the following three common steps: (1) filtering out data irrelevant to requirements, (2) classifying text based on the relevance to different stakeholder groups, or (3) classifying text by categories of technical issues, such as bug reports and feature requests. We accomplished this classification using rule-based approaches and machine learning, mostly within the supervised learning paradigm.

Another scholar an aspect for the requirement elicitation, namely an automated prioritization of the requirement. Avesani [10] introduced a framework based on a requirements prioritization process that interleaves human and machine activities, enabling an accurate prioritization of requirements. Figure 1 [10] depicts the basic process that the evaluator undertakes. The types of data involved in the process are depicted as rectangles, namely: Requirements represent data in input to the process, that is the finite collection of requirements that have to be ranked; Requirements pair is a pair of candidate requirements whose relative preference is to be specified; Preference is the order relation between two alternative requirements elicited from the stakeholder. The preference is formulated as a Boolean choice on a pair; Ranking criteria are a collection of order relations that represents ordering induced by other criteria (e.g. the cost for the realization of the requirements, the estimated utility) defined on the initial set of requirements; Final ranking represents the resulting preference structure on the set of requirements. This final ranking may become the input to a further iteration of the process.

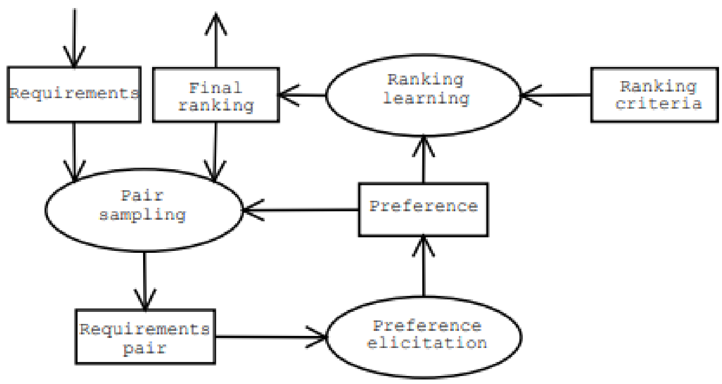

Figure 1: Basic Iteration of Requirements Prioritization Process *[10]*

We enhanced our approach of InnoCrowd - an AI Based Optimization of a Crowdsourced Product Development - [6] and we adopted new ideas from formerly published studies to develop a new method for a semi-automated requirement elicitation for product innovation in data driven marketing and applied it to the e-commerce domain. This method will use the base for the data classification related to our two main aspects mentioned in the next sub chapter, e-commerce and innovation in e-commerce.

3. Need for Base of data analysis for our contribution

We need to identify a base of data analysis by selecting only what pertains to e-commerce and then finding a subset that has significance for innovation.

To identify the data analysis base, we can use ontology for the e-commerce domain, we can use classification structure of a product, or we can use technological aspect as indicator for the relevance to the innovation for example. Initially, we chose a data analysis base from our previous work in the e-commerce domain. Later on, we optimized it.

3.1. Base of data analysis for e-commerce aspect

A study of user behaviour in e-commerce had been conducted by [8] in order to examine the causality and the relationship between the implementation of the theory of technology acceptance in forming individual or group attentions which ultimately give impact on consumer behavior in online shopping at the e-commerce marketplace. The following results could benefit industry managers and researchers of marketing management and international business:

- Perceived Convenience has a significant positive effect on Behavior Intention to use by 35%.
- Habit has a significant positive effect on Behavior Intention to use of 16.8%.
- Effort Expectation has a significant positive effect on Behavior Intention to use of 14.6%.
- Social Influence has a significant positive effect on Behavior Intention to use of 11.4%.
- Facilitating Condition has a significant positive effect on Behavior Intention to use of 9.9%.
- Performance Expectation has a significant positive effect on Behavior Intention to use of 8.8%.
- Perceived Risk has a significant positive effect on Behavior Intention to use of 7.8%.
- Trust has a significant positive effect on Behavior Intention to use of 7.2%.

Variables	Item Question	Code
Social Influence	Shopping online can improve my self-image	Image
	Shopping online can improve my lifestyle	Social Factors
Performance Expectancy	Buying items using e-commerce is faster	Help
	Buying goods using e-commerce is more effective	Effective
Effort Expectancy	Interacting with e-commerce is easy to understand	Interaction
	Interacting with e-commerce is clearly the procedure	Operation
	I feel like I use e-commerce.	Expertise
Habit	I repeatedly shop online using e-commerce	Repetition
	I often shop online using e-commerce	Habit
Facilitating Condition	I have sufficient knowledge of shopping online	Knowledge
	I have enough information to shop online	Resource
	I shop online using e-commerce because it fits with the available facilities	Compatible
Perceived Risk	Shopping using e-commerce products is always in line with advertising	Social
	Shopping using e-commerce time is relatively fast.	Performance
	Transaction process shop online fast	Time
	Online shopping is not expensive	Financial
	Secure buyer data when shopping online using e-commerce.	Security
Perceived Convenience	Shopping online can be from anywhere	Place
	Shopping online can be anytime	Time
	Easy online shopping where to access	Acquisition
	Easy shopping online transactions	Use
	Shopping online is easy to execute purchases	Execution
Trust	Shop online using e-commerce, the service is excellent	Ability
	Shop online using e-commerce products according to specifications	Expertise
	Shop online using e-commerce products in accordance with expectations	Integrity
Behavior Intention	I will shopping use-commerce next week	Do it
	I hope to use e-commerce in shopping	Hope
	I will probably use shopping using e-commerce in shopping	Receive
	I feel the need to use e-commerce in shopping	Adopt
Use Behavior E-commerce	I often shop online using e-commerce	Frequency
	Almost every week I shop online using e-commerce	Use
	I prefer shopping online using e-commerce in searching for goods	Receive
	I will always use e-commerce again in shopping	Utilization

Table 1: Variables for E-Commerce

Table 1 shows the facts in detail. Performance Expectancy (PE) we measured with four questions in dimensions (e.g., effective and helpful).

For effort expectancy (EE) we used six questions with dimensions (e.g., operations, interactions, skills). Habit (H) we measured with three questions on dimensions (e.g., implementation of behavior and repetition). Facilitating condition (FC) which consists of six items of questions on dimensions (e.g., resources, knowledge, compatible). Perceived risk (PR) which consists of ten items of questions on dimensions (e.g., performance risk, time risk, financial risk, security risk). Perceived Convenience (PC) consists of ten items of questions on dimensions (e.g., time, place, acquisition, use, execution). Trust (T) consists of six items of questions on dimensions (e.g., expertise, ability, integrity). Behavior Intention (BI) consists of seven items of questions (e.g., emphasis on doing, expect to use, must accept and use, motivation to adopt). Use behavior e-commerce (UB) consists of seven item questions on dimensions (e.g., the frequency of shopping online, the hope of reusing, having to accept and reuse and suppress the use of e-commerce applications).

We used these factors as our data base analysis for the classification of the text data in the e-commerce domain.

3.2. Base of data analysis for innovation aspect

We collected information on the identification of aspects that may lead to innovations in e-commerce. We selected these aspects as most relevant to innovation:

- Fast delivery and instant pickup
- AI powered personalization
- Voice powered shopping
- Shopping using AR / VR Technology
- Chatbots to handle customer queries
- Blockchain
- Digital storefronts

Nowadays customers will not accept long delivery times. Anything that enables fast delivery and instant pickup is a good requirement for e-commerce success.

We can use artificial intelligence to provide customer with a high degree of personalization. Using AI, e-commerce companies can show products that are linked to the user's interest, which they are more likely to purchase.

By taking the solution from Amazon into account in the usage of Amazon's Alexa, we can see the very bright future of voice powered shopping. It is predicted that with the soaring popularity of this technology, voice powered shopping will become a common solution for e-shopping.

Today more and more e-commerce store owners woo shoppers with VR and AR technologies.

The chatbot service in the e-commerce store gives fast response, nonstop availability and accurate information. It will be also a standard feature in e-commerce.

Blockchain technology can help add transparency to supply chains. For example, Walmart, allows customers to track drone-delivered packages in real time.

Digital storefronts try to combine real and digital experience. For example, they provide a "smart in store dressing room," where the customer can virtually try on apparel.

Similar to the mentioned data base analysis for the classification of the text data in e-commerce domain we used these seven aspects to classify the innovation potential of the text content.

3. Methodology with the base of data analysis for e-commerce and innovation aspects

We propose to adopt ideas from RE studies mentioned in state of the art, to consider results regarding machine learning and AI based marketing and to combined them to develop a new method for a semi-automated requirement elicitation and applied it to the e-commerce domain. This method will use the base for the data classification related to our two main aspects mentioned in the previous chapter, e-commerce and innovation in e-commerce. Our method consists of the steps depicted in Figure 2.

Figure 2. Method Process of New Approach

Our method combines the state-of-the-art solutions with new aspects related to innovation, namely on the fourth step "innovation potential classification". The sixth step "final requirement elicitation" adapts the state of the art to prioritize the requirement and structure it. The next chapter explains each step and the application of our method in detail using a case study.

4. Results applying the method on a case study

Schindler & Parent GmbH supports this academic research project in providing the research team use cases. Our use case for the first research paper is a product whose applied results could be of interest for many Schindler &

Parent GmbH’s partners. For a generic research in requirements engineering domain that can be applied in many branches, we need requirements that can be turned into specifications. We decided to work with use case product fully automated coffee machine. So, we applied each step of our proposed new method for this product.
First, we collected the data related to e-commerce from scraping tools Mention Lyrics®, TalkWalker®, Octoparse®, and self-written program in Python using Anaconda® as development environment on fully automated coffee machines of Siemens®, Philips® and delonghi®.

At the beginning of the data collection, we needed to define the scope of data related to e-commerce. Table 2 shows our starting position for data collection. Table 2 shows leading e-commerce companies that together provide a sufficient number of data to analyze requirements. This dataset appeared large enough to support our project. After having carried up the first step of our method, we could get a first dataset with 11184 items. For further research we intend to get more items.

Ranking	Brand	2020 Brand Value	YoY % Change	Country	Sector
#1	Amazon	$220B	17.5%	United States	Retail
#2	Google	$160B	11.9%	United States	Tech
#3	Apple	$140B	-8.5%	United States	Tech
#4	Microsoft	$117B	-2.1%	United States	Tech
#5	Samsung	$94B	3.5%	South Korea	Tech
#6	ICBC	$80B	1.2%	China	Banking
#7	Facebook	$79B	-4.1%	United States	Media
#8	Walmart	$77B	14.2%	United States	Retail
#9	Ping An	$69B	19.8%	China	Insurance
#10	Huawei	$65B	4.5%	China	Tech

Table 2: Leading e-commerce companies

All collected items had to be prepared for machine learning training: data cleaning according to machine learning and artificial intelligence process (Machine Learning and Artificial Intelligence 2020) which is commonly called data preprocessing. In this second step, we cleaned up the text data (data preprocessing) from any unnecessary text components such as “,” “?” “and” “then”, which are not relevant for the NLP text analysis. We use the tool Orange data mining® for this preprocessing. Preprocess Text splits the text into smaller units (tokens), filters them, runs normalization, creates n-grams and tags tokens with part-of-speech labels. Steps in the analysis are applied sequentially and can be reordered. The preprocessing step includes: transformation, tokenization, normalization and filtering.

The third step used the base of data analysis mentioned in Figure 2. We used the following aspects: perceived convenience, habit, effort expectation, social influence, facilitating condition, performance expectation, perceived risk and trust. During this step, we manually classified the data according to these aspects. Beforehand, we trained the machine learning algorithm with a goal of 85% accuracy.

We separated the dataset into two parts, 80% for the training of the algorithm and 20% for validation / testing of the algorithm. We compared the performance of several machine learning algorithms by using the tool Orange data mining®. Figure 3 shows the commonly used machine learning algorithms & techniques [1]

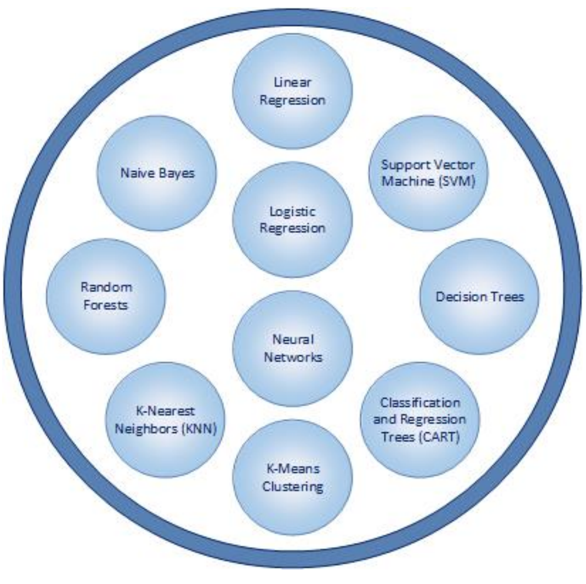

Figure 3: Commonly used machine learning algorithms & techniques

We tested with algorithm Neural Network and k-Means and found that the algorithm Neural Network gave the best performance, an accuracy of 91%. The neural network had the following configuration:

- Neurons per hidden layer: 120
- Activation: ReLu
- Solver: Adam
- Alpha: 0,00010
- Max iterations: 220

The fourth step used the base of data analysis for innovation mentioned in the second sub chapter with the following aspects: fast delivery and instant pickup, AI powered personalization, voice powered shopping, shopping using AR / VR Technology, chatbots to handle customer queries, blockchain, and digital storefronts. As in the third step we use the same dataset and classified manually the text according to the seven aspects related to innovation in e-commerce. We had a more limited text related to these aspects, because we created a subset from results of the third step. For training and validation of the dataset, we used only neural network as the algorithm. We were satisfied with the accuracy of 87%.

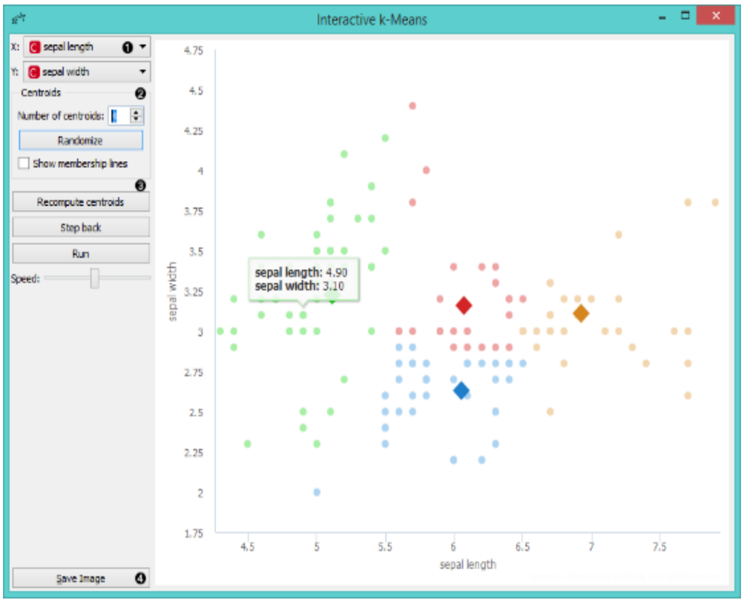

Figure 4: Dataset Automated Clustering Results Using Algorithm k-Means

For the fifth step, we used the tool Orange data mining® to automatically cluster the dataset. It uses the algorithm k-Means for the automated clustering (see Figure 4). This clustering gave a clear structure to the requirement texts processed by four previous steps.

During the last (sixth) step, we need to analyze if the machine-based clustering result is meaningful and better than the requirements structure given by the fifth step manually. We have to compare it with the machine based clustering.

The importance of a requirement can be identified by calculating the frequency of this element being mentioned by the crowd. The importance of a requirement can be identified by calculating the frequency of this element being mentioned by the crowd. It is helpful for e-commerce companies to know which requirements are relevant to their e-commerce product or tool.

5. Conclusions and future works

In this paper we present our method for a semi-automated requirement elicitation based on usable reviews on products in the e-commerce und social media domain by using machine learning for innovation and product politics.

The first results of our research have been published in this paper. This research is not finished. We have to analyze the requirements structure given by the fifth step manually and compare it with the machine-based clustering learnt by comments. (see Figure 4).

In order to get a good data mining result, we need to incorporate the correct domain knowledge. It is based on the state of the art with enhancements of the analysis of e-commerce and innovation aspects to produce valid requirements. We achieved an accuracy of 87 %.

To build upon this work, we plan to consider a more complex framework related to the e-commerce domain to improve machine learning performance. We also need to extend our research work to other marketing domains like individualized content marketing and communication.

References

[1] Batta Mahesh. 2019. *Machine Learning Algorithms -A Review*.

[2] Bohr, A. and Memarzadeh, K., Eds. 2020. *Artificial intelligence in healthcare data*. Academic Press, Amsterdam.

[3] Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., and Wang, Y. 2021. Setting the future of digital and social media marketing research: Perspectives and research propositions. *International Journal of Information Management* 59, 102168.

[4] Dwivedi, Y. K., Ismagilova, E., Sarker, P., Jeyaraj, A., Jadil, Y., and Hughes, L. 2021. A Meta-Analytic Structural Equation Model for Understanding Social Commerce Adoption. *Information systems frontiers*, 1–17.

[5] Farina, M. 2020. Using Conversation Analysis for Examining Social Media Interactions. In *Machine Learning and Artificial Intelligence*. IOS Press, 172–177.

[6] Kusumah, I., Rohleder, C., and Salinesi, C. 2022. InnoCrowd, An AI Based Optimization of a Crowdsourced Product Development. In . Springer, Cham, 267–278. DOI=10.1007/978-3-030-94335-6_19.

[7] Lim, S., Henriksson, A., and Zdravkovic, J. 2021. Data-Driven Requirements Elicitation: A Systematic Literature Review. *SN COMPUT. SCI.* 2, 1, 1–35.

[8] MANSUR, D. M., CHAMIDAH, N., OESMAN, Y. M., Putra, A. H. P. K., SULE, E. T., and KARTINI, D. 2019. Moderating of the Role of Technology Theory to the Existence of Consumer Behavior on e-commerce. *journalofdistributionscience* 17, 7, 15–25.

[9] Mei Ling Goh, Tan Seng Huat, Elaine Ang Hwee Chin, and Mei Qi Yap. 2020. CUSTOMER SATISFACTION AND BRAND SWITCHING INTENTION OF MOBILE SERVICE AMONG UNIVERSITY STUDENTS. *Management & Accounting Review (MAR)* 19, 2.

[10] Paolo Avesani, Cinzia Bazzanella, Anna Perini, and Angelo Susi. 2004. Supporting the Requirements Prioritization Process. A Machine Learning approach. In , 306–311.

[11] Robert V Kozinets. 2010. Netnography: The Marketer's Secret Ingredient How Campbell's and other marketers are using the tools of anthropology online. *MIT's technology review*.

[12] Robert V. Kozinets. 2020. Netnography Today : A Call to Evolve, Embrace, Energize, and Electrify. In *Netnography Unlimited*. Routledge, 3–23. DOI=10.4324/9781003001430-2.

[13] Ruggieri, R., Savastano, M., Scalingi, A., Bala, D., and D'Ascenzo, F. 2018. The impact of Digital Platforms on Business Models: an empirical investigation on innovative start-ups. *Management & Marketing* 13, 4, 1210–1225.

[14] Sharma, G. and Lijuan, W. 2014. Ethical perspectives on e-commerce: an empirical investigation. *Internet Research* 24, 4, 414–435.

[15] Sharma, G. and Lijuan, W. 2015. The effects of online service quality of e-commerce Websites on user satisfaction. *The Electronic Library* 33, 3, 468–485.

[16] Tallón-Ballesteros, A. J. 2020. *Machine Learning and Artificial Intelligence*. *Proceedings of MLIS 2020*. Frontiers in Artificial Intelligence and Applications Ser v.332. IOS Press Incorporated, Erscheinungsort nicht ermittelbar.

[17] Taylor, S. P. 2017. What Is Innovation? A Study of the Definitions, Academic Models and Applicability of Innovation to an Example of Social Housing in England. *JSS* 05, 11, 128–146.

[18] Zhou, L. 2021. *Essays on the Economics of Innovation: Incentives, Diffusion, and Disparity* 10813. EPFL. DOI=10.5075/epfl-thesis-10813.